VACCINE

Visual Analytics for Command, Control and Interoperability Environments
A U.S. Department of Homeland Security
Science and Technology Center of Excellence

VACCINE ANNUAL REPORT – YEAR 4
Addendum B - Publications
April 1, 2012 – June 30, 2013
Cooperative Agreement No. 2009-ST-061-CI0001

PURDUE
UNIVERSITY

Georgia Tech | **UHD** | **IU** | **FIU** | Stanford | Mississippi State | JUSTICE INSTITUTE of BRITISH COLUMBIA | Penn State | **W**

UNCCHARLOTTE | SIMON FRASER UNIVERSITY THINKING OF THE WORLD | **UBC** | Swansea University Prifysgol Abertawe | MORGAN STATE UNIVERSITY | **VT** | Texas

Universität Stuttgart | NAVAJO TECHNICAL COLLEGE | Oxford | Carleton UNIVERSITY | VICTORIA UNIVERSITY | | UNIVERSITY OF MANITOBA |

HOMELAND SECURITY UNIVERSITY PROGRAMS
TODAY'S RESEARCH & EDUCATION, TOMORROW'S SECURITY

Addendum B – Publications

Interactive Table of Contents

PURDUE UNIVERSITY - DELP

PURDUE UNIVERSITY - EBERT

SIMON FRASER UNIVERSITY

UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE

Arizona

State

University

Bristle Maps: A Multivariate Abstraction Technique for Geovisualization

SungYe Kim, Ross Maciejewski, *Member, IEEE,* Abish Malik, Yun Jang, *Member, IEEE,*
David S. Ebert, *Fellow, IEEE,* and Tobias Isenberg, *Member, IEEE*

Abstract—We present Bristle Maps, a novel method for the aggregation, abstraction, and stylization of spatio-temporal data that enables multi-attribute visualization, exploration, and analysis. This visualization technique supports the display of multi-dimensional data by providing users with a multi-parameter encoding scheme within a single visual encoding paradigm. Given a set of geographically located spatio-temporal events, we approximate the data as a continuous function using kernel density estimation. The density estimation encodes the probability that an event will occur within the space over a given temporal aggregation. These probability values, for one or more set of events, are then encoded into a bristle map. A bristle map consists of a series of straight lines that extend from, and are connected to, linear map elements such as roads, train, subway lines, etc. These lines vary in length, density, color, orientation, and transparency—creating the multivariate attribute encoding scheme where event magnitude, change, and uncertainty can be mapped as various bristle parameters. This approach increases the amount of information displayed in a single plot and allows for unique designs for various information schemes. We show the application of our bristle map encoding scheme using categorical spatio-temporal police reports. Our examples demonstrate the use of our technique for visualizing data magnitude, variable comparisons, and a variety of multivariate attribute combinations. To evaluate the effectiveness of our bristle map, we have conducted quantitative and qualitative evaluations in which we compare our bristle map to conventional geovisualization techniques. Our results show that bristle maps are competitive in completion time and accuracy of tasks with various levels of complexity.

Index Terms—Data transformation and representation, data abstraction, illustrative visualization, geovisualization.

✦

1 INTRODUCTION

As data dimensionality increases, the encoding of variables and their relationships is often abstracted down to a representative subset for analysis in a single display, or dispersed across a series of coordinated multiple views [1–3]. Moreover, many techniques have been developed to visually encode multiple data attributes/variables for each data sample to enable interactive analysis, ranging from discrete glyph attribute encoding [4] to more spatially continuous color, transparency, and shading encodings [5–7]. As the number of visualized variables increases, the amount of information that can be effectively displayed becomes limited due to over-plotting and cluttering [8]. This is especially a problem in geographical visualization as a key attribute of the data is the location within the two-dimensional map space.

In geographical visualization, data can be described at any given location on a map. The data being described can come from an aggregated measurement, a direct event occurrence, or various other means. In dense data sets, plot-

- *S. Kim, A. Malik and D. S. Ebert are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. E-mail: {inside|amalik|ebertd}@purdue.edu.*
- *R. Maciejewski is with Arizona State University, Tempe, AZ, USA. E-mail: rmacieje@asu.edu.*
- *Y. Jang is with Sejong University, Seoul, South Korea. E-mail: jangy@sejong.edu.*
- *T. Isenberg is with INRIA, Saclay, France. E-mail: tobias.isenberg@inria.fr.*

ting events as symbols on the map (e. g., Fig. 1(a)) leads to cluttering and is often unable to convey a meaningful sense of event magnitude within the data. Aggregation of the data by defined boundaries, such as county or census tract boundaries (e. g., Fig. 1(b)), leads to a loss of specificity in data location and runs afoul of the Modifiable Areal Unit Problem [9]. Furthermore, it is known that the level of data aggregation can affect aspects of task complexity such as information load and the user's ability to recognize patterns within the data [10]. In order to combat problems associated with areal aggregation, dasymetric mapping focuses on using zonal boundaries that are based on sharp changes in the statistical surface being mapped [11]. However, even when grouping data into small spatial quadrats, data can either be over-aggregated or under-aggregated. A third option is to estimate the discrete event points as a continuous function (e. g., Fig. 1(c)); such a mapping, however, only allows for the use of color as a means of representing data variables. As an encoding based on underlying network data, Fig. 1(d) shows a traditional line map. However, its representation is still restrained by the color and thickness of the lines.

In order to increase the amount of information that can be visualized within the constraints of a thematic map, this paper explores a novel method of multivariate encoding. Inspired by ideas of symbolic encoding from Spence [12] and choices of visual encodings by Wilkinson [13], we have developed the bristle map (Fig. 1(e)), a novel method for the aggregation, abstraction, and stylization of geographically located spatio-temporal data. The bristle map consists of a series of straight lines extended from and connected to

Fig. 1: Data abstraction in geovisualization. In this image, we show crimes in West Lafayette and Lafayette, Indiana where the blue line represents the Wabash River. (a) Plotting events as points. (b) Aggregation of points by areal units. (c) Approximation of a continuous domain from point sampling. (d) Approximation of a continuous domain using solid lines applied to roads. (e) Our abstraction using a series of bristle lines applied to roads.

linear map elements (roads, train lines, subway lines, etc.) that have some contextual relationship with the data being visualized. We vary these lines with respect to their color, length, density, and orientation to allow for a unique encoding scheme that can be used to create informative maps. With respect to the other representations shown in Fig. 1, our technique utilizes the underlying geographical context as a part of its symbology, thereby directly incorporating geographical elements within its encoding scheme. One of the major advantages of the bristle map technique is that the basis domain of the data (e. g., street network) remains highly visible regardless of the color scale being used. If one compares Fig. 1(c) and (e), the street network in Fig. 1(e) is clearly visible because the lines only 'bristle off' to one side, whereas in Fig. 1(c) some streets are hardly discernible due to the dark colors.

To demonstrate our technique, we focus on categorical spatio-temporal event data (e. g., emergency department logs, crime reports). In such data, events consist of locations in time and space where each event fits into a hierarchical categorization structure. These categories are typically processed as time series and snapshots of time are aggregated and typically visualized on a choropleth map [14]. Past work [6, 15] has shown that the use of kernel density estimation [16] is highly suitable in the spatial analysis of such data. Thus, our approach incorporates kernel density estimation as a means of estimating the underlying distribution of spatio-temporal events. Using the estimated distribution in an area for a given category (or categories) and temporal unit, we incorporate the underlying geographical network structure into the visual encoding. Bristles are extended from this underlying structure, and the color, length, density, transparency, and orientation of each bristle is mapped to a particular variable (or set of variables). Schemes presented in this paper include combinations of the following mappings:

- length, density, and color as data magnitude,
- orientation and coloring for bivariate mapping,
- color and length for bivariate mapping,
- color and density for bivariate mapping, and
- length and transparency for temporal variance.

Given the available parameters for visual encoding within the bristle map, other encodings also exist, which illustrate the flexibility and power of our technique. Our work focuses on showing how bristle maps can be used to show spatial and temporal correlations between variables, encode uncertainty in a unique way, and maintain geographical context through linking our visual encoding directly to geographical components. As such, the bristle map is a powerful multivariate encoding scheme that is adaptable to various attribute encodings to create richly informative visualizations.

2   RELATED WORK

Many techniques in multivariate data visualization focus on a means of reducing clutter and highlighting information through a variety of approaches including filtering (e. g., [17]), clustering (e. g., [18]), and sampling (e. g., [19]). In this section, we focus particularly on techniques within geographical visualization for improving the understanding of thematic/statistical maps, as Wilkinson [13] noted that the problem of multivariate thematic symbology for maps is that they are not only challenging to make, but also challenging to read.

In geographical visualization, the most common means of data representation is the choropleth map in which areas are shaded or patterned in proportion to a measured variable. Such maps are typically used to display only one variable, which is mapped to a given color scale. Other research has focused on encoding multivariate information into choropleth maps (such as uncertainty) with textures

Fig. 2: The bristle map generation pipeline. Beginning with data events, a continuous abstraction is created. We also create a topology graph from contextually important linear features (in this case roads). Next, bristles are extended from these features based on the continuous abstraction and the topology. Clutter reduction is performed when generating each bristle, and finally the resultant bristle map is generated.
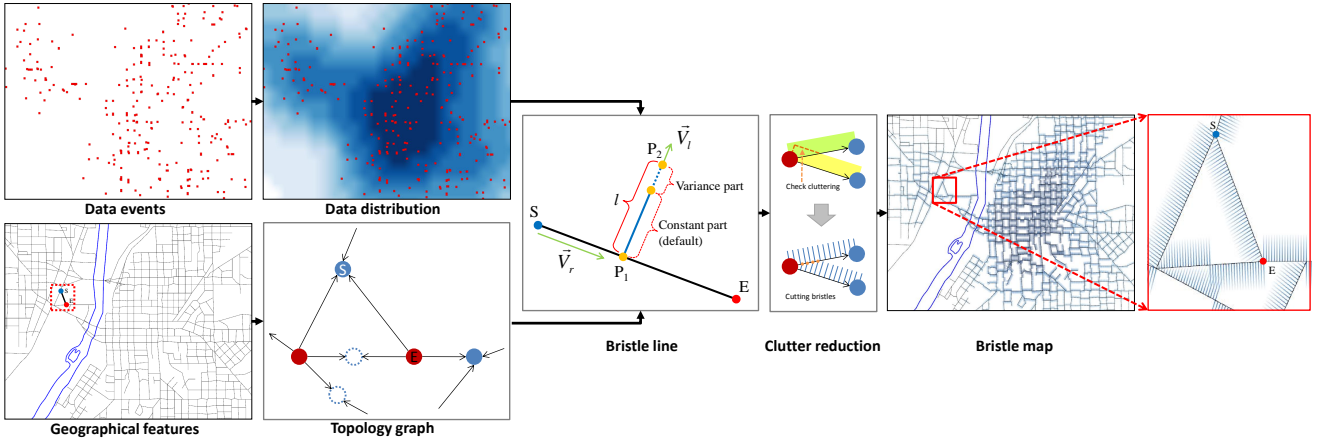
and patterns [20], creating bivariate color schemes for visualizing interactions between two variables [21, 22], or animating choropleth maps to enhance the exploration of temporal patterns and changes [23]. We present bristle maps as a robust alternative to these schemes in which multivariate attributes are instead mapped to a variety of graphical properties of a line (length, density, color and orientation), as opposed to utilizing a bivariate color scheme, texture overlays, or animation.

More recent geographical visualization techniques have included extensions to choropleth mapping ideas. Hagh-Shenas et al. [24] compared the effectiveness of visualizing geographically referenced data through the use color blending (in which a single composite color conveys the values of multiple color encoded quantities) and color weaving methods (in which colors of multiple variables are separately woven to form a fine grained texture pattern). The results from their study indicate color weaving to be more effective than color blending for conveying individual distributions in a multivariate setting. Saito et al. [25] proposed a two-tone pseudo coloring method for visualizing precise details in an overview display. Under this scheme, each scalar value is represented by two discrete colors. Sips et al. [26] focused on revealing clusters and other relationships between geo-spatial data points by their statistical values through the over-plotting of points. This work was later extended [27] to combine a cartogram-based layout to provide users with insight to the relative geo-spatial positioning of the dataset while preserving cluster information and avoiding over-plotting. Other cartogram techniques include the WorldMapper Project [28] which is used to represent social and economic data of the countries of the world. In each of these, novel data visualization techniques are created; however, the distortion of spatial features (country boundaries, roads) is often undesirable. While these techniques focus on displaying large amounts of aggregate data on small screens, our technique focuses on enhancing details of geographical context within the data. A similar concept of preserving data context is found

in Wong et al.'s [29] GreenGrid in which they visualize both the physics of the power grids in conjunction with the geographical relationships using graph based techniques.

Along with the previously described map schemes and cartogram distortions, there has been work in the use of heatmaps based on spatial data. Fisher [30] applied heatmaps to visualize the trends of the interactions of users with interactive maps that are based on their view of the geographic areas. Maciejewski et al. [6] used heatmaps as one of the tools to find aberrations or hotspots that facilitate the exploration of geo-spatial temporal datasets. Work by Chainey et al. [15] illustrated a number of different mapping techniques for identifying hotspots of crime and demonstrated that kernel density estimation provides analysts with an excellent means of predicting future criminal activities.

In conjunction with previous visualizations, other research has focused on expanding the dimensionality of the data being displayed by utilizing three-dimensional visuals. Van Wijk and Telea [7] utilized color and heightfields to visualize scalar functions of two variables. Tominski et al. [31] explored embedding 3D icons into a map display as a means of representing spatio-temporal data. In contrast, our work focuses on a two-dimensional encoding scheme that incorporates a variety of the visual variables described by Bertin [32] and Wilkinson [13] as a means of representing multivariate data.

Finally, it is important to note that our technique is akin to traditional traffic flow maps (e.g., Fig. 1(d)) seen in a variety of atlases; however, provides more generalized schemes. In traffic flow maps, the amount of data that can be displayed is restrained by the color and the width of the line representing linear elements (i.e., roads) on the map. Our work is similar to that of the traffic flow maps in that we utilize width (specifically, matched to the length in our bristle maps) and color as underlying visual variables of our encoding. However, our work also incorporates bristle density as a means of further encoding parameters. In the following sections, we compare our encodings to a variety of methods including the point, color, and flow line maps.

3 BRISTLE MAP GENERATION

In Fig. 1, we developed our motivation for the need to directly incorporate geographic features to the underlying data in order to better preserve contextual information. It is clear that the aggregation of data into arbitrary geographical areas obscures data, while the continuous approximation of an underlying data source can lead to incongruent mappings with respect to geographic features. Furthermore, both these mappings are limited in the fact that only color and texture are available for variable encoding, limiting the amount of data that can be displayed to either a single variable or possibly two variables in the case of a bivariate color map. The goal of this work is to create visual encodings for higher order structures.

The bristle map was inspired by the Substrate simulation of Tarbell [33] and abstract renderings of map scenes in work by Isenberg [34]. Given these images, our work focuses on using the underlying visual properties to intelligently encode information for display. In *The Grammar of Graphics* [13], Wilkinson discusses the combination of several perceptual scales into a single display. Here, he notes the idea of separable dimensions of the data is a key issue, where discriminations between stimuli are of key importance in the visualization. The Substrate aesthetic directly lends itself to this approach as color, line length, and orientation are distinct classes within Wilkinson's table of aesthetic attributes and each of these visual parameters directly contributes to the substrate aesthetic.

Fig. 2 illustrates the bristle map generation pipeline. Given underlying data events, we compute a continuous distribution. We also create a topology graph from given geographically relevant linear content for clutter reduction described in Section 5. As an example of geographical content, if the underlying data was water pollution we could use a city sewage map for the geographic components, for our crime data examples we use roadways. Each linear geographic component consists of a series of line segments, and we extend *bristle lines* from these line segments. These bristle lines emerge perpendicularly from the underlying geographical line segment and are allowed to vary in length, density, color, transparency, and orientation, to facilitate multivariate data encoding. The third stage of the bristle map generation pipeline (Fig. 2) illustrates the bristle line concept for each geographical line segment, $\overline{SE}$, and $\overline{P_1P_2}$ defines our generated bristle line. Each bristle line is created using the vector equation of a line as shown in Equation 1.

$$P_2 = P_1 + \vec{V_l}L_l = P_1 + \vec{V_l}(t \times L_{lmax}) \qquad (1)$$

Here, P_1 is a point on the contextually relevant geographic line segment, $\overline{SE}$, $\vec{V_l}$ is a unit vector perpendicular to the line $\overline{SE}$, and L_{lmax} is the maximum length of the $\overline{P_1P_2}$. L_l is the length of the $\overline{P_1P_2}$ determined by a parameter t.

Each line from P_1 to P_2 is drawn in such a manner that it will either encode different properties of a multivariate data set, or use a data reinforcement technique where properties are encoded to the same variable to provide redundant cues. We utilize three encoding properties for each bristle; length,

TABLE 1: Parameters, corresponding variables, and ranges.

Parameters	Potential variables	Range
Base position (P_1)	Geographic location	(Double, Double)
Length 1 (constant portion)	Data magnitude	Double
Length 2 (variance portion)	Temporal variance, accuracy	e. g., monthly/yearly
Color	Data magnitude	Discrete, Continous
Transparency	Temporal variance, accuracy	Double [0.0, 1.0]
Orientation ($\vec{V_l}$)	Temporal difference, data type	Clock-wise, Counter clock-wise
Density	Average data magnitude on an area ($\overline{SE}$)	Double

color, and orientation. The length of a line $\overline{P_1P_2}$ is separated into two portions; a constant component, which is proportional to the magnitude of the variable being encoded, and a variance component. It captures temporal variance or other properties such as level of certainty. The color of a bristle $\overline{P_1P_2}$ is proportional to the underlying variable distribution to be encoded at point P_1. When the variance component is used, its transparency is adjusted as a means of visually distinguishing it from the constant component. Orientation of the bristle line is always perpendicular to $\overline{SE}$ and is utilized for bivariate comparison (i. e., day/night, two data types) and/or clutter reduction. To summarize, length and color represent a local data magnitude property at point P_1. We also choose to encode redundant information into the density of the number of bristles placed on a given line segment where the density of the bristles along $\overline{SE}$ is decided by an average data value on a line segment $\overline{SE}$.

For each visual encoding, the underlying data is assumed to be continuous over a given geographical segment, such that for all points between any two nodes on the underlying contextual geographic structure, a data distribution value is associated with the point. In the case of a discrete data set (e. g., crime locations), the choice of an appropriate means of data interpolation with regards to the underlying geographic information is dependent on the data analysis being performed. Based on the recommendations of Chainey et al. [15], we apply a kernel density estimation (KDE) [16] to approximate the underlying distribution of crimes over the geographic features. The kernel density estimation procedure used is defined by the following equation:

$$\hat{f}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{\mathbf{x}-X_i}{h}\right) \qquad (2)$$

Here, the window width of the kernel placed on point x is proportional to a window bandwidth, h, and the total number of samples, N. We utilize the Epanechnikov kernel [16], Equation 3:

$$K(\mathbf{u}) = \frac{3}{4}(1-\mathbf{u}^2)1_{(||\mathbf{u}||\leq 1)} \qquad (3)$$

where the function $1_{(||\mathbf{u}||\leq 1)}$ evaluates to 1 if the inequality is true and zero for all other cases.

Thus, given a multivariate dataset where locations in space and time correspond to a series of categorized events,

we can create bristle maps that encode various properties of the data. Note that this technique relies on the data being contextually relevant to an underlying geographical network. For example, crime event data with its 2D geographical coordinates is recorded and hence defined by addresses on streets; thus it is contextually relevant to a street network. Data sets in which this contextual relationship does not exist should utilize other visual encoding schemes. Table 1 shows the parameters in our bristle map and their corresponding potential variables being encoded to each parameter. In the following section, we present a series of potential parameter combinations for various bristle map encodings and discuss the various results.

4 ENCODING SCHEMES

The bristle map is a powerful visual encoding scheme that lends itself to a variety of data encodings, examples of which we present next. For demonstration purposes we employ categorical spatio-temporal police reports collected in Tippecanoe County (specifically West Lafayette and Lafayette, IN, USA), from 1999 to 2010. The data set contains the date, time, crime type (e.g., armed/unarmed aggravated assault, armed robbery, burglary, homicide, noise, other assaults, rape, rape attempted, residential entry, robbery, theft, vandalism, and vehicle theft), and the address of each recorded criminal event. Note that other datasets can be easily encoded with bristle maps, and our choice of data was only made to illustrate the technique.

Utilizing this multivariate crime data set we discuss potential encoding schemes for multivariate spatio-temporal data. We then provide illustrations of each described encoding scheme with respect to our crime data set. Encoding schemes presented in this section include the use of bristle color, length, and density to encode data magnitude, the use of bristle orientation to inform temporal comparison, and the encoding of temporal variance in the bristle lengths.

4.1 Color, Length, and Density as Data Magnitude

Here, we discuss our technique for encoding the color, length, and density of the bristles into two separate variable groups. As both color and length (size) fall into two distinct categories of aesthetics according to Wilkinson [13], the use of separate variables for both categories allows for a distinguishable visual data encoding. In both cases, we assign data magnitude to both a color scale and a length scale. We note that such an encoding scheme has the potential to portray data more effectively than visualizations that map each data variable to a single display parameter. As noted in the arguments for the use of redundant color scales by Rheingans [35], the use of different display parameters is able to convey different types of information. Furthermore, by combining encodings in a redundant manner, it is possible to reinforce the encoding scheme. The utility of redundant color scales was confirmed by Ware [36].

In our encoding scheme, each bristle line's length, L_l, is calculated using Equation 4 based on a parameter, t, and

the maximum length, L_{lmax}.

$$L_l = t \times L_{lmax} = (\alpha \times \kappa_{P_1} + \beta \times \upsilon_{P_1}) \times L_{lmax} \qquad (4)$$

For this visual encoding of the bristles, the parameter t is defined by the ratio of the data value at P_1, which we call κ_{P_1}, the ratio of the temporal variance at P_1, υ_{P_1}, and a set of tuning parameters (α and β) which provide weights to the constant and variance components as shown in Fig. 2. In this work, we use α=1.0 and β=0.3. Note that the choice of encoding the variance at a 30% value was chosen through trial and error by generating visualizations that the authors found to be the most useful and aesthetically pleasing. For problems where determining exact data values from the visual encoding is required (as opposed to approximating high and low rates), the variance portion is removed from the equation entirely by using β=0.0. As such, by creating the encoding scheme with diverse parameters, we are able to generate more aesthetic choices and visualizations. It is important to note that not all encodings will be appropriate and are most likely task dependent.

The L_{lmax} portion of Equation 4 is defined in Equation 5.

$$L_{lmax} = \rho \times log_b \left(\frac{1}{N_r} \sum_{i=0}^{N_r-1} L_{\overline{SE}} \right) \qquad (5)$$

In this equation, we take the average length of all line segments (where N_r is the total number of line segments in the map) and calculate L_{lmax} using a non-linear function such that the length of bristle lines does not grow in an unbounded manner when zooming in. Moreover, L_{lmax} is modified by the parameter ρ, where ρ is the ratio of the current zoom level to the initial zoom level, to decouple our technique with the zoom level. In this work, we use b=15 for the base of a log function.

Next, we determine the number (or density) of bristles, N_l, to be drawn on each line segment $\overline{SE}$ using Equation 6.

$$N_l = \rho \left(\frac{\zeta}{\lambda} L_{\overline{SE}} \right) \kappa_{\overline{SE}} \qquad (6)$$

Here, N_l is calculated using two user-defined constants λ and ζ, where λ is the unit geographical length (distance) and ζ is the number of bristle lines per unit geographical length. We use λ=0.0009 and ζ=3-15 in our current visualization. As the bristle density may also be used to encode data magnitude parameters in bristle map generation, N_l should be proportional to the ratio of average data value on $\overline{SE}$, $\kappa_{\overline{SE}}$. Moreover, we also apply ρ such that N_l will be independent of the zoom level to preserve the extent of density.

For color, we allow users to choose either a continuous or a sequential color scheme from Color Brewer [37]. Then, data is linearly mapped to a probability that a crime of type A will occur at geographic point B, where the probability is estimated from the underlying data distribution using kernel density estimation as described in Section 3.

Fig. 3 illustrates our length, density, and color encoding using the previously described crime data set. Burglary is

Fig. 3: (Left) Our bristle map encodes burglary rates with both bristle color and bristle density. (Right) A line map encoding burglary rates as both line color and line thickness. Compared to the line map, our bristle map provides a distinguishable visualization by incorporating bristle density. For example, bristle lines on the right top area are easily identified, whereas thickness in the line map on the same area is too small to clearly be perceived.

encoded with the red color scheme, and color is proportional to the probability (calculated from the underlying point distribution using kernel density estimation) that a burglary occurred at a given location. Fig. 3 (left) shows our bristle map encoding for burglary rates with a color scheme and bristle density, and Fig. 3 (right) shows a line map encoding the same information with a color scheme and line thickness for comparison to our bristle map. Compared to this line map, our bristle map provides the advantages of additional dimensionality through the density of bristle lines. In this scheme, one is able to easily encode two variables in different combinations of bristle map parameters (i.e., color and density with a constant length, color and length with a constant density), and provide users with distinguishable visual parameters that seem to focus attention to various details.

4.2 Multivariate Encoding: Separating Length, Density and Color, and using Orientation

In the previous section, we illustrated how our method can be utilized for univariate encoding by using a redundant encoding scheme. However, a major benefit of bristle maps is the ability to encode multivariate attributes. One example of this is seen in day versus night time comparison.

Here, one can utilize the orientation to separate two temporal components of a single variable by mapping the temporal components to different orientations of the geographic feature. For instance, it is likely that the rates of data variable will be different with respect to day and night occurrences. We illustrate this visual encoding in Fig. 4. We separate the events into day (6:00 am–6:00 pm) and night (6:01 pm–5:59 am) and map the daytime rates

to red and one orientation, and nighttime rates to blue and the other orientation. In Fig. 4 (right) we illustrate a bristle map encoding of one variable (burglary) during 2009 where length, density and color represent the magnitude of the burglary as well as the encoding of day and night parameters is explored as line orientation.

In Fig. 4 (right), we show areas of high/low nighttime crime, high/low daytime crime, and combinations there within. In contrast, a traditional heatmap using a univariate color scheme can only show either daytime crime (Fig. 4 (left top)), or nighttime crime (Fig. 4 (left middle)). Hence, several heatmaps are needed to see day and night variations as shown in Fig. 4 (left column). Viewers must mentally combine the images to locate regions of the map that have high crime levels at daytime and nighttime, thereby increasing their cognitive load.

Another means of reducing the cognitive burden would be to create a heatmap of the difference between night and day. Fig. 4 (left bottom) shows the difference of day and night data, and the divergent color scheme shows where high daytime or high nighttime crimes occur. For instance, in Fig. 4 (left bottom) the right area indicates higher rates during day, the left area shows higher rates during night, and the border area between the blue and red color schemes only indicates that day and night rates were approximately equal, regardless of them being low or high. Moreover, you need other color maps to explore areas where one occurs similarly high or low during day and night time.

Bristle map encodings have benefits in this situation. When we explore a daytime versus nighttime bristle map in Fig. 4 (right), we see that there exists distinct temporal profiles along the road lines where we see exclusively dominant areas during either day or night. For instance,

Fig. 4: Encoding daytime versus nighttime variations. (Left column) From top to bottom, color maps showing day, night and the difference of day and night burglary rates. (Right) Our bristle map separating the burglary rates into their day and night components with opposite orientation along roads. Note that a color map cannot present two components (i. e., both daytime and nighttime burglary rates) at the same location, hence three color maps are needed to see day and night variations simultaneously. Our bristle map can present such information within one bristle map by using different orientations of bristle lines.

see the diagonal road from the top center to the right center (Main Street, Lafayette, IN) showing that daytime burglary dominates along this road. Another observation is made on the horizontal road at the center of the map (Central Street, Lafayette, IN). Along this road, daytime burglary rates increase from left (west) to right (east), whereas nighttime burglary rates decrease from left to right. For the center area in Fig. 4 (left bottom), where the blue and red color schemes meet, we also see in Fig. 4 (right) that it has relatively equally high rates during both day and night. Such a comparison allows people to understand the differences between the data; however, when subtracting, areas of nearly equal daytime and nighttime crimes will be colored the same. Thus, areas that are safe during both day and night, and areas that are highly dangerous during both day and night will appear the same in the difference color map. In contrast, bristle maps allow viewers to quickly observe trends related to both day and night.

Another example of multivariate encoding using our bristle map is done by separating and/or combining bristle parameters. For instance, bristle density (or length) encodes a variable A, and color encodes a variable B while being presented on one orientation. Similarly, another two variables (C and D) could be encoded and presented on the other orientation. However, this type of parameter combination should be determined carefully so as not to increase viewers' cognitive load. Its effectiveness would depend on several factors such as data type and analysis

purpose. In Section 6, we conduct experiments to explore the effectiveness of different parameter combinations.

4.3 Encoding Data Variance

As introduced in Fig. 2, each bristle can include a portion generated for temporal variance of data, see Equation 4. To present the temporal variance of the data over time, we compute both the monthly and yearly mean and variance values. For a given discrete data set during time periods N_T, we first calculate continuous distributions over time. Then, we determine mean and standard deviation values with respect to the underlying data distribution for the entire data set over a given temporal aggregation. Thus, we calculate the mean μ and variance σ values from time varying data K_i, where $i \in [0, N_T - 1]$. Note that μ and σ are computed only once as they represent constant values for a given dataset. Mean and variance values for each grid point j are calculated using Equation 7 and 8, respectively. Variance is then used to weight the parameter β in Equation 4 such that given the data magnitude at the current time K_{cur}, we compute the ratio of variance at the current time, $\tilde{\sigma}$ as shown in Equation 9. As such, the parameter t in Equation 4 can be detailed as shown in Equation 10 to represent the length of bristle lines with respect to temporal variance.

$$\mu[j] = \frac{1}{N_T} \sum_{i=0}^{N_T - 1} K_i[j] \qquad (7)$$
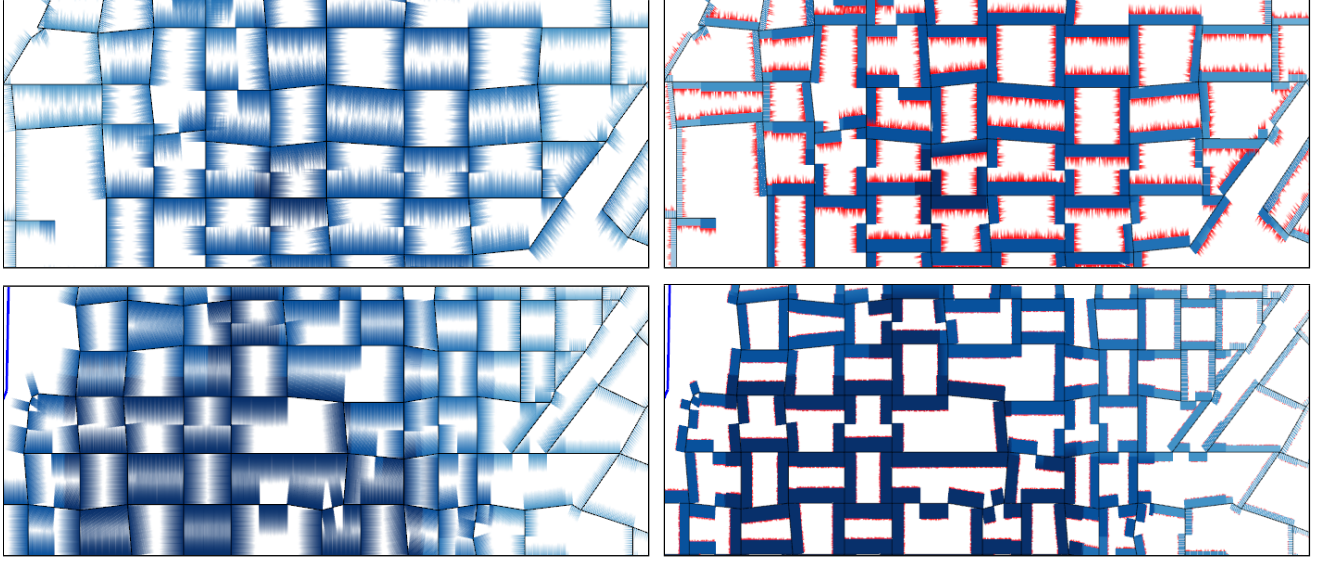
Fig. 5: Encoding data variance of vandalism-graffiti in Lafayette, IN, USA, in 2010 creating an uncertainty aesthetic. Yearly variance of vandalism-graffiti is represented in (a) a residential area and (c) a commercial area without distinguishing the variance component in the bristle length. (b) and (d) show the results using a highlight color for the variance portion and full alpha values for the constant portion of bristle lines. Here, we clearly see that our bristle map can encode the temporal variance and create an uncertainty aesthetic using the variance component.

$$\sigma[j] = \sqrt{\frac{1}{N_T} \sum_{i=0}^{N_T-1} (\mu[j] - K_i[j])^2} \qquad (8)$$

$$\tilde{\sigma}[j] = \frac{1}{\sigma[j]} |\mu[j] - K_{cur}[j]| \qquad (9)$$

$$t = \alpha \times \kappa_{P_1} + \beta \times \upsilon_{P_1} = \alpha \times \kappa_{P_1} + \beta \times \left(\frac{\tilde{\sigma}_{P_1}}{\tilde{\sigma}_{max}}\right) \qquad (10)$$

Furthermore, the variance term, υ_{P_1}, in parameter t in Equation 4 can also be revised to encode an uncertainty factor by using randomness. We may also encode an uncertainty factor by using color and transparency to enhance the variance component. When using color and transparency, we use a highlight color for the variance component, and then fade out the variance component over the bristle length with a full alpha value for one end point and an alpha value weighted by the variance for the other end point. The constant portion of the bristle is assigned an alpha value of 1 to both end points as it represents an exact data value. Hence, according to the data type and analysis purpose, the encoding of parameter t and the use of the variance portion can be different and should be assessed with respect to the visual message trying to be conveyed. Fig. 5 illustrates the application of encoding the data variance of vandalism with the uncertainty factor. In Fig. 5(a, c), we use the same color scheme for the constant and variance portions of bristle lines. To enhance the variance component in Fig. 5(b, d) we highlighted the variance portion in a different color and assigned full alpha values for the constant portion of bristle lines. Fig. 5(a, b) show the same area. In this area, the bristle length shows large fluctuations, indicating a high

yearly variance Fig. 5(c, d) show another area. In this area, the bristles are of a nearly constant length, indicating low yearly variance. When considering that the area in Fig. 5(a, b) includes residential areas, while the area in Fig. 5(c, d) includes the downtown Main street, an art theater, and the City Hall in Lafayette, IN, our bristle map shows that the residential areas have higher yearly variance of vandalism (graffiti) when compared to commercial areas.

5 BRISTLE CLUTTER REDUCTION

Although our bristle map can encode various characteristics from multivariate data, it often suffers from clutter around the intersections of road lines. In order to minimize cluttering, we employ two strategies in our bristle map generation pipeline (Fig. 2); 1) using topology among road lines to determine bristle orientation to minimize clutter and 2) cutting bristle lines crossing neighbor road lines.

5.1 Using Topology

Each bristle map contains an underlying topology of the contextual geographic network that the data is mapped to. In the topology graph, each node is defined as either 'outward' or 'inward' as illustrated in Fig. 6. Using the topology graph, we choose each segment's bristle line orientation such that the overlap of the bristles at intersections will be minimized, thereby reducing the clutter. If the encoding scheme requires both sides of the edge to contain bristles, then clutter at each intersection is inevitable. However, in cases where bristles map to only one side of an edge, we use the right-hand rule to decide the orientation. Hence bristle lines on edges connected to neighboring outward and inward nodes are generated in a manner that provides a reasonable reduction in clutter (Fig. 6).

Fig. 6: To minimize clutter, a topology graph consisting directed edges as road lines and outward (red) and inward (blue) nodes on the intersection of lines is used to decide bristle line orientation.

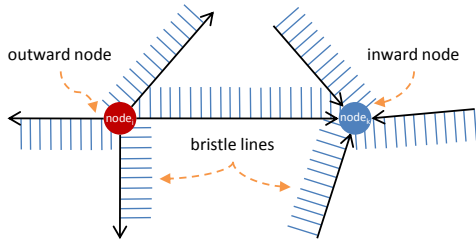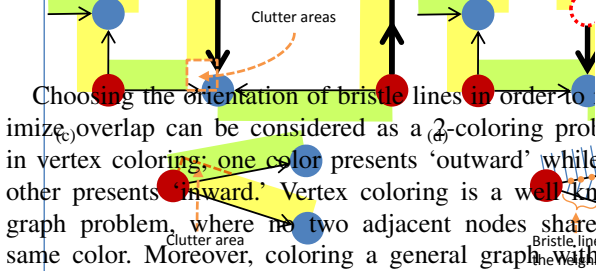Choosing the orientation of bristle lines in order to imize overlap can be considered as a 2-coloring problem in vertex coloring; one color presents 'outward' while the other presents 'inward.' Vertex coloring is a well-known graph problem, where no two adjacent nodes share the same color. Moreover, coloring a general graph with the minimum number of colors is known to be an NP-complete problem. In our case, the minimum number of colors should always be 2 but such 2-colorability is not guaranteed for general road lines. While deciding the orientation of bristle lines, we often have undesirable topology generating inevitable overlap of bristle lines. Fig. 7 (upper row) shows such a bad topology example and our strategy to solve this issue. In Fig. 7(a), we see two clutter areas caused by an undesirable configuration of neighbor nodes which guarantee bristle overlap. To solve this, we consider the addition of a virtual node in a topology graph as shown in Fig. 7(b), thereby allowing for an orientation switch midway across the edge and reducing the clutter. For neighboring two inward nodes (blue), we add a virtual outward node (red dotted circle) at the road line connecting two inward nodes resulting in splitting bristle lines on the road line. Similarly, a virtual inward node (blue dotted circle) is added for neighboring two outward nodes (red).

5.2 Avoid Crossing Neighbors

Another cluttering case is illustrated in Fig. 7(c). When two road lines intersecting with less than a 90° angle have bristle lines, some of the bristle lines overlap as illustrated in Fig. 7(c). For this case, we forbid bristle lines to cross neighbor road lines by placing the end point of a bristle line on the neighbor road line as shown in Fig. 7(d). We first check the intersection of bristle blocks (colored boxes in Fig. 7) for the current road line on which we are generating bristle lines and its neighboring road lines by using the topology graph. If the blocks are intersected, we then check if a bristle line crosses the neighbor road lines by utilizing the intersection algorithm of 2D line segments [38]. This idea is based on the theory of amodal completion (or amodal perception) [39] in psychology that describes how the human visual system completes parts of an object even when it is only partially visible. Although the length of a bristle line represents data magnitude, benefits from

Fig. 7: Two pairs of the cluttering cases and our methods to minimize clutter. Colored box areas on a side of each edge line indicate the orientation for bristle lines. (a) Case 1: bad topology, where two inward nodes (blue) share a line and two outward nodes (red) share a line, generates inevitable clutter. (b) Virtual nodes (dotted circles) are added to split an edge line. (c) Case 2: a small angle between edge lines causes a clutter area. (d) Bristle lines crossing a neighbor edge line are cut on the neighbor line.

cutting the length to avoid clutter dominate the side effects from data misunderstanding that could be caused by clutter. Moreover, when using redundant encoding utilizing bristle length and density as data magnitude, bristle density could help viewers complete parts of the bristle lines. Fig. 8 shows four image pairs before and after applying our clutter reduction strategies. Some improvements could also be considered in the future. For instance, our strategies still generate cluttered bristle lines in cases where road lines are very dense or close to others. We perform experiments in Section 6 to see how people understand the differences before and after clutter reduction. Here we note that the experiments performed were for comparison and identification tasks. In these task types, line direction (as will be shown in the experiments) had little impact on the user results. However, in a cluster/delineate task in which users are asked to segment the data, the splitting of direction may influence the user's perception of cluster boundaries. As such, we recommend that map designers take caution in employing this scheme and use it only in appropriate map contexts. Future work will explore other schemes and design issues to handle neighbor crossings and influence on map design.

6 EVALUATION

To evaluate the effectiveness of our bristle maps, we conducted two quantitative controlled experiments. These studies are both comprised of an introductory session, and a training session. In the first study, five tasks were conducted to evaluate the efficiency of bristle maps compared to existing visualization methods (point, color (kernel density estimated, KDE), and line maps as shown in Fig. 1(a), (c), and (d)) and post-task questionnaires for qualitative feedback. In the second study, two tasks were conducted to evaluate the accuracy of users in estimating values from each of the map types (point, KDE, bristle and line) as well as evaluating the perceived aesthetics of each image. Prior to each study, a pilot study was also conducted to
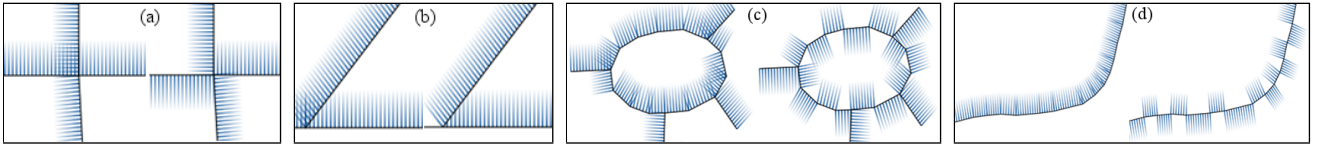
Fig. 8: Before/after image pairs of our clutter reduction. Each pair shows a case of (a) changing bristle orientation using topology, (b) cutting bristle lines crossing neighbor road lines, (c) circular roads, and (d) curved roads.
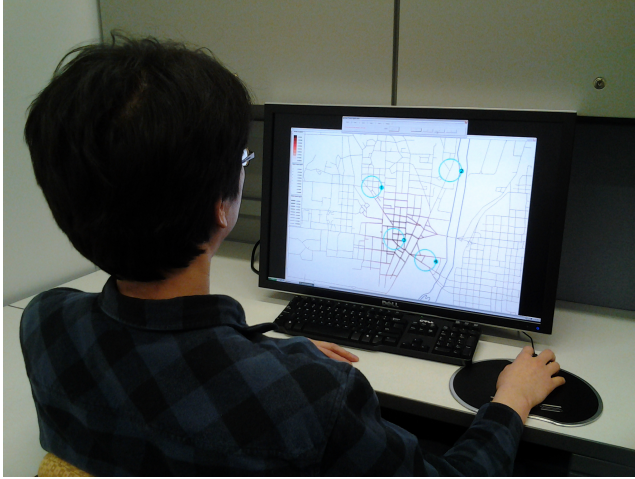
Fig. 9: Example setup for our experiment.

ensure that each task contains a fair comparison among the techniques.

Participants: In the first study, thirty graduate students (23 males, 7 females) in engineering, science, and statistics from our university participated in the study. All participants reported that they had experience in visualizing data on geographical maps using colors or icons (e. g., paper maps, online map services). The experience varied from almost daily (11 participants), 1–2 times a week (17 participants) to 1–3 times a month (2 participants). For the identification/accuracy tasks and aesthetic comparisons (Tasks 6 and 7), a secondary study was run on twenty-six undergraduate students in engineering from our university.

Apparatus: The experiment was performed on a 30" monitor using our experimental application running on Windows XP, as shown in Fig. 9, where all visualizations were generated with 2228×1478 resolution. Each visualization was overlaid with numbered circles as shown in Fig. 9. Participants selected one of the numbers to answer the question in each trial using buttons in the interface panel on the top of the screen. Criminal incident reports collected in West Lafayette and Lafayette, Indiana, USA from 1999 to 2010 were used in each trial, but different types of crimes were selected to generate visualizations in the training phase and in the actual study.

Design: We employed a repeated measure design of tasks incorporating variations of the images shown in Fig. 1(a), (c), and (d) and line maps similar to those of Fig. 3 (right). Table 2 shows the number of data sets, techniques (cases as shown in Fig. 8 for Task 5), and trials in each task. For example, in Task 1 we utilized 18 different data sets to

compare 5 different techniques (i. e., point map, color map, line map, and bristle maps using two different encoding schemes). Hence, each participant performed $18 \times 5 = 90$ trials in Task 1. In Task 3, we compared 6 different techniques (i. e., point map, bivariate color map, line maps in two different encoding schemes, and bristle maps in two different encoding schemes) with 15 data sets, resulting in 90 trials. Due to the difficulty of creating good examples to be used from our real crime data, we used fewer data sets in Tasks 4 and 5. In summary, each participant performed a total of 374 trials in Tasks 1 to 5, and it generally took 90 minutes.

Since the design of Tasks 1–5 focused on questions of comparing regions, a secondary study was also conducted. This study was again a repeated measure design of tasks incorporating variations of the images shown in Fig. 1(a), (c), and (d) and line maps similar to those of Fig. 3 (right). However, here the subjects were asked to identify the values of regions in the image. Areas of homogeneous visual variables were circled in each image and the subjects were asked to approximate the amount of crime per region. As a final task, the subjects were simultaneously presented with a point map, color map, bristle map and line map and asked to rank order the images based on their aesthetic values.

For all Tasks, trial order was varied using a magic square method [40] in each task. Completion time and participants' answers were recorded for a quantitative metric. The collected data from each task was subjected to an analysis of variance (ANOVA) test to determine if the average time and accuracy of task completion were significantly different among techniques. A post-hoc Tukey HSD test was then performed to determine significance between the techniques. P-values reported in this study come from the resultant Tukey HSD test. Before the study, participants were introduced to our experiment application and the techniques through an introductory session and a training session. During the training session, participants could ask questions and receive guidance in the use of the experiment application and analysis of each visualization. Once the training was completed, participants moved to the actual study. After completing each task (Tasks 1 to 4) participants were asked to answer the questionnaire to rate the efficiency of the techniques using a five-point Likert scale [41]. After completing Task 5, participants were also asked to describe their impression with regards to visual complexity for before and after image pairs applying our clutter reduction. In the questionnaire, we stated that the visual complexity is high if a participant felt any kind of difficulty or confusion in understanding the density, length

TABLE 2: The number of data sets, techniques (cases in Fig. 8 for Task 5), and trials.

	Data sets	Techniques (or cases)	Trials
Task 1	18	5	90
Task 2	18	5	90
Task 3	15	6	90
Task 4	12	4	48
Task 5	7	8	56
Task 6	2	4	24
Task 7	2	4	2

and color of bristle lines that encode the underlying data. Finally, after finishing all tasks, participants were asked to rate the overall efficiency among techniques.

Hypotheses: In this experiment, we hypothesized that our bristle maps would be better than or equally as good when compared to the other techniques in terms of task completion time and accuracy. Specifically, we hypothesized that our bristle maps would be better than other techniques as the complexity level of tasks increased from univariate to multivariate. The rationale of this assumption is that the line map and bivariate color map use at most two variables, whereas the several encoding parameters in our bristle map have the potential to create effective encoding combinations. We also hypothesized that our clutter reduction strategies would be useful to minimize cluttering on areas where a large number of bristle lines are created. In our follow-on experiment exploring identification of values, we hypothesized that bristle maps would be as accurate as all other representations in determining values. We also hypothesized that bristle maps would be ranked higher in terms of their aesthetics.

Tasks: We tested seven tasks: three for univariate, bivariate, and multivariate data encoding, respectively, one for temporal variance encoding, one for the clutter reduction, one for accuracy comparisons among the rendering styles and one for aesthetic comparisons.

In Task 1, when given four regions highlighted in circles on the map, participants were asked to "find the region with the highest crime rate" in different visualizations representing spatio-temporal crime data using point, color, line-T (data encoded in the line (T)hickness), bristle-CLD (a redundant data encoding using (C)olor, (L)ength, and (D)ensity), and bristle-LD (a redundant data encoding using (L)ength and (D)ensity).

In Task 2, four regions were highlighted in circles on the map. Participants were asked to "find the region with the highest crime rates at both (or either) day and night time," using point (encoding day/night time crime rates in different colors), color, line-TO (data encoded as line (T)hickness and using (O)rientation for day/night crime rates), bristle-CLDO (redundant data encoding using (C)olor, (L)ength and (D)ensity, and using (O)rientation to indicate day/night crime rates), and bristle-LDO (data encoded using (L)ength and (D)ensity, but in a constant color, using (O)rientation to indicate day/night crime rates). The point map had differently colored points for day and night time crime rates, and two maps (day and night time color maps) were given in different colors for the color map.

In Task 3, four regions were highlighted in circles on the map. Participants were asked to "find the region with the highest crime rates for both (or either) two crimes (crime 1 and 2)," using point map (encoding two crimes in different colors), bivariate color map (Color-B), line-TO (a data encoding using (T)hickness in different colors, and using (O)rientation to indicate crime types), line-CT (encoding crime 1 using (C)olor and crime 2 using (T)hickness), bristle-LDO (a redundant data encoding using (L)ength and (D)ensity, and using (O)rientation to indicate crime types), and bristle-CD (an encoding using (C)olor to indicate crime 1 and (D)ensity to indicate crime 2, with constant length).

In Task 4, participants were given two regions highlighted in circles on the map. Then, they were asked to "find the region with the highest temporal variance" in different visualizations using point maps, color maps, line maps, and bristle-LDV (a redundant data encoding using (L)ength and (D)ensity, and representing (V)ariance in the variance part of a bristle line). For the point, color, and line maps, multiple images were displayed on the screen to provide visualizations during several years. Our bristle map embedded the variance in the variance part of the bristle length as shown in Fig. 2 (third stage) and 5 (right column).

In Task 5, given two regions predefined in circles on bristle maps, participants were asked to "answer if crime rates on this given two regions look either different or the same as each other." Fig. 8 shows representative image pairs before and after applying our clutter reduction method. In trials, participants compared each case in Fig. 8 to a base case (i. e., bristle lines on a single straight road).

In Task 6, subjects were presented with a series of images with a single predefined circle which covered an area consisting of homogeneous visual variables (i. e., identical color, bristle length, thickness, etc.). A univariate encoding was explored, and the Bristle-CLD settings were utilized for the bristle map. Participants were asked to estimate the amount of crime in the area using the provided scale (or scales in the case of bristle and line maps). Time and accuracy of the results were measured.

In Task 7, subjects were presented simultaneously with four images representing the same data set. These images consisted of a point map, a color map, a bristle map and a line map. Subjects were asked to rank order the images in order of most to least aesthetically pleasing.

7 RESULTS AND DISCUSSION

After all tasks were completed, times and answers collected during the study were analyzed using a single-factor ANOVA. A post-hoc Tukey HSD test was then performed to determine significance between the techniques. P-values reported in this study come from the resultant Tukey HSD test. For accuracy, the percentage of correct answers was computed.

Task 1: A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining

TABLE 3: Tukey HSD results for Task 1.

Time	p-value $<$	Point map	KDE map	Line-T
	Bristle-CLD	**.00001**	**.00001**	**.00042**
	Bristle-LD	**.00001**	**.00001**	**.01811**

Accuracy	p-value $<$	Point map	KDE map	Line-T
	Bristle-CLD	**.00001**	.1554	**.01851**
	Bristle-LD	**.00001**	.3214	**.00602**

TABLE 4: Tukey HSD results for Task 2.

Time	p-value $<$	Point map	KDE map	Line-TO
	Bristle-CLDO	**.01713**	.70091	.05943
	Bristle-LDO	**.02024**	.81621	.07166

Accuracy	p-value $<$	Point map	KDE map	Line-TO
	Bristle-CLDO	**.00001**	.07062	.36692
	Bristle-LDO	**.00001**	.99999	**.01283**

areas with highest crime rates within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps and bristle maps. There was a significant effect of visualization type on time at the $p < .05$ level for the conditions $[F(4, 145) = 35.366, p = .0000001]$ and a significant effect of visualization type on accuracy at the $p < .05$ level for the conditions $[F(4, 145) = 3266.782, p = .0000000006]$. Because statistically significant results were found, we computed a Tukey post-hoc test with results reported in Table 3. In Table 3 p-values $< .05$ indicate that groups were statistically different from one another.

The result showed that the bristle maps groups were both significantly different than the point, color and line maps in terms of speed (at the $p < .05$ level). Specifically, the bristle map groups average times were 50.7 seconds and 56.6 seconds for the CLD and LD conditions respectively, which was slightly faster than the Line-T condition at 69 seconds and much faster than the point map condition at 102.6 seconds. However, the color map group was the fastest at 34.6 seconds.

For accuracy, the bristle maps groups were both significantly different than the point map group in terms of accuracy (at the $p < .05$ level). Specifically, the bristle map groups accuracy ratings were 99.6% and 99.8% for the CLD and LD conditions respectively, which was much higher than the point map condition with accuracy of 41.4%. No accuracy differences were found when compared to the other groups. See Table 8 for more specific results.

The comparison between color maps and bristle maps showed that color maps were better than the bristle map in terms of average time, and were not significantly different in terms of accuracy. This shows that bristle maps as a redundant encoding scheme has the same potential to convey data as single parameter encoding schemes; however, traditional schemes such as color maps may allow for a quicker comparison in the univariate case.

Comparing Bristle-LD and Line-T, we saw that the length of the bristle map matches the thickness of the line map. Hence, the bristle density was useful to find answers in Task 1 in terms of completion time and accuracy. Some participants also mentioned bristle density in their qualitative feedback as *"Bristle map is especially good when density of the bristles is also used"* and *"In bristle map, length and density were more noticeable than color difference."* In this univariate encoding test, the point map showed the worst results and the color map was the best results in terms of time and accuracy as shown in Table 8.

Task 2: A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with highest crime rates at both day and nighttime within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps and bristle maps. There was a significant effect of visualization type on time at the $p < .05$ level for the conditions $[F(4, 145) = 2.717, p = .032]$ and a significant effect of visualization type on accuracy at the $p < .05$ level for the conditions $[F(4, 145) = 89.89, p = .0000002]$. Because statistically significant results were found, we computed a Tukey post-hoc test with results reported in Table 4. In Table 4 p-values $< .05$ indicate that groups were statistically different from one another.

As we hypothesized, the result showed that the bristle maps groups were both significantly different than the point maps in terms of speed (at the $p < .05$ level). Specifically, the bristle map groups average times were 86.3 seconds and 87.2 seconds for the CLDO and LDO conditions respectively, which was slightly faster than the point map condition at 106.2 seconds.

For accuracy, the bristle maps groups were both significantly different than the point map group in terms of accuracy (at the $p < .05$ level). Specifically, the bristle map groups accuracy ratings were 90.5% and 93.3% for the CLDO and LDO conditions respectively, which was much higher than the point map condition with accuracy of 63.1%. See Table 8 for more specific results.

The comparison between color maps and bristle maps showed that color maps were better than the bristle map in terms of average time, and were not significantly different in terms of accuracy. This shows that bristle maps as a redundant encoding scheme has the same potential to convey data as single parameter encoding schemes; however, traditional schemes such as color maps may allow for a quicker comparison in the univariate case.

Findings also indicated that Bristle-LDO was better than Line-TO in terms of accuracy, whereas Bristle-CLDO was not significantly different from Line-TO in terms of accuracy. This indicated that the bristle density seems to be useful in finding correct answers in Bristle-LDO, but it was not in Bristle-CLDO. Further testing in combinations of visual variables and the ability to determine levels of sparseness will be done in the future.

Task 3: A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas

TABLE 5: Tukey HSD results for Task 3.

Time	p-value $<$	Point map	KDE-B	Line-TO	Line-CT
	Bristle-LDO	**.00009**	**.00515**	.58128	.73239
	Bristle-CD	**.01131**	**.03469**	.15506	.20693

Accuracy	p-value $<$	Point map	KDE-B	Line-TO	Line-CT
	Bristle-LDO	**.00001**	**.00001**	.27189	**.02771**
	Bristle-CD	**.00001**	**.00001**	.07002	.41194

TABLE 6: Tukey HSD results for Task 4.

Time	p-value $<$	Point maps	KDE maps	Line maps
	Bristle-LDV	**.00001**	**.00001**	**.00001**

Accuracy	p-value $<$	Point maps	KDE maps	Line maps
	Bristle-LDV	**.00001**	**.00001**	**.00001**

with highest crime rates in two types of crimes within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps and bristle maps. There was a significant effect of visualization type on time at the $p < .05$ level for the conditions $[F(5, 174) = 6.655, p = .00001]$ and a significant effect of visualization type on accuracy at the $p < .05$ level for the conditions $[F(5, 175) = 144.24, p = .00000001]$. Because statistically significant results were found, we computed a Tukey post-hoc test with results reported in Table 5. In Table 5 p-values $< .05$ indicate that groups were statistically different from one another.

The result showed that the bristle maps groups were both significantly different than the point maps and color maps in terms of speed (at the $p < .05$ level). Specifically, the bristle map groups average times were 88.2 seconds and 94.5 seconds for the LDO and CD conditions respectively, which was faster than the point map condition at 118.3 seconds and the color map condition at 115.3 seconds.

For accuracy, the bristle maps groups were both significantly different than the point map group and the color map group in terms of accuracy (at the $p < .05$ level). Specifically, the bristle map groups accuracy ratings were 94.4% and 90.4% for the LDO and CD conditions respectively, which was much higher than the point map condition with accuracy of 26.6% and the color map condition with accuracy of 72.6%. See Table 8 for more specific results.

Note that we separated parameters for different crime types in Bristle-CD; (C)olor encodes crime 1 and (D)ensity encodes crime 2. Bristle-CD showed a significant effect compared to the bivariate color map as shown in Table 5. However, generation on this type of bristle maps should be selected carefully since one parameter could dominate the other. For instance, when we use color and length to separate two crime data, short bristle length for low crime rates in crime 2 removes bristle lines in dark color for high crime rates in crime 1. In our experiment, we selected color and density for two crimes, with constant length of bristles.

Task 4: A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with high temporal variance within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps and bristle maps. There was a significant effect of visualization type on time at the $p < .05$ level for the conditions $[F(3, 116) = 42.051, p = .00001]$ and a significant effect of visualization type on accuracy at the $p < .05$ level for the

conditions $[F(3, 116) = 42.33, p = .00001]$. Because statistically significant results were found, we computed a Tukey post-hoc test with results reported in Table 6. In Table 6 p-values $< .05$ indicate that groups were statistically different from one another.

The result showed that the bristle maps groups were both significantly different than the point maps, line maps and color maps in terms of speed (at the $p < .05$ level). Specifically, the bristle map groups average time was 48.4 seconds for the LDV condition, which was faster than the point map condition at 194 seconds, the color map condition at 171.8 seconds, and the line map condition at 178.9 seconds.

For accuracy, the bristle maps groups were both significantly different than the point maps, line maps and color maps in terms of speed (at the $p < .05$ level). Specifically, the bristle map groups accuracy rating was 94.7% for the LDV condition, which was much higher than the point map condition with accuracy of 53.6%, the color map condition with accuracy of 72.6% and the line map condition with accuracy of 75.5%. See Table 8 for more specific results.

As we hypothesized, we found that the representation of temporal variance in bristle maps was significantly faster and accurate in terms of both average time and accuracy compared to providing several images of the point, color and line maps. Moreover, we found that techniques showed the increasing pattern from the point maps to Bristle-LDV as shown in Table 8. This indicates that changes among several images would be better perceived in line patterns than in points or colors.

Task 5: A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with high temporal variance within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps and bristle maps. There was no significant effect of visualization type on time at the $p < .05$ level for the conditions $[F(1, 56) = .328, p = .569]$ and no significant effect of visualization type on accuracy at the $p < .05$ level for the conditions $[F(1, 56) = .315, p = .315]$. In Task 5, we found that bristle lines with and without clutter reduction did not differ significantly w.r.t. both average time and accuracy for all cases (Fig. 8). This means that the base bristle lines and bristle lines before applying clutter reduction and the base and bristle lines after applying our clutter reduction are perceived similarly by participants. Moreover, when told that the bristle line orientation does not encode data, the opposite orientations of bristle lines on a single straight
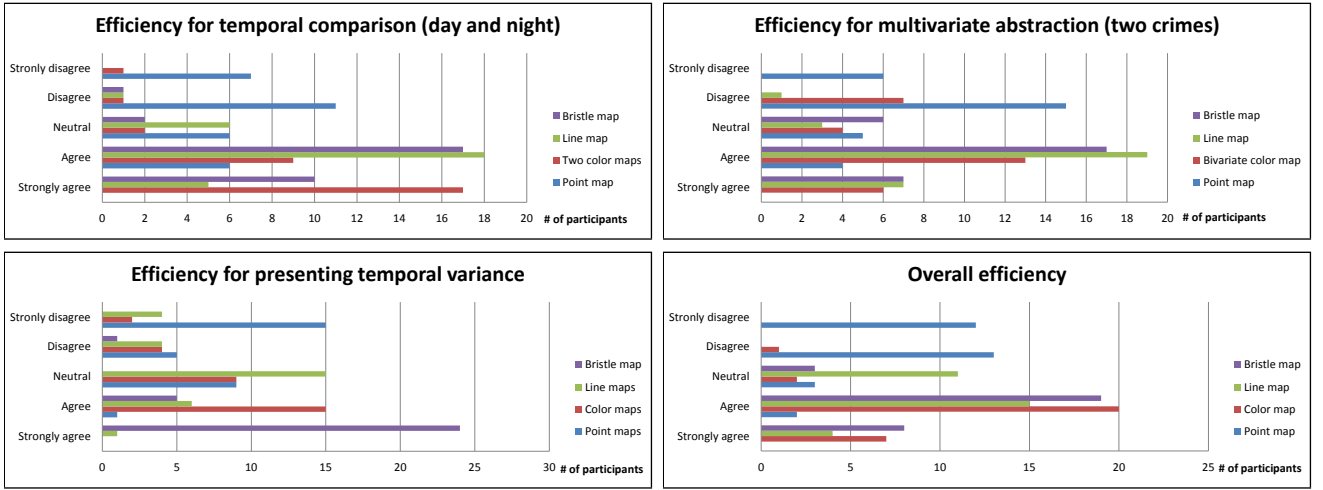
Fig. 10: Results from qualitative feedback for Tasks 2, 3, and 4 as well as overall efficiency.

TABLE 7: Average Rank Ordering by Aesthetics

	Point	KDE Map	Bristle	Line
Average	2.26	2.6	2.79	2.34
Std Dev	1.18	.97	1.19	1.10

road caused by virtual nodes (Fig. 7(b)) did not affect accuracy (87.7%). Other cases showed 42-58% of accuracy.

Task 6: For Task 6, we hypothesized that subjects would be as accurate as all other representations in determining values In Task 6, we found that bristle maps did not differ significantly w.r.t. accuracy when compared with point map, color map and line map identification (ANOVA results of p-value=.18093, F=1.63). However, we found that bristle maps did differ significantly w.r.t. time when compared with point map, color map and line map (ANOVA results of p-value=.0314, F=2.622). Particularly, we found line maps and heat maps to both be significantly faster than point maps and bristle maps in identifying values (Tukey HSD test value of $p < .05$). Overall, these results indicate that in terms of accuracy, all geographical representations were equally useful; however, participants were (on average) over 1 second quicker in value judgments on both line maps and colors maps. This is most likely due to the fact that participants were quicker at making color judgments as compared to counting points and mentally linking multiple variables for the bristle maps.

Task 7: In Task 7, we found that users had a highly variable rating of which image appeared to be more aesthetically pleasing. The average positions and standard deviations are summarized in Table 7. Here we find that while bristle maps have a slightly higher average ranking, there is no significant difference between the aesthetic ordering. A one-way between-subjects ANOVA was conducted to compare the rankings of map visualizations by subject in determining which visualization was ranked highest in aesthetics. There was no significant effect of visualization type on aesthetics at the $p < .05$ level for the conditions

$[F(3,183) = 1.79, p = .149]$.

Qualitative evaluation: Fig. 10 shows the results from qualitative feedback. Among the 30 participants, 27 participants (90%) agreed or strongly agreed that the bristle map was efficient for day and night time comparison in Task 2, 26 for two color maps and 23 for line map. 24 participants (80%) agreed or strongly agreed that the bristle map was efficient for the comparison of two crimes in Task 3, 26 for the line map and 19 for the bivariate map. 29 participants (96.6%) agreed or strongly agreed that the bristle map was efficient for temporal variance representation. In the question for overall efficiency, 27 participants (90%) agreed or strongly agreed that bristle maps and color maps were overall efficient, and 19 (63.3%) for line maps. For point maps, 25 participants (83.3%) disagreed or strongly disagreed.

Participants were also asked to answer visual complexity and preference questions regarding the before (NCR) and after (CR) image pairs applying our clutter reduction. For the circular case (Fig. 8(c)), 96.5% of participants felt that NCR has higher visual complexity and 78.5% preferred CR. For the curved case (Fig. 8(d)), 65.5% of participants answered that CR has a higher visual complexity and 64% preferred NCR. While both cases use a technically identical clutter reduction algorithm, participants reported different visual complexity and preference for them. This indicates that our clutter reduction could be improved by considering the complexity of the underlying network structure.

Summary and Limitations: As a univariate encoding, the bristle maps were significantly different (in terms of speed and accuracy) than the point, color and line maps. In the case of the point and line maps, bristle maps use resulted in a higher average correctness and speed; however, the color map for the univariate case had the fastest response and accuracy totals. This seems to indicate that the redundant encoding scheme is actually not beneficial in these cases. As such, use of bristle maps for single variable encoding is not recommended.

With regards to bivariate and multivariate encoding,

TABLE 8: Average time and accuracy.

	Technique	Average time (seconds)	Accuracy (%)
Task 1	Point map	102.6 ± 40.9	41.4 ± 4.3
	KDE map	$\mathbf{34.6 \pm 7.8}$	$\mathbf{100 \pm 0}$
	Line-T	69 ± 21.9	98.1 ± 3
	Bristle-CLD	50.7 ± 15.3	99.6 ± 1.4
	Bristle-LD	56.6 ± 17.1	99.8 ± 1
Task 2	Point map	106.2 ± 34.7	63.1 ± 6.7
	KDE map	90 ± 30.9	93.1 ± 5.15
	Line-TO	100.5 ± 28.5	87.9 ± 10.4
	Bristle-CLDO	$\mathbf{86.3 \pm 30.6}$	90.5 ± 8.5
	Bristle-LDO	87.2 ± 28	$\mathbf{93.3 \pm 5.2}$
Task 3	Point map	118.3 ± 41.1	26.6 ± 12.5
	KDE-B	115.3 ± 47.9	72.6 ± 7.9
	Line-TO	$\mathbf{84 \pm 22.8}$	$\mathbf{96.4 \pm 6.2}$
	Line-CT	86.1 ± 22	86.8 ± 16.7
	Bristle-LDO	88.2 ± 23.7	94.4 ± 7.6
	Bristle-CD	94.5 ± 28.3	90.4 ± 16.7
Task 4	Point maps	194 ± 73.5	53.6 ± 17.6
	KDE maps	171.8 ± 58.8	61.9 ± 15.6
	Line maps	178.9 ± 61.8	75.5 ± 13.5
	Bristle-LDV	$\mathbf{48.4 \pm 14.8}$	$\mathbf{94.7 \pm 13.4}$

bristle maps and line maps outperformed color and point maps. This is not surprising as bristle and line maps are able to combine variables into a single image, where as in the case of point and color maps, the user must mentally combine the two images together. Bristle-(C)LD also showed a significant effect of the bristle density compared to Line-T. As a bivariate encoding, using orientation in bristle maps was not significant compared to two color maps. However, in the comparison with the bivariate color map, Bristle-LDO showed a significant effect in terms of average time and accuracy. As such, we have that Bristle-(C)LDO as a bivariate encoding scheme created a middle level of cognitive load in-between two color maps and a bivariate color map. Bristle maps also showed potential as a multivariate encoding technique in a single view. Based on the results in Task 3, a point map using various colors and a multivariate color map would considerably increase users' cognitive load. In Tasks 1-3, we also observed that there is no significant effect between the bristle maps using the different encodings. The representation of temporal variance in the bristle map was significantly different from other methods. Our results also showed the differences among point, color and line maps. Participants could better find the region with higher temporal variance when using line maps than using point and color maps. In the qualitative evaluation, 90% of the participants agreed or strongly agreed the overall efficiency of bristle maps to find answers. However, users also strongly preferred the color map in these cases as well.

Finally, we found that with regards to accuracy in identifying values, no technique outperformed any others. However, users were significantly faster in identifying values in both the color and line map scenarios. We hypothesize that in both cases the user focused only on the color, where as in the point map case they needed to count the points and in the bristle map case they needed to reconfirm the univariate

value by double checking several of the encoding legends.

Overall, this technique would be recommended when encoding large amounts of multivariate spatio-temporal point data. As the number of point samples increase, aggregation techniques are need to allow for quick summaries of the data, and, as is evidenced by our studies, pure spatial location representation by glyphs results in too much overlap for accurate measurement and evaluation. As the number of variables increase, color map representations allow for the encoding of variables only along a single visual variable (resulting in bivariate color maps or small multiple plots).

In using multivariate encodings, it is extremely important to understand the interaction effects that the visual variables will introduce in one another. Research into the perceptual interactions among different visual variables was performed by Acevedo and Laidlaw [42]. They measured the perceptual interfernce of icon size, spacing and brightness, noting that brightness outperforms spacing and size while being subject to interferences from both spacing and size. Acevedo and Laidlaw also noted that spacing also outperformed size, which contradicted some previous results; however, this result seems to align with our participants noting that the bristle spacing was a useful cue. Their results were reportedly due to the spacing sampling along a sinusoidal curve. The sampling of our bristles follow a uniform pattern within classification bins. Thus, there seems to be sufficient scientific evidence to justify using sparsity as a discriminating variable in the case of the bristle maps; however, further studies on this are warranted. Stone [43] has also studies the effect of size in color perception, noting that color appearance changed dramatically with the size being viewed. As such, it may be better to utilize fewer map classifications (color bins) when using bristle maps in order to increase the perceptual distance between each color being visualized.

The main limitations of the bristle map technique is that the combinations of data encoding can potentially prove overwhelming for the designer, and a poor choice on variable encoding can result in a suboptimal visualization. In particular, previous studies have provided results that can be used to predict that certain combinations of visual variables will either enhance or impede map reading. For example, the combination of length and density form an emergent property akin to Bertin's definition of grain. Such effects cannot be ignored; however, bristle maps can be encoded to take advantage of such combinations, as shown in Tasks 3 and 4.

Finally, with regards to scalability of the bristle map technique, in areas of dense roadways, different aggregation methods would need to be considered. As the roads become dense, the ability to plot lines of perceptually different length would become untenable. However, a solution to this would be to draw only the most important roads, thereby removing smaller roads from the analysis, or utilizing bristle maps in a focus+context manner.

To summarize, we posited several different hypotheses. First, we hypothesized that our bristle maps would be

better than other techniques as the complexity level of tasks increased from univariate to multivariate. As the number of variables under analysis increased, the bristle maps outperformed more traditional analyses in terms of speed. However, at lower levels of complexity, traditional techniques such as density estimated heatmaps were found to be best in terms of speed. Second, we hypothesized that our bristle maps would be as accurate as all other representations in determining values. This hypothesis was verified through our user study where subjects estimated the magnitude of crimes on a map using points, density estimated heat maps and bristle maps. Results showed that there were no significant differences between map styles. Finally, we also hypothesized that bristle maps would be ranked higher in terms of their aesthetics. This hypothesis was refuted as our results did not demonstrate any significantly different rating for any of the different mapping techniques.

8   CONCLUSIONS

In this work, we have described our novel multivariate data encoding scheme, the Bristle Map. This scheme provides a novel approach for encoding color, length, density, and orientation as data variables and allowing the user to explore correlations within and between variables on a single view. Given the number of parameters available within this encoding, this article has presented only a subset of potential encodings and examples. Here, we have shown the use of encoding bristle lines with redundant information, multivariate attributes for variable comparison, and temporal variance. We also showed a means of potentially encoding data uncertainty. To minimize overlap of bristle lines, we generated a topology graph from underlying geographical line features and employed strategies for clutter reduction. Then, to evaluate the effectiveness of bristle maps, we performed an evaluation study, where we explored different visual encoding combinations within the bristle maps and compared with existing techniques in several tasks. Based on our experiment results, we believe that our bristle map technique has much potential to increase the amount of information that can be visualized on a single map for geovisualization.

ACKNOWLEDGMENTS

REFERENCES

[1] A. MacEachren, D. Xiping, F. Hardisty, D. Guo, and G. Lengerich, "Exploring High-D Spaces with Multiform Matrices and Small Multiples," in *Proc. InfoVis*. Los Alamitos: IEEE Computer Society, 2003, pp. 31–38. doi> 10.1109/INFVIS.2003.1249006

[2] C. North and B. Shneiderman, "Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata," in *Proc. AVI*. New York: ACM, 2000, pp. 128–135. doi> 10.1145/345513.345282

[3] C. Weaver, "Cross-Filtered Views for Multidimensional Visual Analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 192–204, Mar./Apr. 2010. doi> 10.1109/TVCG.2009.94

[4] D. S. Ebert, R. M. Rohrer, C. D. Shaw, P. Panda, J. M. Kukla, and D. A. Roberts, "Procedural Shape Generation for Multi-Dimensional Data Visualization," *Computers & Graphics*, vol. 24, no. 3, pp. 375–384, Jun. 2000. doi> 10.1016/S0097-8493(00)00033-9

[5] S. Bachthaler and D. Weiskopf, "Continuous Scatterplots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1428–1435, Nov./Dec. 2008. doi> 10.1109/TVCG.2008.119

[6] R. Maciejewski, S. Rudolph, R. Hafen, A. M. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert, "A Visual Analytics Approach to Understanding Spatiotemporal Hotspots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 205–220, Mar./Apr. 2010. doi> 10.1109/TVCG.2009.100

[7] J. J. van Wijk and A. Telea, "Enridged Contour Maps," in *Proc. VIS*. Los Alamitos: IEEE Computer Society, 2001, pp. 69–74. doi> 10.1109/VISUAL.2001.964495

[8] R. J. Phillips and L. Noyes, "An Investigation of Visual Clutter in the Topographic Base of a Geological Map," *Cartographic Journal*, vol. 19, no. 2, pp. 122–132, Dec. 1982. doi> 10.1179/000870482787073225

[9] S. Openshaw, "The Modifiable Areal Unit Problem," in *Concepts and Techniques in Modern Geography*. Norwich, UK: Geo Books, 1984, vol. 38.

[10] M. Swink and C. Speier, "Presenting Geographic Information: Effects of Data Aggregation, Dispersion, and Users' Spatial Orientation," *Decision Sciences*, vol. 30, no. 1, pp. 169–195, Jan. 1999. doi> 10.1111/j.1540-5915.1999.tb01605.x

[11] C. L. Eicher and C. A. Brewer, "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation," *Cartography and Geographic Information Science*, vol. 28, no. 2, pp. 125–138, Apr. 2001. doi> 10.1559/152304001782173727

[12] R. Spence, *Information Visualization*. Reading, MA, USA: Addison-Wesley, 2001.

[13] L. Wilkinson, *The Grammar of Graphics*, 2nd ed. Heidelberg/Berlin: Springer-Verlag, 2005.

[14] A. M. MacEachren, *How Maps Work: Representation, Visualization, and Design*. Guilford Press, 1995.

[15] S. Chainey, L. Tompson, and S. Uhlig, "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime," *Security Journal*, vol. 21, no. 1–2, pp. 4–28, Feb.–Apr. 2008. doi> 10.1057/palgrave.sj.8350066

[16] B. W. Silverman, "Density Estimation for Statistics and Data Analysis," in *Monographs on Statistics and Applied Probability*. New York: Chapman & Hall, 1986, no. 26.

[17] C. Ahlberg and B. Shneiderman, "Visual Information Seeking using the FilmFinder," in *Proc. CHI*. New York: ACM, 1994, pp. 433–434. doi> 10.1145/ 259963.260431

[18] Y.-H. Fua, M. O. Ward, and E. A. Rundensteiner, "Structure-Based Brushes: A Mechanism for Navigating Hierarchically Organized Data and Information Spaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, no. 2, pp. 150–159, Apr.–Jun. 2000. doi> 10.1109/2945.856996

[19] A. Dix and G. Ellis, "*By Chance*: Enhancing Interaction with Large Data Sets Through Statistical Sampling," in *Proc. AVI*. New York: ACM, 2002, pp. 167–176. doi> 10.1145/1556262.1556289

[20] A. MacEachren, "Visualizing Uncertain Information," *Cartographic Perspectives*, no. 13, pp. 10–19, Fall 1992.

[21] R. Dunn, "A Dynamic Approach to Two-Variable Color Mapping," *The American Statistician*, vol. 43, no. 4, pp. 245–252, Nov. 1989. doi> 10.1080/ 00031305.1989.10475669

[22] J. Olson, "Spectrally Encoded Two-Variable Maps," *Annals of the Association of American Geographers*, vol. 71, no. 2, pp. 259–276, Jun. 1981. doi> 10.1111/ j.1467-8306.1981.tb01352.x

[23] A. MacEachren and D. DiBiase, "Animated Maps of Aggregate Data: Conceptual and Practical Problems," *Cartography and Geographic Information Systems*, vol. 18, no. 4, pp. 221–229, Oct. 1991. doi> 10.1559/ 152304091783786790

[24] H. Hagh-Shenas, S. Kim, V. Interrante, and C. Healey, "Weaving Versus Blending: A Quantitative Assessment of the Information Carrying Capacities of two Alternative Methods for Conveying Multivariate Data with Color," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1270–1277, Nov./Dec. 2007. doi> 10.1109/TVCG.2007.70623

[25] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda, "Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data," in *Proc. InfoVis*. Los Alamitos: IEEE Computer Society, 2005, pp. 173–180. doi> 10.1109/ INFOVIS.2005.35

[26] M. Sips, J. Schneidewind, D. A. Keim, and H. Schumann, "Scalable Pixel-Based Visual Interfaces: Challenges and Solutions," in *Proc. IV*. Los Alamitos: IEEE Computer Society, 2006, pp. 32–38. doi> 10. 1109/IV.2006.95

[27] C. Panse, M. Sips, D. Keim, and S. North, "Visualization of Geo-spatial Point Sets via Global Shape Transformation and Local Pixel Placement," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 749–756, Sep./Oct. 2006. doi> 10.1109/TVCG.2006.198

[28] D. Dorling, A. Barford, and M. Newman, "Worldmapper: The World as You've Never Seen it Before," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 757–764, Sep./Oct. 2006. doi> 10.1109/TVCG.2006.202

[29] P. C. Wong, K. Schneider, P. Mackey, H. Foote, G. Chin, R. Guttromson, and J. Thomas, "A Novel Visualization Technique for Electric Power Grid Analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 3, pp. 410 –423, May/Jun. 2009. doi> 10.1109/TVCG.2008.197

[30] D. Fisher, "Hotmap: Looking at Geographic Attention," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1184–1191, Nov./Dec. 2007. doi> 10.1109/TVCG.2007.70561

[31] C. Tominski, P. Schulze-Wollgast, and H. Schumann, "3D Information Visualization for Time Dependent Data on Maps," in *Proc. InfoVis*. Los Alamitos: IEEE Computer Society, 2005, pp. 175–181. doi> 10.1109/ IV.2005.3

[32] J. Bertin, *Semiology of Graphics*. Redlands, California: ESRI Press, 2011.

[33] J. Tarbell, "Substrate," Web site & simulation: http://www.complexification.net/gallery/machines/ substrate/, 2003, accessed February 2012.

[34] T. Isenberg, "Visual Abstraction and Stylisation of Maps," *The Cartographic Journal*, vol. 50, no. 1, pp. 8–18, Feb. 2013. doi> 10.1179/1743277412Y. 0000000007

[35] P. Rheingans, "Task-Based Color Scale Design," in *Proc. SPIE*, vol. 3905. SPIE, 2000, pp. 35–43. doi> 10.1117/12.384882

[36] C. Ware, "Color Sequences for Univariate Maps: Theory, Experiments and Principles," *IEEE Computer Graphics and Applications*, vol. 19, no. 5, pp. 41–49, Sep./Oct. 1988. doi> 10.1109/38.7760

[37] C. A. Brewer, *Designing Better Maps: A Guide for GIS Users*. Redlands, CA, USA: ESRI Press, 2005.

[38] M. Prasad, "Intersection of Line Segments," in *Graphics Gems II*, J. Arvo, Ed. Boston: Academic Press, 1991, pp. 7–9.

[39] A. Michotte, G. Thinès, and G. Crabbé, *Les Complements Amodeux des Structures Perceptives (Amodal Completions of Perceptual Structures)*. Louvain: Institut de Psychologie del'Université de Louvain, France: Studia Psychologica, 1964.

[40] M. S. Farrar, *Magic Squares*. Charleston, SC, USA: BookSurge Publishing, 1996.

[41] R. A. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 5–55, 1932.

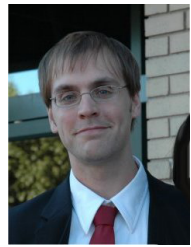[42] D. Acevedo and D. Laidlaw, "Subjective Quantifi-

cation of Perceptual Interactions among some 2D Scientific Visualization Methods," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 1133–1140, Sep. 2006. doi> 10.1109/TVCG.2006.180

[43] M. Stone, "In Color Perception, Size Matters," *IEEE Computer Graphics & Applications*, vol. 32, no. 2, pp. 8–13, Mar./Apr. 2012. doi> 10.1109/MCG.2012.37

David S. Ebert is a professor in the School of Electrical and Computer Engineering at Purdue University, a University Faculty Scholar, director of the Purdue University Rendering and Perceptualization Lab, and director of the Purdue University Regional Visualization and Analytics Center. His research interests include novel visualization techniques, visual analytics, volume rendering, information visualization, perceptually based visualization, illustrative visualization, and procedural abstraction of complex, massive data. Ebert has a PhD in computer science from Ohio State University and is a fellow of the IEEE and member of the IEEE Computer Society's Publications Board.

SungYe Kim received her Ph.D. in Electrical and Computer Engineering from Purdue University in May, 2012. She also received her masters degree in Computer Science and Engineering from Chung-Ang University, South Korea in 2000. She is currently a graphics software engineer at Intel Corporation. Prior to this, she was employed as a research engineer at the Electronics and Telecommunications Research Institute from 2000 to 2006. Her research interests are computer graphics, illustrative visualization, visual analytics and information visualization.

Tobias Isenberg is a senior research scientist with INRIA in France. He received is doctoral degree from the University of Magdeburg, Germany. Previously, he held positions as assistant professor for computer graphics and interactive systems at the University of Groningen, the Netherlands, and as postdoctoral fellow at the University of Calgary, Canada. He works on topics in interactive non-photorealistic and illustrative rendering as well as computational aesthetics and explores applications in scientific visualization. He is a member of the IEEE.

Ross Maciejewski received his Ph.D. in Electrical and Computer Engineering from Purdue University in December, 2009. He is currently an Assistant Professor at Arizona State University in the School of Computing, Informatics & Decision Systems Engineering. Prior to this, he served as a Visiting Assistant Professor at Purdue University and worked at the Department of Homeland Security Center of Excellence for Command Control and Interoperability in the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) group. His research interests are geovisualization, visual analytics and non-photorealistic rendering.

Abish Malik is a Ph.D. student in the School of Electrical and Computer Engineering at Purdue University and a research assistant at the Purdue University Rendering and Perception Lab. He received his B.S. degree in Electrical Engineering from Purdue University in 2009. His research interests include visual analytics, correlation and predictive data analytics.

Yun Jang is an assistant professor of computer engineering at Sejong University, Seoul, South Korea. He received the masters and doctoral degree in electrical and computer engineering from Purdue University in 2002 and 2007, respectively, and received the bachelors degree in electrical engineering from Seoul National University, South Korea in 2000. He was a postdoctoral researcher at CSCS and ETH Zürich, Switzerland from 2007-2011. His research interests include interactive visualization, volume rendering, visual analytics, and data representations with functions. He is a member of the IEEE.

Florida International University

A multimedia information fusion framework for web image categorization

**Wenting Lu · Lei Li · Jingxuan Li · Tao Li ·
Honggang Zhang · Jun Guo**

Abstract With the rapid development of technologies for fast Internet access and the popularization of digital cameras, an enormous number of digital images are posted and shared online everyday. Web images are usually organized by topic and are often assigned appropriate topic-related textual descriptions. Given a large set of images along with the corresponding texts, a challenging problem is how to utilize the available information to efficiently and effectively perform image retrieval tasks, such as image classification and image clustering. Previous approaches on image categorization focus on either adopting text or image features, or simply combining these two types of information together. In this paper, we improve our previously reported two multi-view classification approaches—(**Dynamic Weighting** and **Region-based Semantic Concept Integration**) for categorizing the images under the "supervision" of topic-related textual descriptions—by proposing a novel **multimedia information**

W. Lu · H. Zhang · J. Guo
Pattern Recognition and Intelligence System Lab, Beijing University of Posts
and Telecommunications, Beijing 100876, People's Republic of China

H. Zhang
e-mail: zhhg@bupt.edu.cn

J. Guo
e-mail: guojun@bupt.edu.cn

W. Lu · L. Li · J. Li · T. Li (✉)
School of Computing and Information Sciences, Florida International University,
Miami, FL 33199, USA
e-mail: taoli@cs.fiu.edu

W. Lu
e-mail: wlu@cs.fiu.edu

L. Li
e-mail: lli003@cs.fiu.edu

J. Li
e-mail: jli003@cs.fiu.edu

✪ Springer

fusion framework, in which these two proposed methods are seamlessly integrated by analyzing the special characteristics of different images. Notice that, the proposed framework is a generic multimedia information fusion framework which is not limited to our previously reported two approaches, and it can also be used to integrate other existing multi-view classification methods or models. Also, our proposed framework is capable of handling the large scale image categorization. Specifically, the proposed framework can automatically choose an appropriate classification model for each testing image according to its special characteristics and consequently achieve better classification performance with relatively less computation time for large scale datasets; Moreover, it is able to categorize images without any textual description in real world applications. Empirical experiments on two different types of web image datasets demonstrate the efficacy and efficiency of our proposed classification framework.

1 Introduction

Multimedia information plays an increasingly important role in human's daily activities. With the rapid development of technologies of fast Internet access and the popularization of digital cameras, an enormous number of digital images are posted and shared online everyday. Besides great convenience, how to efficiently and effectively retrieve images that satisfy the needs of web users in large multimedia databases is becoming more and more challenging. Particularly, web image categorization, as a crucial step of image retrieval, attracts much more attention and is very useful in the subsequent procedures, such as indexing and organizing web image databases, browsing and searching web images, and discovering interesting patterns from images [32].

Unlike the traditional scene object databases, which are mainly focusing on visual categorization, web images are usually organized by topic, as shown in Fig. 1. In general, from a classification perspective, the differences between these two kinds of images are as follows:

- Images in the same category for visual scene categorization are visually similar, but object scaling, rotating, occluding and submerging often happen in clutter background; Comparatively, web images in the same category may be different visually but very similar in terms of semantic concepts;
- Images in the same visual category contain the same object or scene and it would occupy most area of the whole image; On the contrary, some web images focus on reflecting the whole event whereas others reflect only one aspect of the whole event. For example, as shown in Fig. 1a, the first image shows the macro overview of the whole event that Mississippi suffered extensive damage from hurricane Katrina, whereas the latter two images reflect two different aspects of the event respectively, one is an American flag flying in front of a damaged home and the other is a line of rescue workers and working dogs.
- The formats of visual scene images in the same database are usually the same, even with the same size, while a web user has the freedom of uploading images

Fig. 1 An illustration of some challenging cases of 4 different image classes for web image categorization. Each row represents an image class, i.e., a topic-related event, and each image in the row either shows the macro overview of the whole event or one single aspect of the event

(a) Hurricane_building_collapse

(b) Hurricane_flood

(c) Oil_spill_seagrass

(d) Oil_spill_Animal_death

of any size and in any format, such as JPEG, BMP, TIFF, GIF, PNG, etc, which makes it more difficult to classify web images.

Therefore, it is not always reasonable to rely on the information of web images and the image techniques suitable for the traditional scene objects to perform web image classification. Fortunately, the textual information is often provided by web users to describe the general content of images, e.g., image titles, headers, or textual descriptions assigned to them. Some researchers utilized textual descriptions of images to categorize web images via simply matching keywords, regarded as the preliminary text-based web image categorization method. However, with the dramatically increasing amount of web images, the textual information of images might be incomplete or even missing, which might result in the ineffectiveness of text-based categorization methods. Moreover, on one hand, different people have different understandings on the semantics of the identical image, which might result in the *weak text similarity* of the images within the same image category and consequently poor performance of web image categorization; On the other hand, several images that describe the same event (each of which reflects only one aspect of the whole event) share the same textual description, which might result in the *strong text similarity* of the images within the same image category and consequently good performance. Simply speaking, in terms of the textual information, the special characteristics of web images can be categorized into three different types: (1) web images without corresponding textual information, (2) images with their individual textual descriptions and (3) images in the same category sharing the same description.

Therefore, in our work, in order to effectively perform web image categorization, we aim to take such special characteristics of different images into consideration and design a multimedia information fusion framework to automatically choose a suitable classification model for each image according to its special characteristics.

In order to verify the feasibility of the above solution, we firstly investigated five different kinds of methods (*Text-based Classification*, *Image-based Classification*, *Feature Integration*, *Semantic Integration* and *Hybrid Integration*) for web image categorization, and provided a comparative experimental study on integrating text and image information to perform the image categorization task in our previous work. Furthermore, we explored the feasibility of using the textual information as a "guidance" for the image categorization by proposing two novel methods (**Dynamic Weighting** and **Region-based Semantic Concept Integration**), which could achieve better performance comparing with the above five existing approaches. The preliminary study of the work has been published as a 6-page paper at the 24th International FLAIRS Conference [21], in which we have verified the efficacy of the above two proposed methods on a small manually collected dataset. However, these two methods are not effective enough, especially when the dataset becomes very large (as shown in Figs. 9 and 10, please refer to its corresponding analysis). Therefore, in this paper, we try to improve our two classification methods by proposing a novel **multimedia information fusion framework** in which these two proposed methods are seamlessly integrated by analyzing the special characteristics of different images. Specifically, the proposed framework can not only choose a suitable classification model for each testing image via a hashing mechanism to determine the quality of its textual annotation and consequently achieve better performance with relatively less computation time for large scale datasets, but also address the problem that the textual descriptions of a small portion of web images are missing. An overview of our proposed framework is depicted as in Fig. 2.
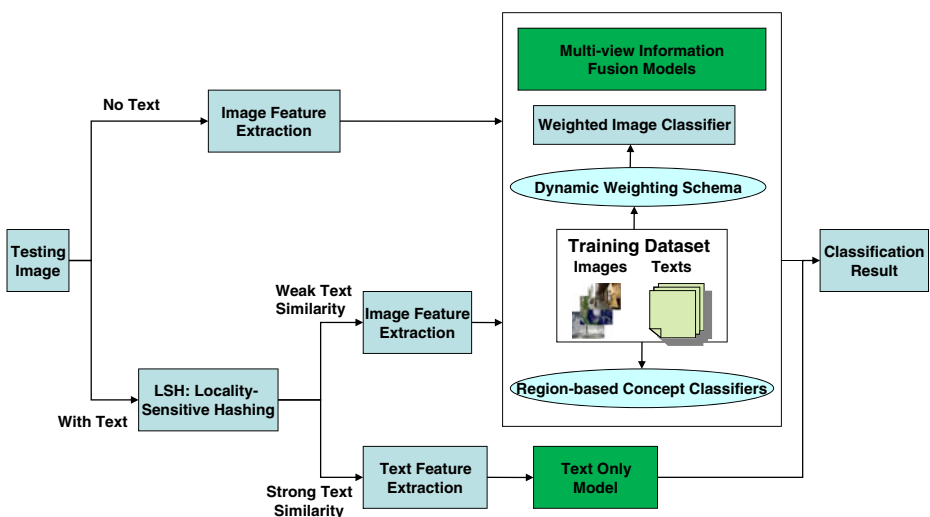
Fig. 2 The multimedia information fusion framework for web image categorization

As shown in Fig. 2, in the training stage, we first train two different multi-view information fusion models based on the weighted image features (*Dynamic Weighting*), or the correlation of semantic concepts between image regions and text concepts (*Region-based Semantic Concept Integration*). The algorithmic details of the training procedure will be described in Sections 3.1 and 3.2. Note that in our proposed framework, although every training image has its corresponding textual description, the framework also can be applied to predict web images without textual descriptions. In the testing stage, when a testing image comes, we first check whether it has a corresponding textual description or not. If not, the framework will extract its image features and choose one of the *Multi-view Information Fusion Models* as its classification model; otherwise, we firstly utilize Locality-Sensitive Hashing (LSH) [10] to determine whether its similarity is *weak* or *strong*. If the text similarity is *strong* (i.e., *the textual information of the testing image is discriminative enough so that the testing image can be classified accurately*), we just extract its text features and choose the *Text Only Model* as its classification model; if the text similarity is relatively *weak* such that only using single textual information can not perform the classification task well, we need to combine the textual information with the corresponding image information together, and consequently choose one of the *Multi-view Information Fusion Models* as its classification model to achieve better classification performance. Which multi-view information fusion model will be chosen could be indicated by web users or randomly chosen by the framework itself. Notice that, such a framework is a generic multimedia information fusion framework which is not limited to our previously reported two approaches, and it can also be used to integrate other existing multi-view classification methods or models. Empirical experiments on two different types of web image datasets demonstrate that our proposed framework is more suitable for the large scale web image categorization in terms of accuracy and efficiency.

The rest of this paper is organized as follows. In Section 2 we review some related works that combine image and text features to perform the image classification task. In Section 3 we give algorithmic details of the proposed multimedia information fusion framework to automatically choose an appropriate classification model for each image and effectively integrate textual information with image information for web image categorization. Section 4 presents a detailed experimental comparison among different categorization approaches and our proposed framework on two different types of web image datasets. Finally we conclude the paper in Section 5.

2 Related work

In real-world applications, various formats of web data are available to us, e.g., text, image, video, audio, etc. Most of the traditional web image categorization approaches often focus on utilizing individual information to categorize images. Due to the simplicity of feature extraction and computation, these techniques are still adopted by the majority of web image applications. However, some problems along with image representations arise in different scenarios. For text-based image categorization, it categorizes web images not by images content themselves, but by texts related to them, so (1) web images cannot be appropriately classified if there is no textual information assigned to them; (2) the manual text labeling is

too subjective due to human assignments, which might result in bias or noise to web image categorization; and (3) using a few words to describe the content of an image is not enough since the limited textual description can only provide a relatively sparse feature space. For image feature based categorization, (1) the extracted image features tend to be substantial, making categorization suffering in high dimensional space; and (2) low-level visual features cannot represent semantic meanings, and hence the distinguishable power of image features is relatively poor when performing the categorization task. Therefore, the performance of traditional web image categorization approaches is very limited.

Web images are naturally multi-modal, in a sense that they are represented by multiple sets of features. Therefore, in order to solve the above problems, many research publications [3, 8, 15, 31, 33] aim to design multi-view learning algorithms to learn classifiers from multiple information sources via integrating different types of features together to perform classification. For example, in [3], Blei et al. described three hierarchical probabilistic mixture models which aim to describe data with multiple types where the instance of one type (such as a caption) serves as a description of the other type (such as an image); in [8], given an initial large set of neural networks, Giacinto et al. proposed an approach which aimed to select the subset formed by the most error-independent nets; In [15], Kalva et al. presented a novel method for the classification of images that combines information extracted from the images and contextual information at classifier level via heuristic rules. Besides classification, multi-view learning algorithms are also exploited and applied in the field of multimedia retrieval and annotation [11, 19, 23, 29]. Specifically, in [29], Wang et al. presented a probabilistic approach to refine image annotations by incorporating semantic relations between annotation words. In [19], Li et al. explored how to leverage the web image collections to fulfill image retrieval and develop a novel multimedia application system named Word2Image. In [23], the authors investigated how to automatically reassign the manually annotated labels at the image-level to those contextually derived semantic regions. In [11], Hare et al. provided a solid, repeatable methodology and protocol for performing evaluations of automatic annotation software using the MIR Flickr dataset together with freely available tools for measuring performance in a controlled manner.

In general, from the fusion perspective, most of the above multi-view learning methods can be categorized into three different groups based on the way that information from different sources is used:

– **Feature integration**
  This approach enlarges the feature representation to incorporate all attributes from different sources and produces a unified feature space. In particular, continuous data types will be converted into discrete levels and categorical data type will be mapped into similar discrete levels. The data are then transformed into the same features space and standard computational methods, such as prediction and clustering, can be performed. The advantage of feature integration is that the unified feature representation is often more informative and allows many different data mining methods to be applied and systematically compared. One disadvantage is the increased learning complexity and difficulty as the data dimension becomes large [30].
– **Semantic integration**
  This approach keeps feature intact in their original form. Computational meth-

ods are applied to each feature set separately. Results on different feature sets are then combined by either voting [4], Bayesian averaging [2], or the hierarchical expert system approach [14]. One advantage of semantic integration is that it can implicitly learn the correlation structure between different sets of features [15, 18]. And the methods are more stable when the size of the training dataset is small [17].

– **Hybrid integration**

This approach can be viewed as a compromise between the feature integration and the semantic integration. The feature sets are kept in their original form and they are integrated at the similarity computation or the Kernel level [16, 27]. Take the integration between two information sources for example, for data objects p_i and p_j, their overall pairwise similarity or affinity is $S_{ij} = A_{ij} + B_{ij}$, where A_{ij} is computed from one information source and B_{ij} is obtained from another information source. Different weights can be used for different data sources. Standard computational methods can then be applied once the overall similarity is computed. This methods are preferred when a dataset has a varying local data distribution [17].

Different from the above three groups of methods which firstly process each information source separately and then combine them together at either feature level, semantic level, or kernel level, we propose a novel multimedia information fusion framework where we first analyze the special characteristics of different web images, and then automatically choose an appropriate classification model for each image according to its special characteristics and consequently achieve better classification performance. Notice that, both *Multi-view Information Fusion Models* integrated in the proposed framework involve the interaction between different information sources. In addition, the proposed framework is a generic multimedia information fusion framework. Besides these two multi-view classification approaches we reported in [21], the above three groups of methods can also be integrated in the framework. The algorithmic details of its versatility will be introduced in Section 3.3.2.

Our contribution   Due to the scalability limitations of our previous reported two multi-view learning methods (Dynamic Weighting and Region-based Semantic Concept Integration), we propose a novel and generic **multimedia information fusion framework** to handle the large scale web image categorization. Specifically, (1) the proposed framework can choose a suitable classification model for each testing image via a hashing mechanism to determine how good its textual annotation is and consequently achieve better performance with relatively less computation time for large scale datasets; (2) the framework can address the problem that the textual descriptions of a small portion of web images are missing; (3) we present an empirical investigation on the time consumed for different categorization methods integrated in the framework, which could further demonstrate the efficiency of our proposed classification framework; and (4) we evaluate the versatility of our proposed classification framework in our experiments. In addition, we provide a systematic experimental study on different methods for combining text information with image features to perform image categorization tasks, and compare their classification performance.

3 Algorithm for web image categorization

In this section, we introduce the algorithmic details of our proposed multimedia information fusion framework. The proposed framework could choose a suitable classification model for each testing image according to its special characteristics and consequently achieve better classification performance with relatively less computation time. Note that in our proposed framework, every image for training has its corresponding textual description, whereas the textual descriptions of a small portion of testing images are missing. In the following, we first introduce the two previously proposed approaches, and then describe the classification framework in detail.

3.1 Dynamic weighting

Image feature extraction techniques tend to extract a huge number of image features based on different criteria. Among these features, some of them might carry significant semantic information about the image, whereas some others might be less crucial for the tasks being executed on the image. Particularly in image classification, the extracted features should be more representative and carry substantial amount of semantic meanings. Therefore, it might be helpful to dynamically assign different weights to different image features so that the features with more importance can be captured and play more meaningful roles on the classification. Some previous works [28] on music information retrieval demonstrate how to learn appropriate similarity

Fig. 3 Framework of dynamic weighting
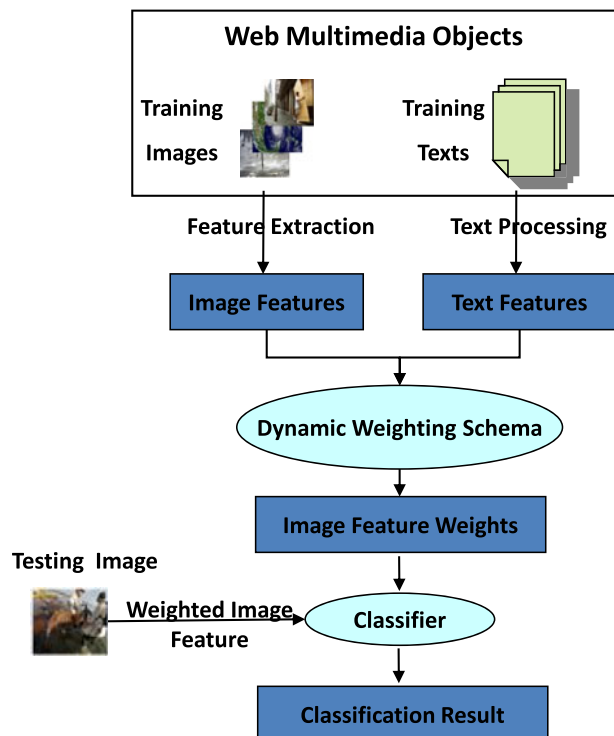
metrics based on the correlation between acoustic features and user access patterns. Motivated by this, we incorporate the concept of dynamic feature weighting into our image classification problem. Figure 3 presents the framework of dynamic weighting.

Specifically in image classification, given that human perception of an image is well approximated by its textual description, a good weighting schema for the extracted image features guided by textual information may lead to a good similarity measurement, and therefore better classification results. Let $\mathbf{m}_i = (\mathbf{f}_i, \mathbf{t}_i)$ denote the i-th image in the image collection, where $\mathbf{f}_i$ and $\mathbf{t}_i$ represent its image features and text features respectively. Let $S_f(\mathbf{f}_i, \mathbf{f}_j; \mathbf{w}) = \sum_l f_{i,l} f_{j,l} w_l$ be the image-based similarity measurement between the i-th and the j-th images when the parameterized weights are given by $\mathbf{w}$, where $f_{i,l}$ is the l-th feature in the image feature set f_i and $f_{j,l}$ is the l-th feature in the image feature set f_j. Let $S_t(\mathbf{t}_i, \mathbf{t}_j) = \sum_k t_{i,k} t_{j,k}$ be the similarity measurement between the i-th and the j-th images based on their textual description features, in general, the words with specific meanings extracted from texts. Here for each k, $t_{i,k}$ denotes whether the k-th word appears in the textual description of the i-th image. To learn appropriate weights $\mathbf{w}$ for image features, we can enforce the consistency between similarity measurements $S_f(\mathbf{f}_i, \mathbf{f}_j; \mathbf{w})$ and $S_t(\mathbf{t}_i, \mathbf{t}_j)$. The above idea leads to the following optimization problem:

$$\mathbf{w}^* = \operatorname{argmin} \sum_{i \neq j} (S_f(\mathbf{f}_i, \mathbf{f}_j; \mathbf{w}) - S_t(\mathbf{t}_i, \mathbf{t}_j))^2 \quad s.t. \mathbf{w} \geq 0. \tag{1}$$

Let p be the number of image features. The summation in (1) can be rewritten as follows:

$$\sum_{i \neq j} (S_f(\mathbf{f}_i, \mathbf{f}_j; \mathbf{w}) - S_t(\mathbf{t}_i, \mathbf{t}_j))^2 = \sum_{i \neq j} \left(f_{i,1} f_{j,1} w_1 + \cdots + f_{i,p} f_{j,p} w_p - \sum_k t_{i,k} t_{j,k} \right)^2$$

$$= \sum_{i \neq j} \left((f_{i,1} f_{j,1} w_1 + \cdots + f_{i,p} f_{j,p} w_p)^2 \right.$$

$$- 2(f_{i,1} f_{j,1} w_1 + \cdots + f_{i,p} f_{j,p} w_p)$$

$$\left. \times \left(\sum_k t_{i,k} t_{j,k} \right) + \left(\sum_k t_{i,k} t_{j,k} \right)^2 \right),$$

Let n be the number of images, and let

$$F = \begin{bmatrix} f_{1,1} f_{2,1} & f_{1,2} f_{2,2} & \cdots & f_{1,g} f_{2,g} \\ \cdots & \cdots & \cdots & \cdots \\ f_{n-1,1} f_{n,1} & f_{n-1,2} f_{n,2} & \cdots & f_{n-1,g} f_{n,g} \end{bmatrix},$$

and

$$T = \begin{bmatrix} \sum_{i \neq j} f_{i,1} f_{j,1} (\sum_k t_{i,k} t_{j,k}) \\ \vdots \\ \sum_{i \neq j} f_{i,g} f_{j,g} (\sum_k t_{i,k} t_{j,k}) \end{bmatrix},$$

where F is a $(\binom{n}{2} \times p)$ matrix and T is a $(p \times 1)$ matrix. Thus, (1) is equivalent to

$$\mathbf{w}^* = \operatorname{argmin}\left[\frac{1}{2} \times 2(Fw)^T(Fw) - T^T w\right]$$

$$= \operatorname{argmin}\left[\frac{1}{2}(w^T(2F^T F)w + (-2T^T)w)\right] s.t.\mathbf{w} \geq 0.$$

This optimization problem can be addressed using quadratic programming techniques [9]. After calculating the dynamic weights $\mathbf{w}^*$ for each image features, we multiply the image feature values with the corresponding weights for each training image, and finally obtain a new feature space with weighted information. These weighted image feature vectors can then be used for training classifiers. In the experiments, we will show how much the classification results are influenced by our dynamic weighting schema.

3.2 Region-based semantic concept integration

In the real-world applications, an image always contains various semantic concepts and these concepts often intersect with each other, which renders semantic information extraction more challenging. In this section, we investigate the feasibility of utilizing the underlying semantic concepts of textual information as a "guidance" to facilitate image categorization. To address the issue mentioned above, we firstly divide original images into different regions to ensure that the content of each region represents almost the same local pattern, and then based on the local semantic patterns of the images, we propose our *Region-based Semantic Concept Integration* method. Figure 4 shows the framework of our proposed approach, which can be divided into four different sub-processes: *semantic concept extraction*, *image segmentation*, *feature extraction* on each region, and *region-based semantic concept classification*. In the following, we provide algorithmic details of these four processes.

3.2.1 Semantic concept extraction

Although there might be bias or noise caused by the subjectivity of manual labeling textual descriptions of web images, human perception on the semantics of an image can be well approximated by its textual description. Therefore, in order to avoid the above bias or noise, we firstly try to find out an "agreement" between text words and image concepts at the semantic level. In this case, the generic terms are good choices to represent various textual description of web images. Specifically, we initially analyze the textual description related to each image, and then obtain some original high-frequency terms in these texts by using MALLET [24], a java-based package for statistical natural language processing. We then compare the semantics of these high-frequency words and summarize them to several more general semantic concepts using WordNet [25], such as "building", "water", "sky", "grass", etc. (shown in Fig. 4). The general concepts are represented as the hypernyms of the high-frequency words in textual descriptions. For example, a flood is an overflow of an expanse of water that submerges land, and a river is a large natural flowing waterway, while a sea generally refers to a large body of salt water. Here, the "flood", "river" and "sea" can be summarized to a more general semantic concept "water". After that, each
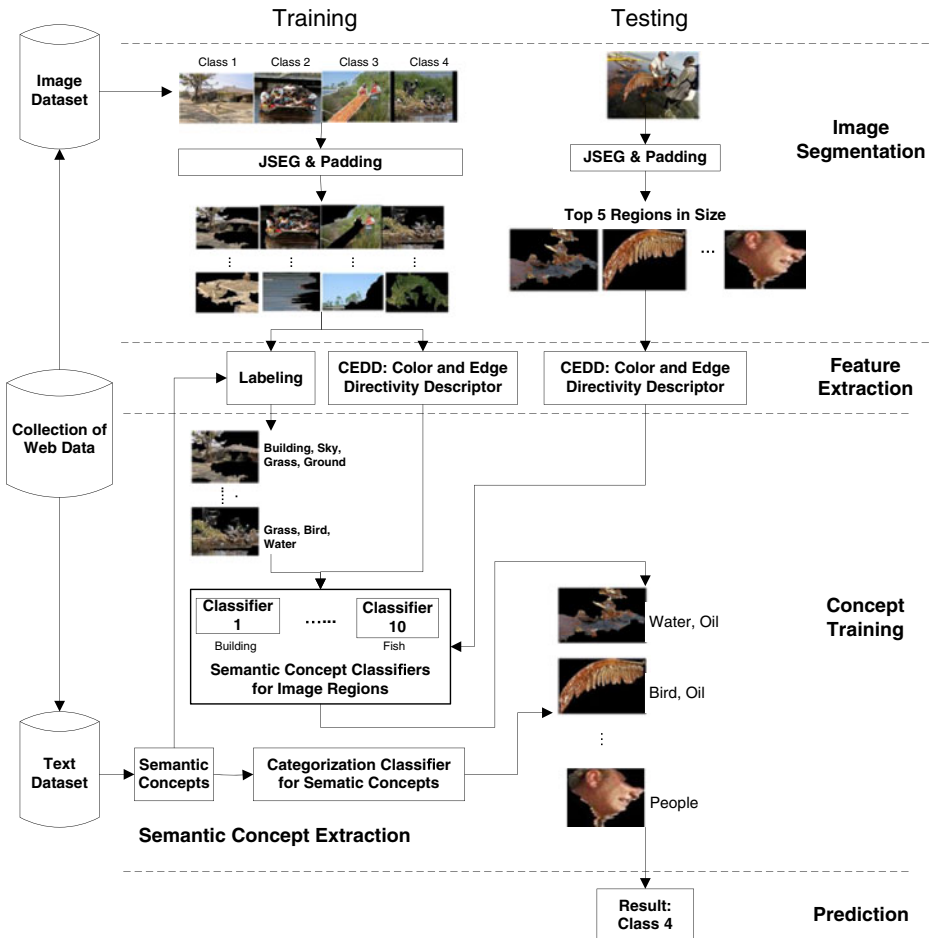
Fig. 4 Framework of region-based concept integration

textual description can be represented by the combination of these concepts. The generalized concepts can provide guidance on how to select image region samples in the training step of semantic concept classifier, as well as to train a concept model that builds the relationship between semantic concepts and original categories, which will be described in Section 3.2.4.

3.2.2 Image segmentation

In order to associate the images with the generalized concepts extracted from the procedure of "Semantic Concept Extraction", we need to segment images into different regions such that each region can be related to one or more semantic concepts. Ideally, image segmentation aims to divide original web images into different regions based on the criterion that each region contains only one object or one part of an object. However, due to the limitation of current segmentation

techniques, it is very difficult to perfectly segment images [20, 22]. In this paper, we use a state-of-the-art segmentation method JSEG (Joint Systems Engineering Group) [7], which segments images based on color and texture information. In the segmentation algorithm, image color space is first quantized into several classes, and a color class-map of each image is then obtained via re-representing each pixel of the original image by its corresponding color class label. After that, the spatial segmentation is performed on this color class-map which can be viewed as a special type of texture composition. Here, a criterion named "J-Value" is used to measure whether the segmentation is reasonable or not. If one image consists of several homogeneous color regions, the color classes will be separated from each other and the value of J is larger. Figure 5 presents some examples of the segmented regions of web images using JSEG. These initial segmented regions are not rectangular; Also, the padding method would pad these unvalued parts with zeros (black parts).

3.2.3 Region feature extraction

After segmenting images into different regions, image feature extraction is performed on each region. Note that color and texture are two of the most general global features in the field of image processing and computer vision.

In order to effectively represent image/region, we adopt CEDD (Color and Edge Directivity Descriptor) [6], which is a new low-level feature descriptor incorporating both color feature and texture feature. In CEDD, a novel but effective method is adopted to integrate a 24-bins color histogram and a 6-bins texture histogram to form a final 144-bins histogram. One of the most important characteristics of CEDD is its low computational power for feature extraction, in comparison with the needs of

Fig. 5 An illustration of 4 different image categories for image segmentation. In each image, the colored irregular but connected part is the segmented region, whereas the black parts are unvalued parts padding with zeros

most of MPEG-7 descriptors. For detailed algorithmic procedure of CEDD, please refer to [6].

3.2.4 Design of semantic concept classifiers

In this step, the design of semantic concept classifiers can be divided into the following two parts:

(1) Semantic concept classifiers for image regions   Based on previous steps of image segmentation and feature extraction, we obtain a set of regions and their corresponding 144-dimension feature vectors. Then we need to train N (the number of generalized semantic concepts obtained from textual descriptions) different semantic concept classifiers respectively using training regions. Each of the training regions is manually labeled with multiple concepts, but each semantic concept classifier is just designed for binary classification identifying whether a region contains this concept or not. For example, as shown in Fig. 4, classifier 1 is designed for identifying whether a region contains the concept "building" or not. If it contains the concept, it will be labeled as *"1"* after passing through the classifier, otherwise it will be labeled as *"0"*. Therefore, given a testing region, it will be fed into these N classifiers respectively and a N-dimension vector is obtained for the input of the categorization classifier.

(2) Categorization classifier for semantic concept   As mentioned in above Section "Semantic Concept Extraction", we have obtained N semantic concepts, and in this part we will design a multi-label classifier to build the relationship between semantic concepts and the original categories in the database. Here we use One-Against-One (OAO) [12] method to design the multi-label SVM classifier. OAO designs an original binary SVM classifier between two random classes of samples, and it needs $k(k - 1)/2$ original binary SVM classifiers. In addition, we use a voting strategy in classification, in which each binary classification is considered to be a voter where votes can be cast for all the regions, and finally each region is designated to be in a class with the maximum number of votes.

For example, in the training stage, we firstly generate 10 most general semantic concepts, such as "building", "water", "sky", "grass", etc. (as shown in Fig. 4). Then the textual description of each training image can be represented by the combination of these 10 concepts. Specifically, each textual description could be represented by a 10-dimension vector where each entity ("1" or "0") represents whether the description contains the general semantic concept or not. After that, these 10-dimension vectors are used to train a multi-label concept model which can be used to predict the original categories of testing image via certain semantic concepts contains in its textual description. In the testing stage, an image is divided into several regions (e.g., five). After each region passing through the 10 concept classifiers respectively, it will be assigned multiple concept labels and also represented by a 10-dimension vector where each entity ("1" or "0") represents whether this region contains the concept or not. Then we integrate these five 10-dimension vectors into a single 10-dimension vector describing whether the original image contains certain concepts or not. Finally, we use the single 10-dimension vector to feed the trained multi-label concept model and obtain the information about which category the testing image belongs to.

3.3 Multimedia information fusion framework for categorization

As mentioned in Section 1, in real world applications, most of web images have corresponding textual descriptions whereas a small portion of images do not have. Moreover, for the images with textual information, in some cases, the textual information is discriminative enough and consequently the similarity between similar textual descriptions is strong, while as in other cases, the similarity between them is relatively weak. Taking both accuracy and efficiency into consideration, we propose a multimedia information fusion framework, which not only seamlessly integrates our two proposed categorization approaches reported recently, but also could automatically choose an appropriate classification model for each different web image to achieve better classification performance. As shown in Fig. 2, for a given testing image with corresponding textual description, we firstly utilize Locality Sensitive Hashing (LSH) to quickly find out its similar textual descriptions of images in the training dataset, and then choose a suitable classification model for the testing image via determining whether the similarity between the testing text and these similar textual descriptions is *weak* or *strong*.

3.3.1 Data preprocessing

Given a testing web image with its corresponding textual description, in order to choose an appropriate classification model to categorize it into a designated class, we need to perform data preprocessing to decide the type of its special property. Naively, we have to compute pairwise similarities between the testing text and all the textual descriptions in the training dataset. However, both time complexity and space complexity of the above calculation of pairwise similarities are very high, and thus computing all the pairwise similarities for every testing image takes too much time, even becomes unrealistic especially when the datasets are large. Moreover, it is not necessary to compute all the similarities each time when a testing image comes, since what we are concerned with in our framework is the quality of the textual description, but not the concrete pairwise similarity value. Specifically, if the quality of the textual description is good and discriminative enough, the testing image can be classified accurately by just choosing the *Text Only Model* as its classification model; if the quality of the textual description is not good enough, we need to combine the textual information with corresponding image information together and choose the *Multi-view Information Fusion Models* to achieve better classification performance. Therefore, in order to reduce the comparisons and improve the efficiency of data preprocessing, we utilize LSH [10] to find similar textual descriptions of images in the training dataset.

LSH is a general probabilistic method to find the nearest neighbors in high dimensional data. The basic idea of LSH is to hash the input items into different buckets so that similar items can be mapped to the same bucket with high probability. In this paper, given a pair of textual descriptions T_1 and T_2, we can find a hash function [13] h such that: if $sim(T_1, T_2)$ is high, then with high probability, $h(T_1) = h(T_2)$; If $sim(T_1, T_2)$ is low, then with high probability, $h(T_1) \neq h(T_2)$. Then we could hash textual descriptions into buckets, and expect that "most" pairs of similar textual descriptions would hash into the same bucket. Specifically, as shown in Fig. 6, we adopt a standard hash function to hash all the textual descriptions of training images into a series of buckets to guarantee that each pair of textual descriptions in the

Fig. 6 A toy example of LSH on textual descriptions of web images
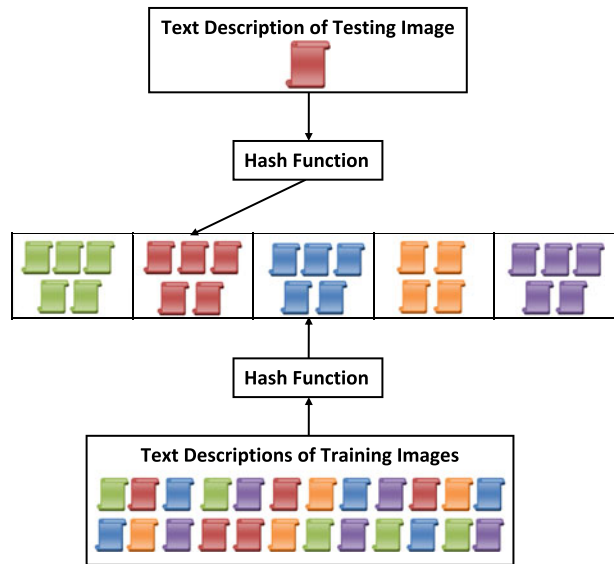
same bucket are sufficiently similar. For a given testing image, we use the same hash function to map it into one of the above buckets. Once the number of the similar textual descriptions in the bucket where the testing text is mapped into is larger than a threshold, we regard that the similarity between the testing text and its similar textual descriptions in training dataset is *strong*; Otherwise, we say that the similarity between them is relatively *weak*.

It is clear that the hash function *h* depends on the similarity metric. Since not every similarity metric has a suitable hash function, for simplicity, in our work, we adopt the Jaccard similarity as our similarity metric to calculate the similarity between the word sets in different textual descriptions. In this way, we can easily avoid the curse of dimensionality in high dimensional data resulted from the sparsity of textual features. Assume that each textual description can be represented as a subset of some universe $U = [n]$ (e.g., the set of vocabulary, where $[n] = \{0, ..., n-1\}$), the Jaccard similarity between two textual descriptions T_1 and T_2 can be defined as:

$$Sim(T_1, T_2) = \frac{|T_1 \cap T_2|}{|T_1 \cup T_2|}. \tag{2}$$

As introduced in [13], the value of $Sim(T_1, T_2)$ can be estimated by precomputing for word set T_1 its sketch $\bar{T}_1 = $min $h_1(T_1),...,$ min $h_k(T_1)$, where the function $h_i (i = 1, ..., k)$ are chosen independently and uniformly at random from some family of functions $\mathcal{H} = \{h_i : [n] \rightarrow [n]\}$. Then we can obtain a good estimation of $Sim(T_1, T_2)$ by comparing sketch $\bar{T}_1$ and $\bar{T}_2$. The benefit of using the sketch is that its size is much smaller than the size of the original textual description, which enables us to significantly reduce the time and space requirements of the procedure. For detailed algorithmic proof of min-wise independent family of hash function, please refer to [13].

Therefore, based on the above analysis, we have three parameters to tune: (1) The number of buckets for the hash function; (2) the threshold of Jaccard similarity in LSH, which directly determines whether each pair of textual descriptions in the same bucket is sufficiently similar or not; and (3) the number threshold of the similar textual descriptions in the bucket where the testing text is mapped in. Here, how to choose appropriate parameters for LSH in our proposed framework is a complex problem to be solved. Specifically, for the former two parameters of LSH, if the number of buckets is set too large, or the threshold of Jaccard similarity is set too high, it is possible that two training images with very similar textual descriptions will be mapped into different buckets, i.e., the measurement of similarity is too strict such that the number of the similar textual descriptions in almost every bucket is less than the third parameter (the number threshold of similar textual descriptions), which will likely result in that the similarity between almost every testing text and its similar textual descriptions in training dataset is regarded as *weak* and consequently choosing inappropriate classification model for it; Otherwise, if the number of buckets is set too small, or the threshold of Jaccard similarity is set too low, dissimilar textual descriptions of training images will also be mapped into the same bucket, i.e., the measurement of similarity is too loose such that it cannot guarantee that each pair of textual descriptions in the same bucket are sufficiently similar, which will probably result in that the similarity between almost every testing text and its similar textual descriptions in training dataset is regarded as *strong* and consequently choosing inappropriate classification model for it. In this paper, taking the impact of these factors into account, we empirically set these parameters based on the experimental evaluation, as shown in Section 4.3.1.

Compared with other tree-based data structures, such as KD-Tree, SR-Tree, LSH could better overcome the curse of dimensionality. Moreover, one of the main advantages of LSH is that it can significantly reduce the time complexity of k-nearest neighbor (KNN) to sub-linear. Therefore, LSH is very suitable for quickly determining whether the similarity between the testing text and its similar textual descriptions in training dataset is *strong* or *weak*: (1) the stronger the similarity between the testing text and its similar textual descriptions in training dataset, the more training textual descriptions in the same bucket with it, and consequently the textual information of the testing image is discriminative enough so that the testing image can be classified accurately. Therefore, we could achieve satisfied classification performance only using textual information, and the *text only model* will be better for the testing image due to its fast speed. (2) While the similarity between the testing text and its similar textual descriptions in training dataset is relatively weak, there are few (even no) training textual descriptions in the same bucket with it. That means the testing image has its individual textual description and the testing text is not similar to almost every training text in the same category with it, and consequently the performance of traditional web image categorization approaches which only use single information source is very limited (as shown in Figs. 11 and 12). In such cases, we need to combine the textual information with corresponding image information together and adopt our two proposed methods to integrate these two kinds of information to perform the classification, and consequently the *multi-view information fusion models* will be more suitable for the testing image due to its better categorization performance.

3.3.2 Categorization via multi-view information fusion models

As mentioned above, in the cases that the testing image has no textual description or in the cases that the similarity between the testing text and its similar textual descriptions in training dataset is relatively weak, our proposed framework will automatically choose the *multi-view information fusion models* to classify these testing images. Here, in order to balance the tradeoff between the classification accuracy and time consumed, whether the *Dynamic Weighting* or the *Region-based Semantic Concept Integration* is chosen as the designated classification model for the testing image could be indicated by web users or randomly chosen by the framework itself.

For the proposed method *dynamic weighting*: given a testing image, we first extract its CEDD feature, and then utilize the dynamic weighting schema to obtain the weighted image features, which will be fed into the train classifiers in Section 3.1 and obtain the classification result.

For the proposed method *region-based semantic concept integration*: when a testing image comes, it will be segmented into several regions. Then the system will choose the top M regions in size automatically (if the actual number of regions for the image is less than M, the system will adopt the actual number of regions) and extract CEDD feature from these regions. Note that the larger the region is, the more information it contains, and if the region is too small, it is difficult to identify the exact semantic meaning and therefore may involve noises. After that, each region will pass through the N different semantic concept classifiers respectively, to identify whether this region contains the concept or not. If the region contains this concept, a concept label will be assigned to this region so that each region will be assigned multiple concept labels, and the region can be represented by a N-dimension vector. Then we integrate these M N-dimension vectors into a single N-dimension vector describing whether the original image contains certain concepts or not. Finally, we use the categorization classifier for semantic concepts to predict which class the original testing image belongs to.

In addition, the proposed framework is a generic multimedia information fusion framework. Besides the above two multi-view classification methods, the three types of existing multi-view classification methods mentioned in Section 2 can also be integrated in the framework. Using such a framework, the classification results of these methods can be significantly improved (as shown in Table 7). Note we can not integrate these types of existing multi-view classification methods (which firstly process each information source separately and then combine them together at either feature level, semantic level, or kernel level) in the framework by simply replacing the above two *Multi-view Information Fusion Models* since they require that each information source (both image and text) must be complete. In other words, when the textual description of each image is complete, we can simply replace the above two models with existing multi-view classification methods. Otherwise, if the textual descriptions of a small portion of testing images are missing, for these images without corresponding textual descriptions, we can only use the *Image Only Model* to classify them; for the other images with textual descriptions, we can determine whether the *Text Only Model* or the *Multi-view Information Fusion Models* should be chosen as an appropriate classification model based on the text similarity. By involving another model—*Image Only Model* which is based on single image feature, the proposed

framework can also be used to categorize images without any textual descriptions in real world applications.

4 Experiments

In this section, our experiments are divided into the following three parts. Firstly, we introduce two different datasets of web multimedia objects (consisting of images and their corresponding textual descriptions) used in our experiments. Secondly, we simply introduce the base classification tools and evaluation measures. Finally, in order to demonstrate the efficacy and efficiency of our proposed multimedia information fusion framework, we report an experimental comparison of five different existing methods for web image categorization mentioned in Section 2, and then compare their performance with that of our proposed framework on both datasets.

4.1 Real world datasets

4.1.1 CNN News (355 objects, 4 topics)

For this dataset, we manually collected 355 colored web images and their corresponding textual descriptions about "the aftermath of disasters" (see Fig. 7 for some examples), which include 4 different topics: (1) Hurricane_building_collapse, (2) Hurricane_flood, (3) Oil_spill_seagrass, and (4) Oil_spill_animal_death. Each category includes 101, 101, 53 and 100 images respectively, and the entire image dataset is split into two parts: about 70 % images (247 in total) are randomly selected for training and the rest 30 % images (108 in total) are taken as the test data.

Note that for *Region-based Semantic Concept Integration* method, we generalize 10 (i.e., $N = 10$) concepts from the textual description set, including "building", "water", "sky", "grass", "oil", "bird", "ground", "people", "helicopter" and "fish". There are 247 training images to be segmented and we use a segmentation threshold 0.55 (Details can be seen in Table 1). Then, with the guidance of these semantic concepts, we choose 1,573 "proper" regions with comparatively large size and containing meaningful object information for training 10 semantic concept classifiers. Moreover, in the testing stage, 513 regions are generated from 108 testing images by automatically choosing the top 5 (i.e., $M = 5$) regions in size (if the actual number of regions is less than 5, the system will adopt the actual number of regions).

4.1.2 Flickr (6, 000 objects, 5 categories)

For this dataset, we randomly choose a subset of 6,000 web images and their corresponding textual descriptions from a public image dataset obtained from Flickr

Table 1 Region information of the training data for CNN News
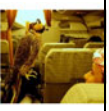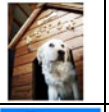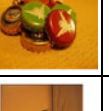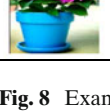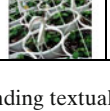
Category	Images	Regions	Regions/images
Topic 1	70	635	9.0714
Topic 2	70	1312	18.7429
Topic 3	37	239	6.4595
Topic 4	70	2946	42.0857
total	247	5132	20.7773
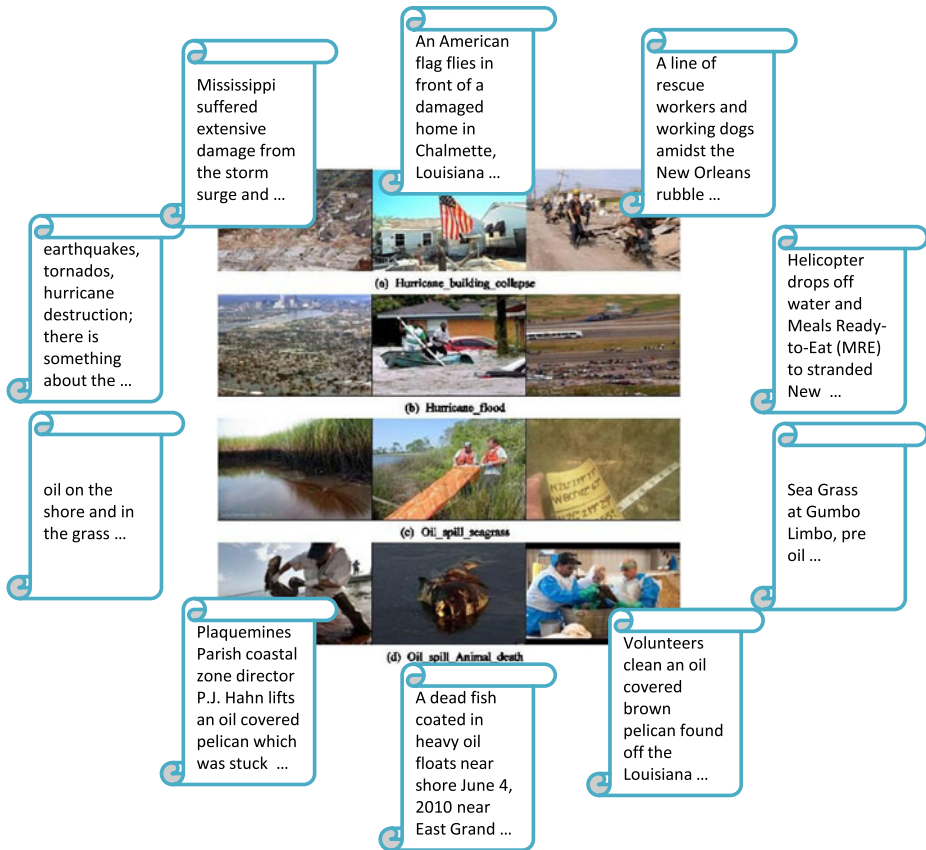
Fig. 7 Example Images and their partial corresponding textual descriptions of the CNN News

[1], as shown in Fig. 8. The textual descriptions of Flickr images are intended to describe the general content of images, in the form of textual tags posted by individual web users. Web users often upload a series of images but assign the same descriptions to the whole image set. Moreover, the textual tags are often ambiguous: the tag "pottedplant" might indicate that the image contains a flower of a pottedplant, a plant of a pottedplant, a pot of a pottedplant, a fertilizer for a pottedplant, or a pot rack for displaying pottedplants, rather than a pottedplant itself. However, sometimes web users might be interested in images related to the above weakly-defined topic [1]. In this paper, we adopt this dataset since images from Flickr contain different image types we defined in Section 1, i.e. images without corresponding textual descriptions, images have their individual descriptions, and images in the same category sharing the same description.

The entire image dataset includes 5 different categories (aeroplane, person, dog, bottle and pottedplant) and each category includes 1,200 images respectively. We split each category into two parts: 800 images (with corresponding textual description) are randomly selected for training and the rest of 400 images are regarded as the test data. Note that for *Region-based Semantic Concept Integration* method,

	Show, plane Airplane Australia Aeroplane, perth, redbull Stunt, airrace		Illustration Photoshop Plane, Flying Drawing Explosion aeroplane		Bird, plane Fun, aeroplane Falcon, Gulfair …		Clouds, plane cockpit aeroplane dials aerobatic flight
	Autumn, dog Nature westie Hund, Eos400d		Dog Toy basket Ground …		Dog, golden retriever Astor, mydog asti kennel		Dog Nagoya Picture …
	Wedding Man, Male Person Brother suit		Man, Water Face, fog Person Dresden …		People, anime Animal, fox Person, kia Drawing Human, draw		Horse Person Dc Washington statues
	Cameraphone shop, bottle Bottles, shopwindow By iphone		Beer, bottle Can, koozie Coozie, cozie		Broken Bottle Dema Person Man, …		Bottle Caps Green Red, …
	Flower Pottedplant Hydrangea Clementine Blue, …		Crochet Pottedplant Daisy Amigurumi anapaulaoli		Pottedplant hangingbasket Hangingbaskets, plant …		Houseplant Pottedplant Potting Zenstyle Corner, …

Fig. 8 Example Images and their partial corresponding textual descriptions of the Flickr

we generalize 11 (i.e., $N = 11$) concepts from the textual description set, including "dog", "person", "aeroplane", "bottle", "flower", "sky", "plant", "ground", "pot", "sofa" and "mountain". There are 4,000 training images to be segmented and we use a segmentation threshold 0.7 (Details can be seen in Table 2). Then, with the guidance of these semantic concepts, we choose 12,164 "proper" regions from 4,000 training images, and 6,168 regions are generated from 2,000 testing images by automatically choosing the top 4 (i.e., $M=4$) regions in size (if the actual number of regions is less than 4, the system will adopt the actual number of regions).

4.2 Classification tools and evaluation measures

4.2.1 Classification tools

In our experiments, we use LIBSVM [5] as our base classification tool. The parameter tuning is done via k-fold cross validation. In the k-fold cross-validation, all the data in data set is divided into k subsets of equal size. The SVM classifier is trained k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the accuracy of the SVM classifier. This process

Table 2 Region information of the training data for Flickr

Category	Images	Regions	Regions/images
1 aeroplane	800	10964	13.705
2 person	800	23930	29.9125
3 dog	800	22192	27.74
4 bottle	800	17832	22.29
5 pottedplant	800	28182	35.2275
total	4000	103100	25.775

is repeated until all the subsets have been used for both training and testing, and the computed average recognition accuracy was used as the performance measure for the designed SVM classifier.

4.2.2 Evaluation measures

We use *Accuracy* as the performance measure in the experiments. Accuracy measures the extent to which each class contains the entities from corresponding true class and is given by:

$$Accuracy = \frac{C}{N} \times 100 \%, \tag{3}$$

where C denotes the number of sample classifies correctly and N means the number of all samples. Generally, the higher the accuracy value, the better the classification result.

4.3 Experimental results

In this section, our experiments are divided into the following three parts. Firstly, we evaluate the efficacy and efficiency of our proposed framework by comparing it with the three single models (*Text Only Model*, *DyW Model* and *Reg Model*) integrated in the framework. Secondly, we report an experimental comparison of the five existing methods mentioned in Section 2 for web image categorization and our previously proposed two classification methods in [21], then compare their performance with that of our proposed multimedia information fusion framework. Finally, in order to illustrate the versatility of our proposed framework, we integrate the three different multi-view classification methods mentioned in Section 2 into the generic framework, and then compare their performance with those of original methods without the framework.

4.3.1 Multimedia information fusion framework

Experiment design   We have evaluated the validity of our two proposed methods in [21], and the initial experimental results demonstrate that these two methods are better than the other three multi-view learning methods mentioned in Section 2. Therefore, we integrate these two proposed methods into our proposed framework as two different multi-view information fusion models. Specifically, the three single models integrated in our proposed framework are:

– *Text Only Model*: For the case that text similarity is *strong*, we extract text features from textual description assigned to the testing image, and then use these text features to feed SVM classifier obtained by training textual descriptions;
– Multi-view Information Fusion Model (1)—*DyW Model*: For the case without text and the case that text similarity is relatively *weak*, we first extract CEDD image features from the testing image, and then multiply the CEDD feature values with the dynamic weights vector $\mathbf{w}^*$ obtained via (1) to compute the weighted image features, and finally use these weighted features to feed SVM classifier obtained by weighted image features of training images;
– Multi-view Information Fusion Model (2)—*Reg Model*: For the case without text and the case that text similarity is relatively *weak*, the system will first

segment the testing image into several regions and choose the top 5 (CNN News) (or 4 (Flickr)) regions in size automatically (if the actual number of regions for the image is less than 5 (or 4), the system will adopt the actual number of regions), and then extract CEDD feature from these regions. After that, each region will respectively pass through the 10 (CNN News) (or 11 (Flickr)) different semantic concept classifiers, and obtain a 10-dimension (or 11-dimension) vector describing whether the region contains certain concepts or not. Then we integrate these 5 (or 4) vectors into a single 10-dimension (or 11-dimension) vector describing whether the original image contains certain concepts or not. Finally, we use the trained categorization classifier for semantic concepts to predict which category the original testing image belongs to.

Experimental results We compare our proposed multimedia information fusion framework which integrates the above 3 models with the 3 base line models (*Text Only Model*, *DyW Model* and *Reg Model*) without using such framework. In this paper, based on the experimental evaluation, for dataset CNN News, we empirically set the number of buckets to be 50, the similarity threshold to be 0.9, and the number threshold of similar textual descriptions to be 10; And for dataset Flickr, the above three parameters are set to be 200, 0.9 and 40 respectively.

Moreover, in our experiments, once one of the *Multi-view Information Fusion Models* has been chosen as the classification model for a testing image, in order to balance the tradeoff between the classification accuracy and time consumed, the framework randomly chooses one model from the *Dynamic Weighting* and the *Region-based Semantic Concept Integration* as the final classification method to predict the testing image. Due to the randomness of selecting models, we respectively run the experiment 5 trials on 108 testing images in CNN News and 2,000 testing images in Flickr, and take the average accuracy as the final result.

The experimental results on both datasets are shown in Tables 3 and 4.

From the comparison results, we have the following observations:

1. Compared with the single models (*Text Only Model*, *DyW Model* and *Reg Model*) in the framework, the classification performance is significantly improved using the multimedia information fusion framework, which demonstrates that our proposed framework could effectively integrate the three single models.

Table 3 Comparison among the 3 base line models and our proposed framework which integrates these 3 models based on the accuracy of web image categorization on CNN News

Category	Without framework			Our
	Text only model (%)	DyW model (%)	Reg model (%)	framework (%)
Topic 1	48.39	67.74	87.10	90.32*
Topic 2	67.74	70.97	54.84	74.19*
Topic 3	25.00	6.25	50.00	62.50
Topic 4	80.00	93.33	100.00	96.67*
Average	59.26	66.67	75.93	83.33*

The results with * are significantly better than the others (under $p < 0.005$). Note that "Topics 1–4" represents Hurricane_building_collapse, Hurricane_flood, Oil_spill_seagrass, and Oil_spill_animal_death respectively

Table 4 Comparison among the 3 base line models and our proposed framework which integrates these 3 models based on the accuracy of web image categorization on Flickr

Category	Without framework			Our
	Text only model (%)	DyW model (%)	Reg model (%)	framework (%)
1 aeroplan	92.75	91.25	95.25	97.50*
2 person	66.75	75.25	90.75	91.00
3 dog	76.25	83.50	82.75	88.75*
4 bottle	86.00	92.25	95.25	96.25*
5 pottedplant	81.75	82.25	93.50	94.50*
Average	80.70	84.90	91.50	93.60*

The results with * are significantly better than the others (under $p < 0.005$)

2. The overall performance on dataset Flickr is better than that of the dataset CNN News because the textual descriptions assigned to the images in Flickr are textual tags which can be regarded as more accurate textual information after preprocessing. Therefore, besides the relatively high performance of the *Text Only Model* itself, under the "guidance" of such more accurate textual descriptions, the results of our two proposed multi-view learning methods are also better.

3. In Table 4, the performance of almost every model on category 2 and 3 is slightly worse than the results on the other 3 categories. This is due to the characteristics of the image dataset. Most of the images in category 2 and 3 contain a lot of ambiguous "semantic noise". For instance, most of the images in category 2 and 3 respectively focus on the concepts "person" and "dog", but "person" and "dog" often appear together in these images. These "noise" would induce our classification models to misclassify these images into the other categories. Even though our proposed framework incorporates the semantic information and shows better results than other single models, "noise" still exists to some extent. And in Table 3, similar situation can be seen that the performance of almost every model on topic 1 and 2 is slightly worse than the results on topic 4.

4. The *Text Only Model* provides better results than multi-view *DyW Model* on topic 3 in Table 3 and category 1 in Table 4 for the reason that most images in topic 3 or category 1 share the same textual descriptions or tags and consequently have high text similarities between them.

5. Comparing these two multi-view information fusion models, the *Reg Model* outperforms the *DyW Model* on topic 1,3 and 4 in Table 3, and category 1, 2, 4 and 5 in Table 4. The reason is that the *Reg Model* could benefit from the semantic information hidden in the text whereas the *DyW Model* only makes use of raw textual information.

In order to evaluate the efficiency of our proposed framework, we further compare the classification accuracy and the time consumed of several different models. The processing time includes: (1) *Text Only Model*: text preprocessing , feature extraction and testing; (2) *DyW Model*: image preprocessing, feature extraction using CEDD, weighted feature calculating and testing; (3) *Reg Model*: image segmentation using JSEG (which costs most time), feature extraction on top 5 (CNN News) or 4 (Flickr) regions using CEDD, concept prediction and integration, and category prediction; (4) *Framework Without LSH*: text preprocessing, pairwise similarity

(a) Classification Accuracy   (b) Time Consumed
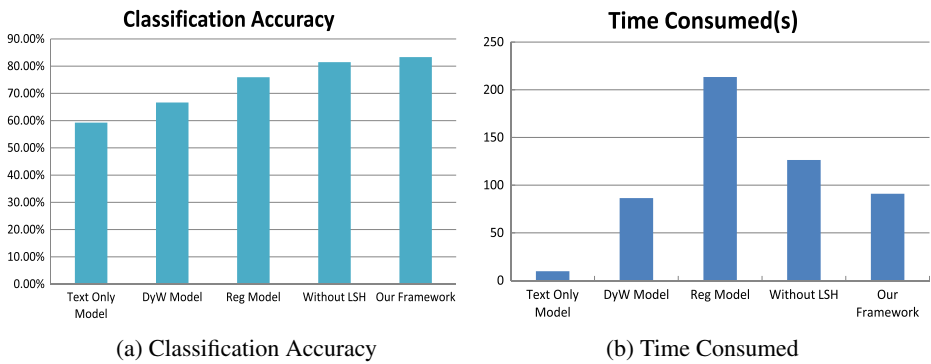
Fig. 9 Comparison of the different models in terms of classification accuracy and time consumed on CNN News

calculating of textual descriptions (which costs most time), model selection and one of the subsequent processing time of the above 3 corresponding models; and (5) *Our proposed Framework*: utilizing LSH instead of the pairwise similarity calculating in model (4), and one of the subsequent processing time of the former 3 corresponding models. For the random selection of these two models, we respectively run our proposed framework 5 trials on 108 testing images in CNN News and 2,000 testing images in Flickr, and take the average accuracy and average time consumed as the final results to illustrate the efficiency of our proposed framework. The comparison results are shown in Figs. 9 and 10, where the time consumed represents the total process time of 108 (CNN News) or 2,000 (Flickr) testing images (Intel(R) Core(TM) i5 CPU M560 2.67GHz, 3.8GB Memory).

From the comparison results, we could observe that:

1. On the one hand, although the *Text Only Model* and the *DyW Model* run fast, the classification accuracy is relatively low; On the other hand, the *Reg Model* has relatively high accuracy but costs much time. In comparison, our proposed framework achieves the best classification performance with relatively

(a) Classification Accuracy   (b) Time Consumed
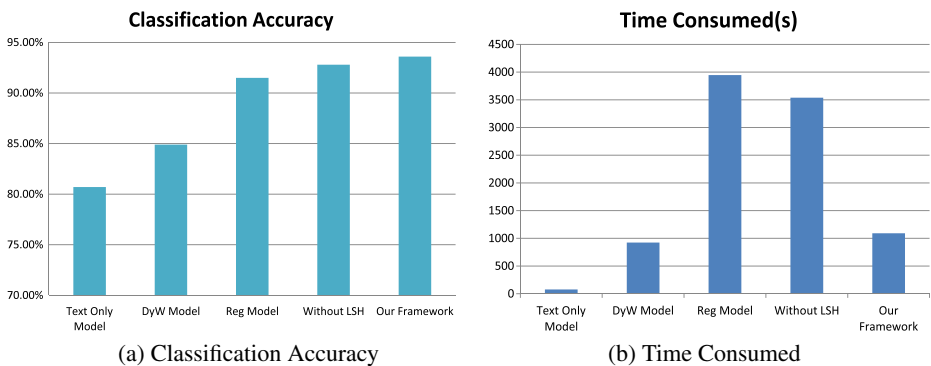
Fig. 10 Comparison of the different models in terms of classification accuracy and time consumed on Flickr

less computation time, which illustrates that the framework could effectively integrate the three models by analyzing the special characteristics of different images.

2. Both of the *Multi-view Information Fusion Models* have their advantages, i.e., the *DyW Model* runs relatively fast, while the classification accuracy of *Reg Model* is higher. Thus, in order to balance the tradeoff between the classification accuracy and time consumed, we randomly select *DyW Model* and *Reg Model* in our experiments. However, in real world application, we allow the web users to indicate which multi-view model is utilized to perform the classification task according to their specific requirements.

3. Compared with the *Framework Without LSH*, our proposed framework outperforms the former in terms of the accuracy and efficiency. Specifically, comparing the results in Figs. 9b and 10b, the time consumed of the *Framework Without LSH* increases rapidly when the dataset becomes large, whereas the time increment of our proposed framework is relatively less. A straightforward explanation for the accuracy increment is that our framework uses LSH to quickly find out similar textual descriptions of the testing textual description in the training dataset. Here, LSH aims to separate the textual descriptions into a set of k-shingles where the probability of any given shingle appearing in any textual description is low. In this way, similar textual descriptions will have more shingles in common, whereas dissimilar textual descriptions share rarely few shingles. And the explanation for the efficiency increment is that the LSH has the ability to reduce the time complexity of KNN to sub-linear. Therefore, taking both classification accuracy and time consumed into consideration, our proposed framework obtains high accuracy with relatively less running time which is more suitable for the large scale datasets in real world applications.

4.3.2 Comparison among different classification methods

Experiment design   For the purpose of comparison, we first implement these two single information source based classification methods, and three different multi-view learning methods for classification mentioned in Section 2, and then compare their classification performance with that of our proposed framework. We choose the above five existing methods since each one of them is well accepted and typical representative of that type of methods. The concrete methods we adopted in our experiments will be described as follows:

– Text-based Classification (*Text* for short): Extract text features from texts assigned to the corresponding images via MALLET [24] and *tf-idf* schema discussed by Salton and McGill [26], and then use these features to feed SVM classifier;

– Image-based Classification (*Img* for short): Extract 144-dimension CEDD image features [6] from the images, and then use these features to feed SVM classifier;

– Feature Integration (*Feat* for short): In the experiments, we adopted one popular method proposed in [30]. Here, we treat unique terms as text features and extract image features using CEDD. Note that for our dataset, the extracted CEDD feature is a 144-dimension vector, while the cardinality of the text features is 1788 (CNN News) or 3530 (Flickr). To balance the contribution of different features to the classification results, we choose the top 144 terms with high frequency as

the text features. We combine the features of text and image together by simply concatenating these two types of features to form a 288-dimension vector as the input of SVM classifier.

– Semantic Integration (*Sem* for short): Train two classifiers based on text features and image features (CEDD) respectively, and then ensemble these two classifiers to be an integrated version via voting schema, similar to the method proposed in [4].

– Hybrid Integration (*Sim* for short): Compute the pairwise similarity using text-based features and image-based features (CEDD) respectively, and then use the weighted summation [27] of these two types of similarities as the similarity measurement between images. Note that different weights can be assigned to the features of different data sources. We tune the weight factor to find the optimal one through the empirical comparison.

Experimental results In Figs. 11 and 12, the comparison results among the above five different methods, our previously reported two classification methods and our proposed framework based on the accuracy of web image categorization are presented.

From the comparison results, we observe that the best performance of web image categorization on two datasets using single-modal approaches is respectively less than 60 % or about 80 %. However, once the text and image data sources are integrated using two-modal information fusion techniques, the categorization performance is improved. The intuitive explanation for the improvement is that two-modal approaches are able to incorporate the advantages of the two data sources together, which leads to better categorization results compared with using only one type of features (here the "advantage" represents the positive contribution of the
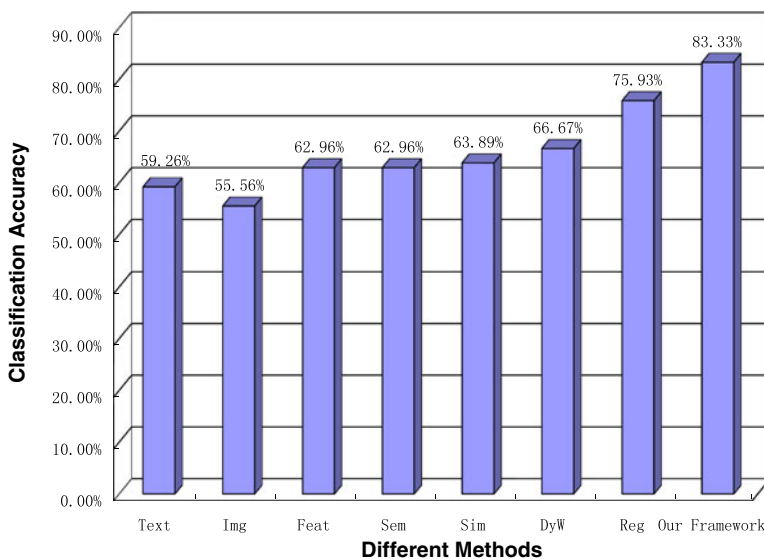
Fig. 11 Comparison among the existing classification methods and our proposed framework based on the accuracy of web image categorization on CNN News
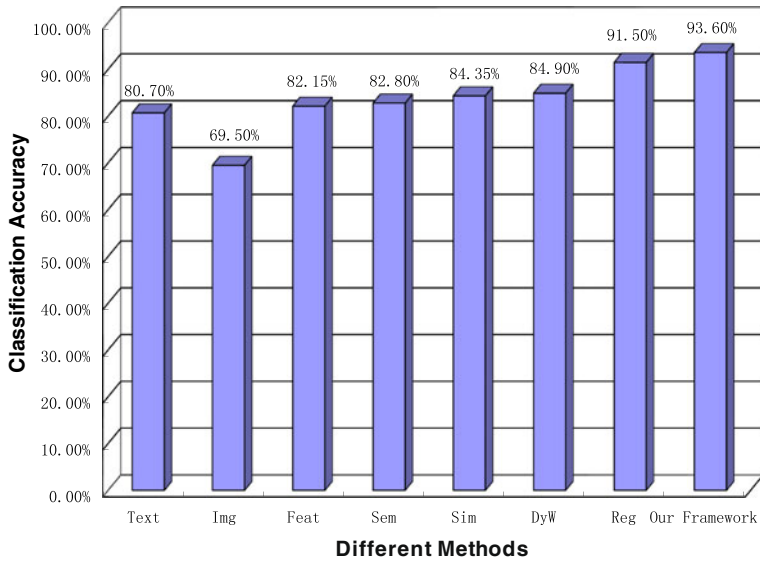
Fig. 12 Comparison among the existing classification methods and our proposed framework based on the accuracy of web image categorization on Flickr

features of certain data source to the image categorization results). In addition, our proposed framework significantly outperforms the others. The reason behind the performance improvement is straightforward: our proposed framework has an important characteristic—employing advantages of one data source to enrich the other data source. In other words, our proposed framework explores inherent connections between two data sources by utilizing text-based information to either find out the best weighting schema for the image-based features or generate the classifiers from the semantic image regions related to text concepts. And compared with the two classification methods we reported previously, our proposed framework provides better results for the reason that the framework can automatically choose a suitable classification model for each testing image according to its special characteristics and consequently achieve better performance.

We further compare the three multi-view learning methods with our proposed framework based on their classification performance over each category of web

Table 5 Comparison among the three multi-view learning methods and our proposed framework on each category in terms of classification accuracy on CNN News

Category	Multi-view learning method			
	Feat (%)	Sem (%)	Sim (%)	Our framework (%)
Topic 1	61.29	58.06	54.84	90.32*
Topic 2	64.52	61.29	67.74	74.19*
Topic 3	31.25	50.00	37.50	62.50
Topic 4	80.00	76.67	83.33	96.67*
Average	62.96	62.96	63.89	83.33*

The results with * are significantly better than the others (under $p < 0.005$)

images. The comparison results are shown in Tables 5 and 6. From the results, we have the following observations:

1. The overall performance on dataset Flickr is better than that on dataset CNN News. The reason can be seen in Section 4.3.1.
2. Compared with the three multi-view learning methods, the categorization results provided by our proposed framework are better than those of others. A straightforward explanation for the accuracy increment is that both of the *Multi-view Information Fusion Models* integrated in our proposed framework involve the interaction between different information sources, i.e., one data source can provide "guidance" for another on how to perform categorization task better, whereas the above three multi-view learning methods process each information source separately and then combine them together at different levels.
3. In Table 5, almost every method provides reasonable performance on topic 1 and 2, and good results on topic 4, whereas the performance on topic 3 is very low. After analysis, we found that: (1) most of the images in topic 3 are about "grass"; however, "grass" appears in almost all the topics, which results in the misclassification of the images; and (2) compared with the other three topics, the size of the training images in topic 3 is relatively small.
4. Compared among the three multi-view learning methods, we could observe that: (1) when the size of the training dataset is relatively small (topic 3 in Table 5), *Sem* outperforms the other two; and (2) As shown in Table 6, the performance of *Sim* on almost every category is slightly better than those of the other two. We think the reason is likely to be the characteristics of our image dataset. Compared with the dataset CNN News, most of the images in dataset Flickr indeed vary greatly, and consequently *Sim* is more preferred.

4.3.3 Versatility of the multimedia information fusion framework

As mentioned above, the proposed framework is a generic multimedia information fusion framework which is not limited to our previously reported two approaches, and it can also be used for other multi-view classification methods or models. In order to illustrate the versatility of our proposed framework, we integrate the three different multi-view classification methods mentioned in Section 2 into the generic framework, and then compare their performance with those of the original methods without the framework, as shown in Table 7.

Table 6 Comparison among the three multi-view learning methods and our proposed framework on each category in terms of classification accuracy on Flickr

Category	Multi-view learning method			
	Feat (%)	Sem (%)	Sim (%)	Our framework (%)
1 aeroplane	94.50	95.75	97.00	97.50*
2 person	72.25	72.75	75.25	91.00
3 dog	78.00	77.50	79.25	88.75*
4 bottle	83.75	84.75	86.25	96.25*
5 pottedplant	82.25	83.25	84.00	94.50*
Average	82.15	82.80	84.35	93.60*

The results with * are significantly better than the others (under $p < 0.005$)

Table 7 Comparison among the three multi-view learning methods with and without the proposed framework in terms of classification accuracy on both datasets

Dataset	Method	Without framework (%)	With framework (%)
CNN News	Feat	62.96	68.52
	Sem	62.96	70.37
	Sim	63.89	72.22
	DyW	66.67	83.33*
	Reg	75.93	
Flickr	Feat	82.15	87.55
	Sem	82.80	89.35
	Sim	84.35	88.70
	DyW	84.90	93.60*
	Reg	91.50	

The results with * are significantly better than the others (under $p < 0.005$)

From the results, we observe that the classification results of the multi-view learning methods are significantly improved to some extent by using such a framework. Specifically, the performance improvements of the three multi-view learning methods on dataset CNN News are 5.56 %, 7.41 % and 8.33 % respectively, and the improvements on dataset Flickr are 5.40 %, 6.55 % and 4.35 % respectively. The experiments illustrate that our proposed framework is a generic multimedia information fusion framework which is not only suitable for our previously reported two approaches, but also can be used to integrate other existing multi-view classification methods or models to achieve better performance.

5 Conclusions

In our work, we studied the problem of combining two data sources (text and image) to effectively perform web image categorization and showed that such combination could lead to better classification results comparing with using individual data sources. Also, we investigated the feasibility of using the textual information as a "guidance" for image categorization by proposing two novel multi-view learning methods (**Dynamic Weighting** and **Region-based Semantic Concept Integration**). Both of the two methods could effectively utilize the image-related text data to find out better scheme to classify the images, and outperformed the previous methods in terms of the accuracy of classification results, but not effective enough especially when the dataset became very large.

Based on these preliminary studies, to further improve the efficiency of our two proposed methods for large scale image datasets, we proposed a novel **multimedia information fusion framework** by seamlessly integrating these two proposed methods into our framework so that they could effectively handle the scalability of web images. The proposed framework was very significant because it could not only automatically choose an appropriate classification model for each testing image according to its special characteristics and consequently achieved better performance with relatively less computation time for large scale datasets, but also could address the problem that the textual descriptions of a small portion of web images are missing. Moreover, our proposed framework is a generic multimedia information

fusion framework which is not limited to our previously reported two approaches, and it can also be used to integrate other existing multi-view classification methods or models. Finally, we evaluated the framework on two different types of web image datasets, one is a manually collected web multimedia object (related to the events after disasters) dataset, and the other is a large public available image dataset from the Flickr. The empirical experiments demonstrated the efficacy and efficiency of our proposed framework so that it could handle large scale image categorization, and provide solid basis for the subsequent procedures of image retrieval and other web applications involving image categorization.

References

1. Allan M, Verbeek J (2009) Ranking user-annotated images for multiple query terms. In: British machine vision conference. URL http://lear.inrialpes.fr/pubs/2009/AV09
2. Bishop C (2006) Pattern recognition and machine learning. Springer, New York
3. Blei D, Jordan M (2003) Modeling annotated data. In: Proceedings of the 26th annual international ACM SIGIR conference on research and development in informaion retrieval. ACM, pp 127–134
4. Carter R, Dubchak I, Holbrook S (2001) A computational approach to identify genes for functional RNAs in genomic sequences. Nucleic Acids Res 29(19):3928
5. Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(3): Article 27
6. Chatzichristofis S, Boutalis Y (2008) Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In: Proceedings of the 6th international conference on computer vision systems, pp 312–322
7. Deng Y, Manjunath BS (2001) Unsupervised segmentation of color-texture regions in images and video. IEEE Trans Pattern Anal Mach Intell 23(8):800–810
8. Giacinto G, Roli F, Fumerga G (2002) Unsupervised learning of neural network ensembles for image classification. In: Proceedings of the IEEE-INNS-ENNS international joint conference on neural networks, vol 3. IEEE, pp 155–159
9. Gill P, Murray W, Wright M (1981) Practical optimization. Academic Press
10. Gionis A, Indyk P, Motwani R (1999) Similarity search in high dimensions via hashing. In: Proceedings of the 25th international conference on very large data bases. Morgan Kaufmann, pp 518–529
11. Hare J, Lewis P (2010) Automatically annotating the MIR Flickr dataset. In: Proceedings of the 2nd ACM international conference on multimedia information retrieval
12. Hsu C, Lin C (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13(2):415–425
13. Indyk P (1999) A small approximately min-wise independent family of hash functions. In: Proceedings of the tenth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics, pp 454–456
14. Jordan M, Jacobs R (1994) Hierarchical mixtures of experts and the EM algorithm. Neural Comput 6(2):181–214
15. Kalva P, Enembreck F, Koerich A (2007) Web image classification based on the fusion of image and text classifiers. In: Proceedings of the 9th international conference on document analysis and recognition. IEEE Computer Society, pp 561–568
16. Lanckriet G, Cristianini N, Bartlett P, Ghaoui L, Jordan M (2004) Learning the kernel matrix with semidefinite programming. J Mach Learn Res 5:27–72
17. Lee W, Verzakov S, Duin R (2007) Kernel combination versus classifier combination. In: Multiple classifier systems, pp 22–31
18. Li T, Ogihara M (2005) Semisupervised learning from different information sources. Knowl Inf Syst 7(3):289–309

19. Li H, Tang J, Li G, Chua T (2008) Word2image: towards visual interpreting of words. In: Proceeding of the 16th ACM international conference on multimedia. ACM, pp 813–816
20. Li L, Socher R, Fei-Fei L (2009) Towards total scene understanding: classification, annotation and segmentation in an automatic framework. In: IEEE conference on computer vision and pattern recognition. IEEE, pp 2036–2043
21. Li L, Lu W, Li J, Li T, Zhang H, Guo J (2011) Exploring interaction between images and texts for web image categorization. In: Proceedings of FLAIRS, pp 45–50
22. Liu Y, Zhang D, Lu G (2008) Region-based image retrieval with high-level semantics using decision tree learning. Pattern Recogn 41(8):2554–2570
23. Liu X, Cheng B, Yan S, Tang J, Chua T, Jin H (2009) Label to region by bi-layer sparsity priors. In: Proceedings of the seventeen ACM international conference on multimedia. ACM, pp 115–124
24. McCallum A (2002) MALLET: a machine learning for language toolkit. http://mallet.cs.umass.edu
25. Miller G (1995) WordNet: a lexical database for English. Commun ACM 38(11):39–41
26. Salton G, McGill M (1986) Introduction to modern information retrieval. McGraw-Hill, Inc., New York, NY, USA
27. Schölkopf B, Smola A (2002) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT
28. Shao B, Ogihara M, Wang D, Li T (2009) Music recommendation based on acoustic features and user access patterns. IEEE Trans Audio Speech Lang Process 17(8):1602–1611
29. Wang Y, Gong S (2007) Refining image annotation using contextual relations between words. In: Proceedings of the 6th ACM international conference on image and video retrieval. ACM, pp 425–432
30. Wu L, Oviatt S, Cohen P (2002) Multimodal integration-a statistical view. IEEE Trans Multimedia 1(4):334–341
31. Wu Y, Chang E, Chang K, Smith J (2004) Optimal multimodal fusion for multimedia data analysis. In: Proceedings of the 12th annual ACM international conference on multimedia. ACM, pp 572–579
32. Yin Z, Li R, Mei Q, Han J (2009) Exploring social tagging graph for web object classification. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 957–966
33. Zhu Q, Yeh M, Cheng K (2006) Multimodal fusion using learned text concepts for image categorization. In: Proceedings of the 14th annual ACM international conference on multimedia. ACM, pp 211–220

Wenting Lu is currently a Ph.D. candidate in the Pattern Recognition and Intelligence System Lab at Beijing University of Posts and Telecommunications. Her research interest is in content-based image retrieval and image categorization.

Lei Li is currently a Ph.D. candidate in the School of Computing and Information Science at Florida International University. His research interest includes data mining and recommendation systems.
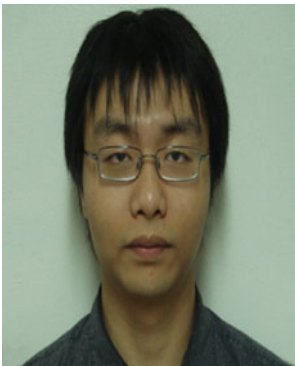
Jingxuan Li is currently a Ph.D. candidate in the School of Computing and Information Science at Florida International University. His research interests are in data mining and social network analysis.

Tao Li received his Ph.D. degree in computer science in 2004 from the University of Rochester. He is currently an associate professor in the School of Computer Science at Florida International University. His research interests are in data mining, machine learning and information retrieval. He is a recipient of USA NSF CAREER Award and multiple IBM Faculty Research Awards.
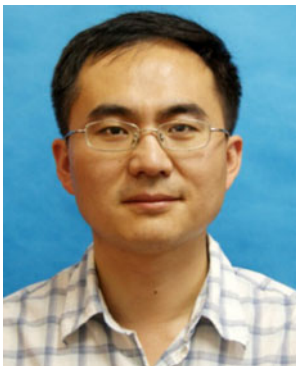
Honggang Zhang received the BS degree from the Department of Electrical Engineering, Shandong University in 1996, the master and PhD degrees from the School of Information Engineering, Beijing University of Posts and Telecommunications (BUPT) in 1999 and 2003, respectively. He is currently an associate professor and director of Web Search Center at BUPT. His research interests include image retrieval, computer vision and pattern recognition.

Jun Guo received B.E. and M.E. degrees from BUPT, China in 1982 and 1985, respectively and Ph.D. degree from the Tohuku-Gakuin University, Japan in 1993. At present he is a professor and the dean of School of Information and Communication Engineering, BUPT. His research interests include pattern recognition theory and application, information retrieval, content based information security, and network management.

A MULTIMEDIA SEMANTIC RETRIEVAL MOBILE SYSTEM BASED ON HIDDEN COHERENT FEATURE GROUPS

Yimin Yang, Hsin-Yu Ha, Fausto C. Fleites, Shu-Ching Chen

School of Computing and Information Sciences,
Florida International University,
1200 SW 8th Street,
Miami, FL 33199, USA

ABSTRACT

A multimedia semantic retrieval system based on Hidden Coherent Feature Groups (HCFGs) is proposed to support multimedia semantic retrieval on mobile applications. The system is able to capture the correlation between features and partition the original feature set into HCFGs, which have strong intra-group correlation while maintaining low inter-correlation. A novel, multi-model fusion scheme is presented to effectively fuse the multi-model results and generate the final ranked retrieval results. In addition, to incorporate user interaction for effective retrieval, the proposed system also features a user feedback mechanism to refine the retrieval results.

Index Terms— Multimedia semantic retrieval, feature, correlation, affinity propagation, multi-model fusion, mobile, application.

1. INTRODUCTION

Due to the proliferation of internet-connected mobile devices such as mobile phones and tablet computers, people commonly upload all kinds of multimedia data to social sites such as Flickr, YouTube, and Facebook. By 2014, the rate of data sharing via mobile devices will be 14 times greater than in 2008 [1]. The multimedia research community has addressed this challenge by developing systems that allow the semantic retrieval of multimedia data. Nevertheless, research on this problem remains active given the difficulty posed by the semantic gap between the low-level representation of multimedia data and their high-level semantic meaning.

A typical concept retrieval framework is built upon the tasks of feature extraction, model training, classification, and ranking. Although much research has been done on each of these tasks [2, 3], significant challenges still remain such as the effective analysis and utilization of multi-source, high-dimensional features. To effectively retrieve meaningful semantics from rapidly growing multimedia data, it is essential to capture the correlations among features to enhance the effectiveness of model training and classification tasks. In order to tackle this problem, researchers usually perform either a linear combination of the original features from different modalities, or use statistical techniques such as principle component analysis (PCA) and independent component analysis (ICA) to transform the original features into another space and select the most "important" features. The problem with these statistical methods is that they try to make each feature independent in the transformed space and may lose some information during model training on the transformed feature set. Overall these methods do not thoroughly explore the correlation between features with different types and may not fully utilize the complementary information from various features. For instance, the tag "tree" implies the color "green" for the semantic concept "forest", which is considered as a "hidden" correlation between features. In addition to the feature analysis problem, another issue is the integration of multiple models in the semantic space by fusing the decisions (scores) from different models. The challenges lie in how to select the training models for different feature types and how to evaluate the confidence of the decision from different models and take that into account when performing final fusion.

With the aforementioned existing problems and challenges, we propose a correlation based feature analysis method to explore Hidden Coherent Feature Groups (HCFGs) and present a novel, multi-model fusion scheme. Specifically, we analyze the correlation between each feature pair and use the affinity propagation algorithm to separate the original feature set into different feature groups (HCFGs), where the intra-group correlation is maximized and the inter-group correlation is minimized. Subsequently, one model is trained for each of the HCFGs, and the HCFGs with best performance in the training phase are chosen for the final score fusion. A mobile system is presented which utilizes the proposed framework as its retrieval engine and features a user feedback mechanism for improving retrieval performance.

The main contributions of the proposed multimedia semantic retrieval system for mobile applications are summarized as follows:

- Incorporates a correlation-based feature analysis method that generates Hidden Coherent Feature Groups (HCFGs) via similarity matrix construction and feature clustering. The generated HCFGs have a strong intra-group relationship while maintaining a low inter-group correlation.
- Presents a multi-model fusing strategy that identifies the best HCFGs and effectively fuses multiple uncorrelated models at the decision level by analyzing the reliability of each model based on training performance.
- Develops a multimedia semantic retrieval mobile system that integrates the correlation-based feature analysis method and the multi-model fusion scheme as well as the user feedback mechanism to provide efficient retrieval services.

The rest of paper is organized as follows. Section 2 introduces the related work on multimedia semantic retrieval mobile systems. Section 3 presents the details of the proposed system. Section 4 discusses the experimental results, and section 5 concludes the paper.

2. RELATED WORK

The related work on multimedia semantic retrieval systems can be generally categorized based on the following two perspectives: (a) from the back-end algorithm point of view and (b) from the system performance point of view. For category (a), the work in [2] thoroughly overviews the state-of-the-art in multimedia semantic retrieval and identifies several prevalent research topics that have potential for improving multimedia retrieval by bridging the well-known semantic gap. These topics are: human-centered computing, learning and semantics, new features and similarity measures, new media, browsing and summarization, indexing, and evaluation/benchmarking. In this paper, we mainly focus on extracting the semantics in multimedia data by exploring the correlations in feature space and present a multi-model fusion scheme for effective retrieval. There are two subtopics involved: feature space analysis and multi-model fusion.

With the purpose of effectively retrieving semantic concepts from multimedia data, many research works have been done to project the original feature space to a low dimensional space using linear or nonlinear mapping methods [4], and further derive the Euclidean distance for each instance pair to represent the pairwise similarity. For example, Huang et al. [5] propose an image retrieval system using only Euclidean distance of image color features to calculate the ranking score for each image per specific concept. In [6], Smaragdis et al. propose to employ the subspace projection on all the features by using PCA and ICA to find out the maximally independent subspaces. Other works use statistical techniques to capture multimedia

correlation in the feature level. In [7], Nefian et al. adopt an early fusion approach to combine audio and visual features for speech recognition by using the coupled hidden Markov model (CHMM) and dynamic Bayesian networks. Recently, Canonical Correlation Analysis (CCA), another powerful statistical technique, has found its application in finding linear mapping that maximizes the cross-correlation between two feature sets [8]. However, besides the correlation among multimedia data instance, the complementary and mutual information among features from multiple modalities should also be extensively exploited as a reference knowing how to integrate them to improve the performance and avoid possible information loss during the transformation between different feature spaces.

Other than correlation captured at the feature level, the correlation among different models and model confidence toward extracting semantic concepts should also be learned. In [9], separate generative probabilistic models are learned for different classifiers, respectively. Then the scores are combined afterwards to yield a final detection score. In [10], Chen et al. propose a fusion strategy to combine ranking scores from both tag-based and content-based models, where the adjustment, reliability, and correlation of ranking scores from different models are all considered. To leverage the correlation from both feature level and model level, Bendjebbour et al. [11] perform fusion at both levels. At the feature level, the mass of a given pixel based on two sensors is computed and fused; while at the decision level, the HMM outputs are combined. In [12], CCA is used to fuse audio-visual features with joint subspace learning at different granularity and the final decision is made based on the Bayesian decision fusion of multiple HMM-based classifiers. Although many attempts have been made to utilize two kinds of correlation among multimedia data, the performance is far from being satisfactory.

Regarding category (b), most of the mobile multimedia retrieval systems mainly focus on improving performance in terms of transmission time. In [13], David et al. propose to first compress low-level feature descriptors such as Compressed Histogram of Gradients (CHoG) and progressively transmit compressed data to avoid having network transmission latency. Another way to expedite multimedia retrieval process is to unify the approach of retrieving and processing various multimedia data. A multimedia query language called MPEG Query Format (MPQF) is introduced in [14] to save complex interpretation among all kinds of description formats by generally expressing multimedia requests. Different from the above-mentioned research work, our proposed framework decides to perform off-line training on server side and upload them periodically with the corresponding concept relationship, thus users can retrieve in real time a set of well-trained models without having end-to-end network latency.

3. PROPOSED FRAMEWORK

The proposed multimedia semantic retrieval mobile system is depicted in Figure 1. The system design follows a Model-View-Controller (MVC) pattern. The Model part (labeled as A) implements the main logic of the system, i.e., a retrieval model built from the fusion of multiple classification models, which are based on hidden feature groups. The usage of the retrieval model consists of training and testing phases. During the training phase, the meta-

using the learned model during the testing phase. All the processed data and the trained models are stored in the production database. The multimedia retrieval Controller (labeled as C) translates user input into operations on the model and controls the data transfer between the front-end user interface and the back-end server through a REST API. Finally, the View (labeled as B) generates and presents output to users. The detailed architecture of the front-end mobile application as well as the user interface will be discussed in section 4.4. Following is a specific description
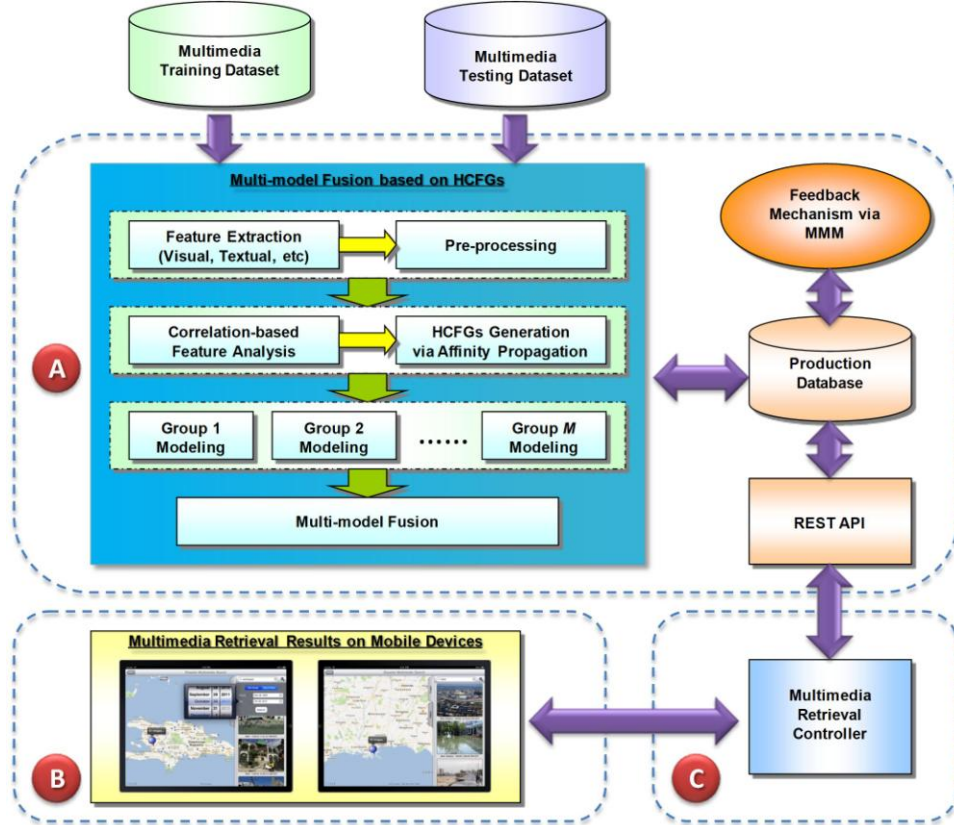
model is trained based on training data with ground-truth information, and unknown multimedia data are classified

of the steps for building a retrieval model based on multi-model fusion.

Fig. 1. Multimedia semantic retrieval mobile system based on HCFGs.

The proposed system builds the retrieval model following a five-step process that consists of (a) feature extraction, (b) pre-processing, (c) correlation-based feature analysis and clustering, (d) model training, and (e) model fusion. Firstly, in the first two steps, the system extracts visual features (e.g., HOG, CEED) from the training data and performs pre-processing to normalize the features and remove those with relatively low variance. Secondly, in the correlation-based feature analysis and clustering step, the system computes a feature similarity matrix based on correlation coefficients for all pairs of retained features and applies the Affinity Propagation (AP) algorithm to cluster the feature set to obtain multiple Hidden, Coherent Feature Groups (HCFGs) that exhibit low inter-group correlation

and high intra-group correlation. Subsequently, the model-training step builds a classification model for each discovered feature group. Finally, the model fusion step combines the individual models using the proposed multi-model fusion strategy (section 3.4). Such a partition of the feature set into HCFGs aims at "untapping" hidden feature groups that will enhance the predictive power of the fused model.

When a query is issued to the system, the system performs feature extraction and pre-processing and groups the features into the same HCFGs identified in the training phase. The HCFGs are then fed to the trained models obtained during the model-training step. The generated testing scores are afterward fused and ranked. The ranked

results are shown via the mobile application. In addition, the system contains a user feedback component that incorporates user interactions in the retrieval process to refine the retrieval results.

3.1. Visual Feature Extraction

The feature set utilized in the proposed system consists of Histogram of Oriented Gradients (HOG), Color and Edge Directivity Descriptor (CEDD), as well as other low-level visual features.

3.1.1. HOG Descriptor

The essential thought behind the HOG descriptors is that local object appearance and shape within an image can be characterized by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. The HOG descriptor maintains a few key advantages over other descriptor methods. It captures local edge or gradient structure that is invariant to a low degree of geometric and photometric transformations in the local area.

3.1.2. CEDD Descriptor

CEDD is a popular low-level feature descriptor which incorporates both color and texture features in a histogram. The size of CEDD is limited to 54 bytes per image, making this descriptor suitable for large image databases. The CEDD histogram is composed of 6x24=144 regions, where the six regions are determined by the texture component and the 24 regions are originated from the color component.

3.1.3. Low-level visual features

The extracted low-level features include 48-dimension features for color histogram in the HSV space; 120-dimension local features for color moment in the YCbCr space; and 260-dimension features for texture wavelet.

3.2. Feature Correlation Analysis (FCA)

In this paper we propose a feature correlation analysis method that explores the interrelationships amongst the features to lay down the basis for the identification of HCFGs (elaborated in section 3.3).

Let $X = \{x_i\}_{i=1}^{N}$ be a given dataset, where $x_i \in R^L$ represents each instance in the dataset, and N and L are the number of instances and the dimension of the feature set $\{\mathbf{f}^i\}_{i=1}^{L}$, respectively. Then the feature matrix $\mathbf{F}$ of $\mathbf{X}$ is represented as

$$\begin{bmatrix} \mathbf{f}_{11} & \mathbf{f}_{12} & \cdots & \mathbf{f}_{1L} \\ \mathbf{f}_{21} & \mathbf{f}_{22} & \cdots & \mathbf{f}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_{N1} & \mathbf{f}_{N2} & \cdots & \mathbf{f}_{NL} \end{bmatrix},$$

where the ith column represents $\mathbf{f}^i$ and rows are instances in $\mathbf{X}$. Let $(\mathbf{f}^j, \mathbf{f}^k)$, $1 \le j$, $k \le L$, be a feature pair, then the correlation coefficient between them can be calculated as follows

$$C_{\mathbf{f}^j, \mathbf{f}^k} = \frac{\sum_{i=1}^{N}(\mathbf{f}_i^j - \overline{\mathbf{f}^j})(\mathbf{f}_i^k - \overline{\mathbf{f}^k})}{\sqrt{\sum_{i=1}^{N}(\mathbf{f}_i^j - \overline{\mathbf{f}^j})^2}\sqrt{\sum_{i=1}^{N}(\mathbf{f}_i^k - \overline{\mathbf{f}^k})^2}}, \quad (1)$$

where $\overline{\mathbf{f}^j}$ and $\overline{\mathbf{f}^k}$ are the mean values of $\mathbf{f}^j$ and $\mathbf{f}^k$ respectively.

The above correlation coefficients analysis method is based on the calculation of the Pearson product-moment correlation coefficient, which implies the assumption of the normally distributed data and the linear relationship between feature variables. However, this is not always the case. In order to take into account the situation where the feature variables follow a non-linear relationship, we propose another correlation estimation method based on the Spearman's rank correlation coefficients, which use the ranks of the observations instead of their values and are calculated as

$$C_{\mathbf{v}^j, \mathbf{v}^k} = \frac{\sum_{i=1}^{N}(\mathbf{v}_i^j - \overline{\mathbf{v}^j})(\mathbf{v}_i^k - \overline{\mathbf{v}^k})}{\sqrt{\sum_{i=1}^{N}(\mathbf{v}_i^j - \overline{\mathbf{v}^j})^2}\sqrt{\sum_{i=1}^{N}(\mathbf{v}_i^k - \overline{\mathbf{v}^k})^2}}, \quad (2)$$

where $\mathbf{r}$ is the rank representation[1] of the feature variable $\mathbf{f}$. Finally, the feature correlation matrix $\mathbf{C}$ is constructed as

$$\begin{bmatrix} C_{\mathbf{v}^1, \mathbf{v}^1} & C_{\mathbf{v}^1, \mathbf{v}^2} & \cdots & C_{\mathbf{v}^1, \mathbf{v}^L} \\ C_{\mathbf{v}^2, \mathbf{v}^1} & C_{\mathbf{v}^2, \mathbf{v}^2} & \cdots & C_{\mathbf{v}^2, \mathbf{v}^L} \\ \vdots & \vdots & \ddots & \vdots \\ C_{\mathbf{v}^L, \mathbf{v}^1} & C_{\mathbf{v}^L, \mathbf{v}^2} & \cdots & C_{\mathbf{v}^L, \mathbf{v}^L} \end{bmatrix},$$

where $\mathbf{v}$ could be the feature variable $\mathbf{f}$ or it's rank vector $\mathbf{r}$. Each element in the matrix presents the correlation coefficient between each feature pair, creating a symmetric matrix, i.e., $C_{\mathbf{v}^j, \mathbf{v}^k}$ equals $C_{\mathbf{v}^k, \mathbf{v}^j}$.

All the correlation coefficients are calculated based only on the positive instances, thus identifying relationships between the features in a supervised manner, i.e., per concept. In addition, the inclusion of the negative instances may hinder the discovery of correlations between feature pairs. An added benefit is the improved computational efficiency of the system, which is an important requirement in mobile systems.

---

[1] The rank representation means the rank of a variable in a feature vector with a specific order (e.g., by value).

3.3. Feature Grouping via Affinity Propagation

Because of its simplicity, general applicability, and performance, the affinity propagation (AP) algorithm has found application in the fields of science and engineering [15], which inspires us to adapt it to our framework for feature clustering. Specifically, we choose to use AP algorithm for the following reasons:

- AP generates clusters with much lower error than other clustering methods, such as k-means and mixtures of Gaussian.
- AP is deterministic, i.e., its clustering results do not depend on initialization, unlike most clustering methods such as k-means.
- AP is able to automatically determine the number of clusters.

Considering each feature as a data point, the input for AP is the similarity matrix S, with each element computed as

$$s(\mathbf{v}^j, \mathbf{v}^k) = C_{\mathbf{v}^j, \mathbf{v}^k}. \tag{3}$$

The AP algorithm propagates affinities by passing two types of messages between two data points (e.g., features $\mathbf{v}^j$ and $\mathbf{v}^k$) [16] as follows:

- The "responsibility" $r(\mathbf{v}^j, \mathbf{v}^k)$ sent from $\mathbf{v}^j$ to $\mathbf{v}^k$, representing how well $\mathbf{v}^k$ serves as the exemplar of $\mathbf{v}^j$ considering other potential exemplars for $\mathbf{v}^j$.
- The "availability" $a(\mathbf{v}^j, \mathbf{v}^k)$ sent from $\mathbf{v}^k$ to $\mathbf{v}^j$, reflecting how appropriate $\mathbf{v}^j$ chooses $\mathbf{v}^k$ as its exemplar considering other potential features that may choose $\mathbf{v}^k$ as their exemplar.

The responsibility and availability are updated iteratively using the following equations

$$r(\mathbf{v}^j, \mathbf{v}^k) \leftarrow s(\mathbf{v}^j, \mathbf{v}^k) - \max_{l:l \neq k}(a(\mathbf{v}^l, \mathbf{v}^j) + s(\mathbf{v}^j, \mathbf{v}^l)), \tag{4}$$

$$a(\mathbf{v}^k, \mathbf{v}^j) \leftarrow \min(0, r(\mathbf{v}^k, \mathbf{v}^k) + \sum_{l:l \notin \{k, j\}} \max\{0, r(\mathbf{v}^l, \mathbf{v}^k)\}). \tag{5}$$

The self-availability is updated as

$$a(\mathbf{v}^k, \mathbf{v}^k) \leftarrow \sum_{l:l \neq k} \max\{0, r(\mathbf{v}^l, \mathbf{v}^k)\}. \tag{6}$$

This message reflects an accumulated confidence that feature $\mathbf{v}^k$ is an exemplar, based on the positive responsibilities sent to the candidate exemplar k from other features.

Finally, the exemplar for feature $\mathbf{v}^j$ is chosen as follows

$$e_j^* \leftarrow \arg\max_{\mathbf{v}^k}(r(\mathbf{v}^j, \mathbf{v}^k) + a(\mathbf{v}^k, \mathbf{v}^j)). \tag{7}$$

Figure 2 illustrates the feature grouping results for four disaster topics (with preference value set to 30 times the minimum similarity, and using the visual features described in section 3.1), where the x-axis and y-axis represent the first and second component of the features in the projected subspace using PCA. Each colored point in the plots represents one feature. All the feature points belonging to the same group are of the same color, and there is a line between the exemplar feature point and each member of the feature group. This figure demonstrates that the proposed features grouping method is capable of capturing the underlying correlation among all the features and separates them into different feature groups. Each of the feature groups potentially implies distinct contexts relating the disaster topic.

(a) Road Debris   (b) Earthquake

(c) Flood   (d) Volcano
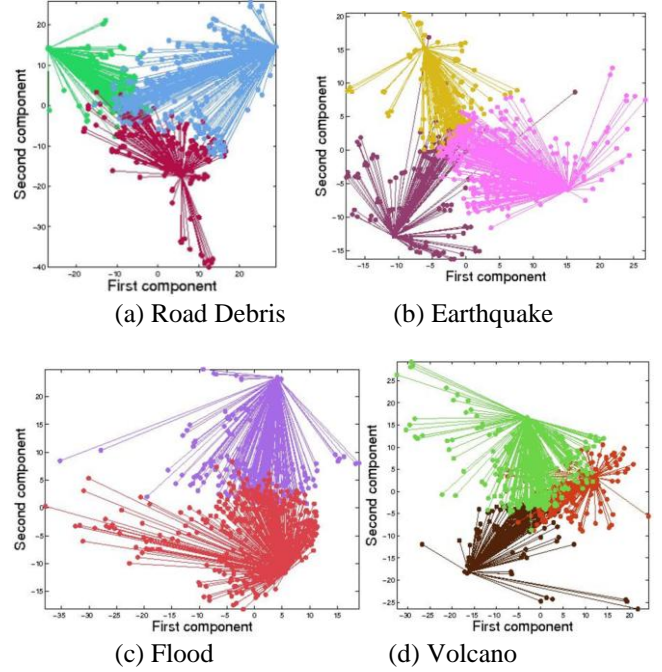
Fig. 2. Feature grouping results for the four disaster semantic concepts.

3.4. Model Fusion

The multi-model fusion procedure is depicted in Figure 3. First the feature correlation analysis and affinity propagation (FCA-AP) algorithm is applied to the original feature set, obtaining M HCFGs; then each HCFG is modeled by a series of classifiers, named A through N, generating a score array, denoted as $\left[Score(t)_g^m\right]$, where t represents each concept, g and m denote the HCFG group id and the model used for training, respectively. The score array is sorted against the training performance evaluated using MAP measurement. Only the top Q scores are kept for the final fusion. This procedure ensures the best HCFGs are selected for the fusion and so as to optimize the final retrieval performance.
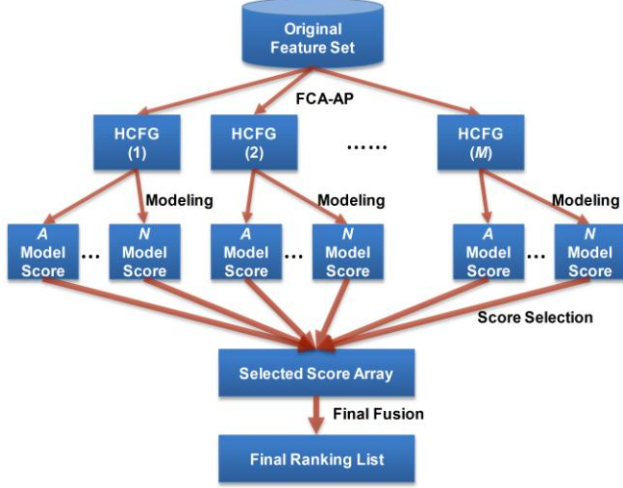
Fig. 3. Multi-model fusion procedure.

The fusion of the selected scores from multiple models are combined using the refined formula from [10] expressed as

$$Score(\boldsymbol{x})= \sum_{q=1}^{Q} \frac{\gamma_q \cdot \beta_q}{\gamma_q + \beta_q} \cdot \left(\frac{Score_q(\boldsymbol{x})}{\alpha_q} \right), \qquad (8)$$

where the parameters are explained as follows:

- α_q denotes the refined scale factor for balancing the ranking score from the qth model. It is calculated as the absolute mean score for all the training instances for that model. We refine this parameter by taking the absolute value to accommodate negative scores.
- β_q expresses the relationship between the testing score for the qth model and the target concept, which is measured based on the correlation value between the testing score interval and the related concept [10].
- γ_q represents the reliability of model q based on training performance. Specifically, it is calculated as the average precision of the qth model evaluated on the instances in the training set.

3.5. User Feedback Mechanism

One important component of our proposed system is the user feedback system based on the Markov Model Mediator (MMM) [17]. The objective is to improve the multimedia semantic retrieval performance by incorporating user interaction. The MMM mechanism is used to model the searching and retrieval process for content-based image retrieval. One distinctive characteristic of MMM model is that it carries out the searching and similarity computing process dynamically, taking into consideration not only the

image content features but also other properties of multimedia data instances such as their access frequencies and access patterns.

3.5.1 Markov Model Mediator (MMM)

MMM is a probability-based mechanism that adopts the Markov model framework and the mediator concept. The MMM mechanism models a multimedia database by a 5-tuple $\lambda = (S, F, A, B, \pi)$, where S is a set of instances called states; F is a set of distinct features of the instances; A denotes the state transition probability distribution, where each entry (i, j) indicates the relationship between instances i and j captured through the off-line training procedure; B is the feature matrix of all instances; and π is the initial state probability distribution.

The training of MMM basically involves the construction of the two statistical matrices, A and π. A sequence of user feedback characterizing access patterns and access frequencies is used to train the model parameters. Specifically, the training of the two parameters are described as follows:

Matrix A: The training of A is based on the intuition that the more frequently two images are accessed together, the more closely related they are. In order to capture the relative affinity measurements among all the instances, a matrix AF is constructed with each element $af_{i,j}$ representing the relative affinity relationship between two instances i and j as

$$af_{i,j} = \sum_{d=1}^{D} P_{i,d} \times P_{j,d} \times AC_d, \qquad (9)$$

where $P_{i,d}$ denotes the feedback pattern of instance i in time period d and AC_d represents the access frequency in that time period.

Matrix π: The preference of the initial states for user feedback can be obtained from the training data set. For any instance i, the initial state probability is defined as the fraction of the number of occurrences of instance i with respect to the total number of occurrences for all the images in the image database from the training data set.

4. EXPERIMENTAL ANALYSIS

4.1. Dataset Description

The evaluation of our proposed framework is based on a disaster dataset, which contains over 10,000 images with the associated tags and descriptions covering 11 disaster topics are crawled from Flickr, which includes both natural disasters such as "Earthquake" and "Floods", and man-

made disasters like "Road debris" and "Oil spill". Table 1 shows the composition of the data set.

Table 1. Disaster image dataset.

ID	Disaster Topic	# of Images
1	Avalanche	624
2	Drought	599
3	Earthquake	884
4	Flood	1,009
5	Ice Storm	1,078
6	Mudflow	266
7	Oil Spill	1,847
8	Volcano	800
9	Tornado	266
10	Gas Explosion	1,019
11	Road Debris	2,009
Total: 10,401		

4.2. Experiment Setup

To thoroughly evaluate the effectiveness of the proposed framework, a series of experiments are conducted. First, the significance of the feature grouping approach is analyzed by discussing the number of feature groups; second the multi-model fusing scheme is evaluated using the disaster image dataset under 3-fold cross validation; finally we compare the overall performance of our fusion framework with the other modeling methods.

The evaluation criteria is the well-known Mean Average Precision (MAP) widely used in the information retrieval society, which is calculated as

$$MAP(T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{1}{n_i} \sum_{j=1}^{n_i} Precision(R_{ij}), \qquad (10)$$

where $|T|$ is the total number of queried concepts, and R_{ij} is the top-j ranked results for concept i.

4.3. Evaluation on the Disaster Image Dataset

4.3.1. Analysis on number of feature groups

The AP algorithm has a heuristic parameter P, called preference, which indicates the preference that an instance is chosen as an exemplar. The work in [15] shows that the number of groups monotonically increases with P

polynomially. The value of P is empirical set to -10 in the following experiments. Figure 4 shows the number of groups for each concept in each of the three folds, which range from 4 to 9. Experimental analysis shows the advantages of our proposed feature grouping method, i.e., the decomposition of features enables parallel processing, which is a very important characteristic for mobile applications. In addition, the feature grouping method keeps all the original information, thus avoiding potential information loss by using the previously discussed subspace analysis methods.
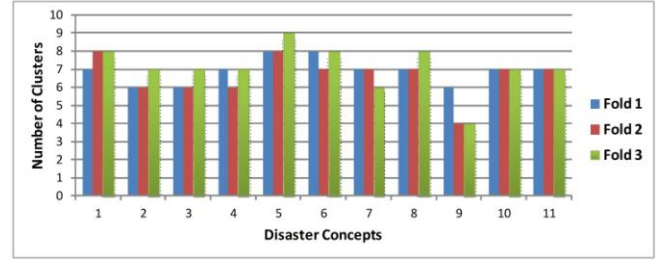
Fig. 4. Number of groups for each concept.

4.3.2. Evaluation on multi-model fusion scheme

Figure 5 shows the MAP values when selecting different number of models for multi-model fusion described in section 3.4. There are two major observations as follows: (1) the MAP values increase as more and more groups (models) being selected for final fusion, which is intuitive because we add more valuable information for final decision; (2) The performance stabilizes when the number of models reaches certain point, in this case, top 6 groups, which indicates that we capture the most important information for final decision with a subset of the original features. It also means that our framework can automatically filter out the irrelevant information, which is not useful for the final decision-making. We further compare the final fusion results with the average performance for all the groups using different modeling methods i.e., LibSVM [18] and Multiple Correspondence Analysis (MCA) [10], as shown in Figure 6. The results demonstrate that the fused scheme outperforms single models by taking advantages of both models. It is worth noting that our framework is adaptive to multiple training models and is able to optimize the overall performance by fusing the most promising HCFGs from different models.
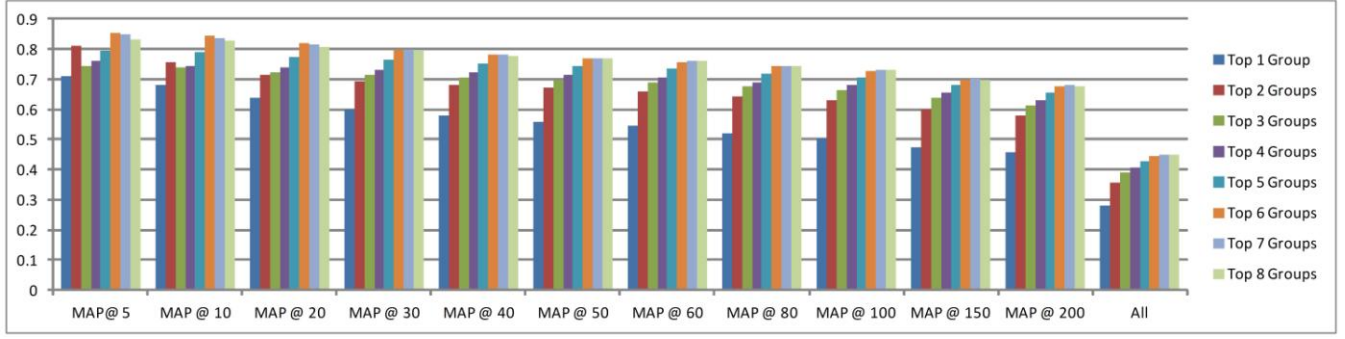
Fig. 5. MAP values for different number of hidden coherent feature groups (HCFG).
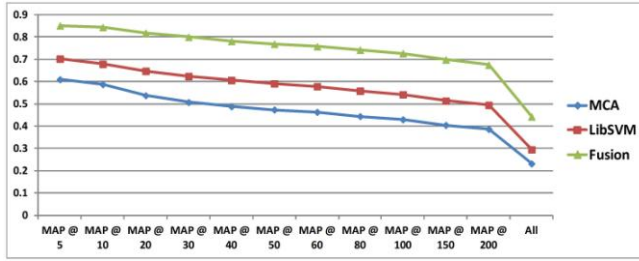
Fig. 6. MAP values for different modeling methods and the proposed fusion scheme.

4.4. Multimedia Retrieval via Mobile Devices

4.4.1. Application architecture

An iPad application has been developed based on our proposed framework, which follows a three-tiered architecture as depicted in Figure 7. The production database is implemented as a PostgreSQL database, which stores all the processing results of the back-end system. The API to access the database and perform complicated data queries is done through the REST API, implemented as a Java Tomcat servlet (using the Restlet framework). Upon these two layers, the Client is implemented in iOS, specifically for Apple's iPad devices.
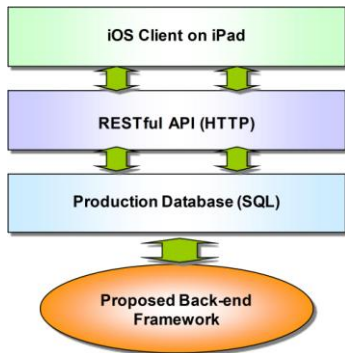
Fig. 7. iPad application architecture.

4.4.2. Application interface

Figure 8 shows two search results with the developed application tested on the disaster image dataset. It allows a user to search for multimedia content based on one or more keywords. Upon submission of the search terms in the mobile application, these terms are sent to our back-end server where a query is generated dynamically to search our database for images that match the given keywords. Relevant information about each image is then sent to the mobile application. This information includes the keywords (concept names as well as their synonyms) associated with the image, its subject, location, description, and URL for retrieving the image for display. The mobile application is designed with a built-in image cache so that when an image is requested to be displayed multiple times, the cache is checked first, before the call to retrieve the image from the servers; this reduces overhead when retrieving and displaying an image multiple times.

In addition to simply search based on keywords, the system also allows the user to specify a date range for the search. This enables the user to search for images that are relevant to a specific disaster event. Once the user has submitted a search, the mobile application groups all the images based on location and displays the on the map to the left. Selecting one of the push pins on the map filters the list of images, showing only the images at the specific location. Moreover, users are allowed to give feedback to the retrieval results with the following three options, (1) thumbs up: system made a correct match, but some image(s) is/are more relevant than others; (2) thumbs down: system made a correct match, but some image(s) is/are less relevant than others; (3) flag: image is completely inappropriate, and should be hidden from all future image lists. Those user feedbacks are collected and processed by the MMM component to further refine the retrieval results.
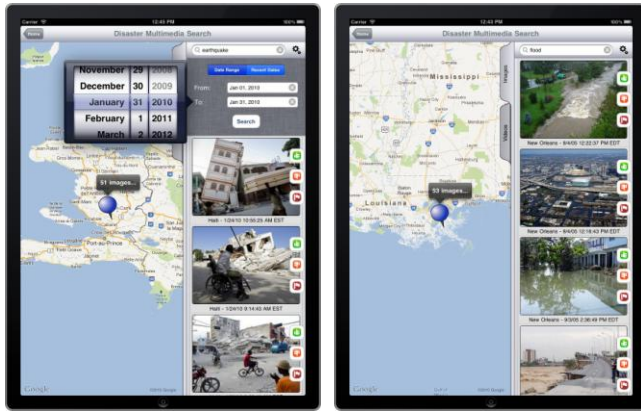
(a)   (b)

Fig. 8. Application interface: (a) search results using keyword "earthquake"; (b) search results using keyword "flood".

5. CONCLUSION

In this paper, we present a novel correlation-based feature analysis method to derive Hidden Coherent Feature Groups (HCFGs) for multimedia semantic retrieval on mobile devices. The proposed framework explores the mutual information from multiple modalities by performing correlation analysis for each feature pair and separating the original feature set into different HCFGs by using the affinity propagation algorithm at the feature level. Then a novel fusion scheme is proposed to fuse the testing scores from selected HCFGs to obtain optimal performance. The experimental analysis and results demonstrate the effectiveness of the proposed framework. An iPad application is developed based on our proposed framework with a user feedback processing system to refine the retrieval results.

6. ACKNOWLEDGMENT

7. REFERENCES

[1] S. Cherry, "Cloud computing drives mobile data growth," IEEE Spectrum, vol. 46, no. 10, pp. 52, 2009.

[2] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 2, no. 1, pp. 1–19, 2006.

[3] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: a survey," Multimedia Systems, vol. 16, no. 6, pp. 345–379, 2010.

[4] J. Yu and Q. Tian, "Learning image manifolds by semantic subspace projection," in Proceedings of the 14th annual ACM international conference on Multimedia, 2006, pp. 297–306.

[5] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih, "Image indexing using color correlograms," in IEEE International Conference on Computer Vision and Pattern Recognition, 1997, pp. 762–768.

[6] P. Smaragdis and M. Casey, "Audio/visual independent components," in Proc. ICA, 2003, pp. 709–714.

[7] [7] A. V. Nefian, L. Liang, X. Pi, X. Liu, and K. Murphy, "Dynamic bayesian networks for audio-visual speech recognition," EURASIP Journal on Advances in Signal Processing, vol. 1900, no. 11, pp. 1274–1288, 2002.

[8] D. Liu, S. Hua, Z. Ou, and J. Zhang, "Ir and visible-light face recognition using canonical correlation analysis," Journal of Computational Information Systems, vol. 5, no. 1, pp. 291–297, 2009.

[9] T. Westerveld, A. P. De Vries, A. Van Ballegooij, F. de Jong, and D. Hiemstra, "A probabilistic multimedia retrieval model and its evaluation," EURASIP Journal on Applied Signal Processing, vol. 2003, pp. 186–198, 2003.

[10] C. Chen, Q. Zhu, L. Lin, and M.-L. Shyu, "Web media semantic concept retrieval via tag removal and model fusion," ACM Transactions on Intelligent Systems and Technology, in press.

[11] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski, "Multisensor image segmentation using dempster-shafer fusion in markov fields context," IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 8, pp. 1789–1798, 2001.

[12] M. E. Sargin, Y. Yemez, E. Erzin, and A. M. Tekalp, "Audiovisual synchronization and fusion using canonical correlation analysis," IEEE Transactions on Multimedia, vol. 9, no. 7, pp. 1396–1403, 2007.

[13] V.R. Chandrasekhar, S.S. Tsai, G. Takacs, D.M. Chen, N.M. Cheung, Y. Reznik, R. Vedantham, R. Grzeszczuk, and B. Girod, "Low latency image retrieval with progressive transmission of chog descriptors," in Proceedings of the 2010 ACM multimedia workshop on Mobile cloud media computing, 2010, pp. 41–46.

[14] M. Doller, R. Tous, M. Gruhne, K. Yoon, M. Sano, and I.S. Burnett, "The mpeg query format: Unifying access to multimedia retrieval systems," MultiMedia, IEEE, vol. 15, no. 4, pp. 82–95, 2008.

[15] Y. Yang and S.-C. Chen, "Disaster image filtering and summarization based on multi-layered affinity propagation," in IEEE International Symposium on Multimedia, 2012, pp. 100–103.

[16] B.J. Frey and D. Dueck, "Clustering by passing messages between data points," Science, vol. 315, no. 5814, pp. 972–976, 2007.

[17] M.-L. Shyu, S.-C. Chen, M. Chen, and C. Zhang, "Affinity relation discovery in image database clustering and content-based retrieval," in Proceedings of ACM Multimedia, 2004, pp. 372–375.

[18] C.C. Chang and C.J. Lin, "Libsvm: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 27, 2011.

Yimin Yang is a Ph.D. candidate at the School of Computing and Information Sciences, Florida International University (FIU), Miami. She received her M.S. degree in Computer Science from FIU in 2012. Her research interests include multimedia data mining, multimedia systems, image and video processing. Yimin can be reached at yyang010@cs.fiu.edu.

Hsin-Yu Ha is a Ph.D. candidate at the School of Computing and Information Sciences, Florida International University (FIU), Miami. She received her M.S. degree in Computer Science from FIU in 2012. Her major research interest is multimedia data mining. Hsin-Yu can be reached at hha001@cs.fiu.edu.

Fausto C. Fleites is a Ph.D. candidate at the School of Computing and Information Sciences, Florida International University (FIU), Miami. He received his M.S. degree in Computer Science from FIU in Spring 2012. His research interests include multimedia indexing, multimedia data mining, and distributed systems. Fausto can be reached at fflei001@cs.fiu.edu.

Shu-Ching Chen has been a Full Professor at the School of Computing and Information Sciences, Florida International University since August 2009. He received his Ph.D. degree in Electrical and Computer Engineering in 1998 from Purdue University, West Lafayette, Indiana. His research interests include distributed multimedia database management systems and multimedia data mining. Dr. Chen can be reached at chens@cs.fiu.edu.

CORRELATION-BASED FEATURE ANALYSIS AND MULTI-MODALITY FUSION FRAMEWORK FOR MULTIMEDIA SEMANTIC RETRIEVAL

Hsin-Yu Ha, Yimin Yang, Fausto C. Fleites, Shu-Ching Chen

School of Computing and Information Sciences,
Florida International University,
1200 SW 8th Street,
Miami, FL 33199, USA
{hha001,yyang010,fflei001,chens}@cs.fiu.edu

ABSTRACT

In this paper, we propose a Correlation based Feature Analysis (CFA) and Multi-Modality Fusion (CFA-MMF) framework for multimedia semantic concept retrieval. The CFA method is able to reduce the feature space and capture the correlation between features, separating the feature set into different feature groups, called Hidden Coherent Feature Groups (HCFGs), based on Maximum Spanning Tree (MaxST) algorithm. A correlation matrix is built upon feature pair correlations, and then a MaxST is constructed based on the correlation matrix. By performing a graph cut procedure on the MaxST, a set of feature groups are obtained, where the intra-group correlation is maximized and the inter-group correlation is minimized. Finally, one classifier is trained for each of the feature groups, and the generated scores from different classifiers are fused for the final retrieval. The proposed framework is effective because it reduces the dimensionality of the feature space. The experimental results on the NUSWIDE-Lite data set demonstrate the effectiveness of the proposed CFA-MMF framework.

Index Terms— multimedia semantic retrieval, feature correlation, maximum spanning tree, multi-modality, fusion

1. INTRODUCTION

Nowadays, the propagation of multimedia data is increasing drastically as a result of advance technology and how people tend to share their life through pictures and videos on a daily basis. This fact has been drawing multimedia research society's attention to implement a comprehensive framework to effectively retrieve a variety of semantic concepts from all kinds of multimedia data, such as images, videos, text, etc.

In order to bridge the semantic gap between the low-level features extracted from multimedia data and their high-level semantic meaning, there are two major challenges researchers have to cope with. First of all, effectively analyzing high-dimensional low-level features in different formats plays an important role in building a good semantic retrieval framework, especially when it comes to scalability issues. To address this issue, researchers usually adopt linear transformations that project the low-level features into a low-dimensional space, reducing the dimensionality of the data as well as the noise contained in the original feature representation. Specifically, statistical measures such as principle component analysis (PCA) and Singular Value Decomposition (SVD) [1, 2] are widely integrated with genetic algorithms (GA) [3, 4] in feature extraction and feature selection to carry out a dimension reduction process. However, projecting all the low-level features into a relatively small universal feature space may be easily affected by outliers and thus valuable information can be lost during dimensionality reduction. Secondly, the correlation between various features and the dependency between modalities should be thoroughly explored since the implication among features would definitely help with semantic retrieval. For example, the tag "sky" implies the color "blue" for the semantic concept "outdoor", which is considered as a "hidden" correlation between features.

Based on the above-mentioned challenges, we propose a Correlation-based Feature Analysis and Multi-Modality Fusion (CFA-MMF) framework for multimedia semantic retrieval. Specifically, we explore the correlations between each feature pair from multiple modalities and the feature space can be reduced by removing features with low correlation toward other features and features with zero standard deviation in the positive instance. Then Maximum Spanning Tree-based Feature Graph Cut (MaxST-FGC) algorithm is used to extract Hidden Coherent Feature Groups (HCFGs), where the intra-group correlation is maximized and the inter-group correlation is minimized. Then one classifier is trained for each of the feature groups, and the generated scores from different classifiers are fused for the final retrieval.

The main contributions of this paper are as follows:

- Propose a correlation-based feature analysis method

that analyzes pair-wise feature relationships and extracts coherent feature groups (i.e., HCFGs) via a graph cut on a correlation-based feature graph.

- Develop a framework that integrates early fusion and late fusion by decomposing all features from multiple modalities into groups, and later fusing the multiple uncorrelated models at the decision level.

The rest of paper is organized as follows. Section 2 introduces the state of the art in multimedia semantic retrieval. Section 3 presents the details of the proposed CFA-MMF framework. Section 4 discusses the experimental results, and section 5 finalizes the paper.

2. RELATED WORK

The related works in the area of multimedia semantic retrieval can be roughly summarized into (1) uni-modality based approaches and (2) multi-modality based approaches, from an information-fusion point of view. In the first category, single modality features, i.e., visual, textual, etc., are extracted for multimedia semantic retrieval. However, due to the versatile characteristics of multimedia data, uni-modality representation cannot properly convey the rich information embedded in the multimedia content. Therefore, many works have been presented for effective fusion of multi-modality features. One common way of multi-modality information fusion is to apply statistical analysis methods to the direct concatenation of features from multiple modalities at feature level. For example, Smaragdis et al. [1] adopt PCA and ICA to obtain the maximally independent audio-video subspaces from the audio-visual concatenated features. Huanzhang et al. [5] apply both PCA and Adaboost as feature selection methodologies to select useful region-based features in object detection. Kusuma et al. [6] exploit the dependency between 2D and 3D facial images and recombined the features from different modalities with the usage of PCA in the first phase. In the second phase, Fishers Linear Discriminant (FLD) was applied to perform another recombination transform into more discriminating data.

Besides analyzing the feature level correlation, the correlation among different models and model confidence toward extracting semantic concepts have also been studied. Liu et al. [7] propose a method called Selective Weighted Late Fusion (SWLF) which used the results trained from a binary classifier to weight the corresponding features in testing data set. Chen et al. [8] propose a fusion strategy to combine ranking scores from both tag-based and content-based models, where the adjustment, reliability, and correlation of ranking scores from different models are all considered. Hofmann et al. [9] propose a fusion method based on probabilistic kernel density estimation to fuse the output of part-based object detectors from multiple camera views in person detection.
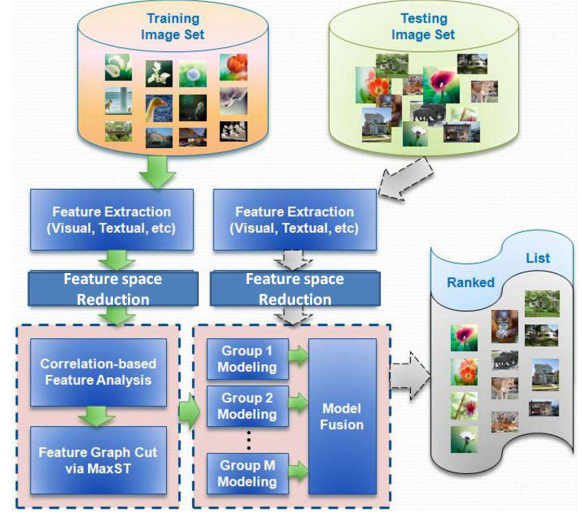
Fig. 1. Correlation based Feature Analysis (CFA) and Multi-Modality Fusion (CFA-MMF) framework.

Despite all the reported advantages reported in existing multi-modality fusion frameworks for multimedia semantic retrieval, they still suffer from the problem of information loss by transformation between different feature spaces and cannot fully utilize the complementary and mutual information among features from multiple modalities. To tackle this problem, we propose the CFA-MMF framework that discovers the HCFGs crossing multiple modalities based on feature correlation analysis and enhance the predictive power of the final fused model.

3. PROPOSED FRAMEWORK

The proposed semantic retrieval framework is depicted in Fig. 1. It builds the retrieval model following a five-step process that consists of (a) feature extraction, (b) pre-processing, (c) correlation-based feature analysis and feature graph cut via MaxST algorithm, (d) model training, and (e) model fusion. Firstly, in the first two steps, the system extracts multi-modality features (e.g., visual, textual, etc.) from the training data and performs pre-processing to normalize the features and remove those with relatively low variance. Secondly, in the correlation-based feature analysis and feature graph cut step, the system computes a feature similarity matrix based on correlation coefficients for all pairs of retained features and applies the MaxST algorithm to analyze the original feature set and obtain HCFGs that exhibit low inter-group correlation and high intra-group correlation. Subsequently, the model training step builds a classification model for each discovered feature group. Finally, the model fusion step combines the individual models using the proposed multi-model fusion strategies. Such a partition of the feature set into HCFGs aims at "untapping" hidden feature groups that

will enhance the predictive power of the fused model. When a query is issued to the system, the system performs feature extraction and pre-processing and groups the features into the same HCFGs identified in the training phase. The HCFGs are then fed to the trained models obtained during the model training step. The generated testing scores are afterward fused and ranked.

3.1. Correlation based Feature Analysis (CFA)

Though the features are extracted from diverse media streams, they may be correlated. For example, in a video shot, the visual frames show a dog barking while the audio channel also records the sound. If the two sources could be effectively integrated in the retrieval system, this kind of multi-modality features may be more discriminant than the single modality feature. On the other hand, the independence among the modalities is also important as it may provide additional cues that help for the retrieval. When fusing multiple modalities, this correlation and independence may equally provide valuable insight based on a particular scenario or context. This section describes the proposed correlation-based analysis method that explores the interrelationship among feature from multiple modalities and constructs the basis for feature graph cut (elaborated in section 3.2).

Suppose a given dataset is denoted by $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^L$ represents each instance in the dataset (N and L are the number of instances and the feature set cardinality, respectively). Then the feature set F is represented as $\{\mathbf{f}^l\}_{l=1}^{L}$, where $\mathbf{f}$ is a feature representation of all the instances in $\mathbf{X}$.

It is worth noting that we do not differentiate the features from multiple modalities and treat each individual feature equally at this point. Let $(\mathbf{f}^j, \mathbf{f}^k)$ $(j, k = 1, 2, \cdots, L)$ be a feature pair, then the correlation coefficient between them can be calculated as follows

$$C_{\mathbf{f}^j,\mathbf{f}^k} = \frac{\sum_{i=1}^{N}(\mathbf{f}_i^j - \overline{\mathbf{f}^j})(\mathbf{f}_i^k - \overline{\mathbf{f}^k})}{\sqrt{\sum_{i=1}^{N}(\mathbf{f}_i^j - \overline{\mathbf{f}^j})^2}\sqrt{\sum_{i=1}^{N}(\mathbf{f}_i^k - \overline{\mathbf{f}^k})^2}}, \quad (1)$$

where $\overline{\mathbf{f}^j}$ and $\overline{\mathbf{f}^k}$ are the mean values of $\mathbf{f}^j$ and $\mathbf{f}^k$, respectively. The initial feature correlation matrix $\mathbf{C}$ is constructed as

$$\begin{bmatrix} C_{\mathbf{f}^1,\mathbf{f}^1} & C_{\mathbf{f}^1,\mathbf{f}^2} & \cdots & C_{\mathbf{f}^1,\mathbf{f}^L} \\ C_{\mathbf{f}^2,\mathbf{f}^1} & C_{\mathbf{f}^2,\mathbf{f}^2} & \cdots & C_{\mathbf{f}^2,\mathbf{f}^L} \\ \vdots & \vdots & \ddots & \vdots \\ C_{\mathbf{f}^L,\mathbf{f}^1} & C_{\mathbf{f}^L,\mathbf{f}^2} & \cdots & C_{\mathbf{f}^L,\mathbf{f}^L} \end{bmatrix}$$

Each element in the matrix presents the correlation coefficient between each feature pair, and the matrix is symmetric, i.e., $C_{\mathbf{f}^j,\mathbf{f}^k}$ equals $C_{\mathbf{f}^k,\mathbf{f}^j}$

The above correlation coefficients analysis method is based on the calculation of Pearson product-moment correlation coefficient, which assumes normally-distributed data and the linear relationship between feature variables. However, this is not always the case. In order to take into account the situation where the feature variables follow a non-linear relationship, we propose another correlation estimation method based on the Spearman's rank correlation coefficients, which use the ranks of the observations instead of their values and are calculated as

$$C_{\mathbf{r}^j,\mathbf{r}^k} = \frac{\sum_{i=1}^{N}(\mathbf{r}_i^j - \overline{\mathbf{r}^j})(\mathbf{r}_i^k - \overline{\mathbf{r}^k})}{\sqrt{\sum_{i=1}^{N}(\mathbf{r}_i^j - \overline{\mathbf{r}^j})^2}\sqrt{\sum_{i=1}^{N}(\mathbf{r}_i^k - \overline{\mathbf{r}^k})^2}}, \quad (2)$$

where $\mathbf{r}$ is the rank representation of the feature variable $\mathbf{f}$.

Finally, we applied the following rules to regulate the correlation matrix:

- Only the feature value with non-zero standard deviation from positive instances were considered in obtaining correlation coefficients toward other features.

- The self correlation coefficients are set to zero (i.e., $C_{\mathbf{f}^j,\mathbf{f}^j} = 0$) for the purpose of later feature graph operation.

- The negative correlation coefficients are replaced by their absolute values. This operation is necessary because we are more concerned with how much two features are correlated than how far they depart from each other. In other words, it is not relevant in which direction two features are correlated.

- All the correlation coefficients are calculated based on positive instances. Therefore the correlation matrix is concept specific. This rule implies the advantage of our feature analysis approach by decreasing the total number of training instances and reducing computation complexity, which is a considerable merit over the other statistical-based methods such as PCA, ICA etc.

By using the proposed correlation-based feature analysis method, we are able to capture the correlations among feature variables from multiple multimedia modalities at different granularity. For example, either one the feature ($\mathbf{f}^j$ or $\mathbf{f}^k$) in the correlation coefficient $C_{\mathbf{f}^j,\mathbf{f}^k}$ may be color feature or texture feature from the visual modality, or the tag feature from textual modality, or even the object location feature, which can be considered as a middle level feature based on visual characteristics.

3.2. Feature Graph Cut via Maximum Spanning Tree

To better cope with feature correlation from different modalities, a graph-based approach Maximum Spanning Tree (MaxST) was leveraged in our framework due to its capability of detecting clusters with irregular boundaries. Let $G(F, E)$ be the general notation of a feature graph constructed based

on the feature correlation matrix, where F is the feature set (section 3.1) and E represents the set of feature correlation coefficients $\left\{C_{\mathbf{f}^j, \mathbf{f}^k}\right\}_{j,k=1}^{L}, j < k$. Prim's method [10] was used for constructing a MaxST over the features under absolute correlation value [11]. Unlike other research works using minimum spanning tree to cluster data instance, we constructed an acyclic subgraph that has maximum sum of edge weights and spans over all the vertices. Next, all the edges included in the MaxST are sorted in ascending order. Finally, M feature groups which have high intra-group correlation and low inter-group correlation are obtained by removing $M - 1$ smallest edges from the MaxST.

3.3. Model Fusion

The final fusion of the scores from multiple models are based on the refined fusion scheme ARC [8] expressed as

$$Score(I) = \sum_{m=1}^{M} \frac{\xi_m \cdot \theta_m}{\xi_m + \theta_m} \cdot \left(\frac{Score_m(I)}{\omega_m}\right), \quad (3)$$

where M is the number of models, and ξ_m represents the reliability of model m based on training performance. Specifically, it is calculated as the average precision of the mth model evaluated on the instances in the training set; θ denotes the relationship between the testing score for the mth model and the target concept, which is measured based on the correlation value between the testing score interval and the related concept [8]; ω_m is a scale factor to balance the ranking score for the mth model, which is refined in this paper by using the absolute mean score for all the training instances.

4. EXPERIMENTAL ANALYSIS

In this section, the performance of the proposed CFA-MMF framework is evaluated based on the NUS-WIDE-Lite dataset [12], which includes 55,615 images as well as the associated tags crawled from the Flickr website, with 27,807 for training and 27,808 for testing. The dataset provides the ground truth for 79 concepts and several low-level features commonly used for evaluation such as 64-dimensional color histogram and 128-dimensional wavelet texture, which were also used in this paper. In addition, we also utilize the textual features extracted using the method proposed in [8].

4.1. Experiment Setup

To elaborately evaluate the effectiveness of the presented CFA-MMF framework from different perspectives, we conduct two sets of experiments. First, the CFA and MaxST-FGC algorithms are tested to show the better performance of our proposed feature analysis mechanism against the original flat concatenation of multi-modality features. Second, the overall framework is evaluated to demonstrate the

superiority of our proposed approach over the other existing multimedia semantic retrieval works. In both experiments, the correlation-based feature analysis is based on Spearman's rank correlation coefficients and M is selected as 2, i.e., we extract 2 HCFGs from the original feature set. The LibSVM modeling [13] method is adopted in this paper for evaluation because it has been proven to be effective for various multimedia analysis tasks in previous works. It can be easily replaced by any other model training approach. Finally, the evaluation criteria is the well-known Mean Average Precision (MAP) widely used in the information retrieval society.

4.2. Evaluation of CFA and MaxST-FGC algorithms

Fig. 2 shows the number of features after applying the proposed framework (denoted as CFA-MMF) and the original feature set including both visual and textual features (with noisy tag removed [8]). With the proposed correlation-based feature analysis method, only features with high correlation toward other features will be used in classification process. Therefore, there are 11 concepts which had feature dimensionality drop down more than 80% as shown in Fig. 3. As shown in the figure, our framework greatly reduced the dimensionality of the feature space (enhancing the computational performance) and eliminated redundant information.

4.3. Evaluation of CFA-MMF Framework

We compare the results of our proposed framework (CFA-MMF) with other research studies investigating semantic retrieval on the NUS-WIDE-LITE data set using all 79 concepts. These related works demonstrated their performance using K-nearest neighbor (KNN) model [14], LibSVM model [15], linear neighborhood propagation (LNP) [16], entropic graph semi-supervised classification (EGSSC) [17], sparse graph-based semi-supervised learning (SGSSL) [18], large-scale multi-label propagation (LSMP) [19], and three retrieval frameworks, i.e. SVD combined with minimum fusion (SVD+MIN), SVD combined with super kernel fusion (SVD+SKF), and multiple correspondence analysis-based tag removal algorithm (MCA-TR+ARC) constructed from [8].

In our framework, both visual features and image annotated tags are considered as discriminant features to overcome the semantic gap problem . In addition, we apply the MCA-TR method to remove noisy tag information using MCA [8]. Each tag feature was assigned a feature weight and the threshold with highest MAP in the training dataset was set up to remove useless tag features. Other algorithms have their parameters set up which were already proved to be best tuned in [19, 8].

The MAP value of our proposed framework against other above-mentioned frameworks is shown in Fig. 4. and it can
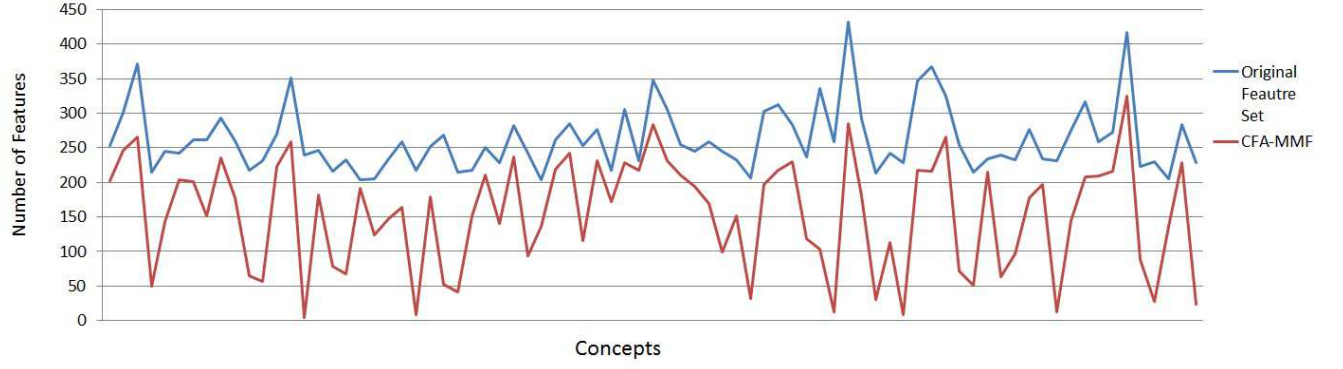
Fig. 2. Dimensionality Reduction of CFA-MMF Over Original Feature Set

Concept	Reduction Rate	Concept	Reduction Rate	Concept	Reduction Rate	Concept	Reduction Rate	Concept	Reduction Rate	Concept	Reduction Rate	Concept	Reduction Rate
airport	-20.16%	flags	**-96.31%**	protest	-59.59%	*tower*	-35.87%	clouds	-26.21%	*military*	-16.61%	surf	-71.65%
animal	-18.27%	flowers	-28.97%	*railroad*	-35.19%	town	-15.81%	computer	**-97.92%**	moon	-20.74%	sunset	-18.46%
beach	-28.49%	food	**-80.22%**	rainbow	**-84.54%**	toy	**-94.37%**	coral	-26.02%	mountain	-25.25%	tiger	-58.80%
bear	-76.74%	fox	**-80.47%**	reflection	-34.77%	*train*	-47.64%	cow	-63.89%	nighttime	-6.06%	zebra	**-89.91%**
birds	-41.63%	frost	-30.28%	*road*	-30.45%	tree	-34.38%	dog	-70.82%	ocean	-18.68%	sports	-53.31%
boats	-15.70%	garden	-16.00%	rocks	-19.01%	valley	-18.99%	*earthquake*	-5.88%	person	-24.51%	leaf	-54.15%
book	-23.28%	glacier	-38.43%	running	-50.21%	*vehicle*	-20.59%	elk	-39.51%	*plane*	-17.65%	cityscape	-17.41%
bridge	-42.37%	grass	-15.96%	*sand*	-69.05%	water	-22.06%	*fire*	-37.18%	plants	-20.82%		
buildings	-19.45%	harbor	-61.16%	sign	**-94.96%**	waterfall	-60.54%	fish	-36.43%	*police*	-34.75%		
cars	-31.54%	horses	-32.84%	sky	-34.03%	wedding	**-87.83%**	swimmers	-76.28%	*statue*	**-96.05%**		
castle	-70.18%	house	-16.09%	*snow*	-38.14%	whales	-33.17%	tattoo	-8.12%	*street*	-37.28%		
cat	-75.76%	lake	-15.09%	soccer	**-85.51%**	window	-19.72%	*temple*	-73.75%	sun	-41.14%		

Fig. 3. Percentage Change in Dimensionality Reduction of CFA-MMF over Original Feature Set

be easily distinguished with at least 4% and at most 30% higher MAP values. Compared with other algorithms, the improvement in performance can be explained as follows. We take advantage of copious information provided along with the image data, which includes features from multiple modalities, and explore the correlation among different modalities to extract a reduced feature set that filters out irrelevant information and identifies feature groups that better fuse information from different modalities.

5. CONCLUSION

In this paper, we have presented a novel correlation-based feature analysis and multi-modality fusion framework for multimedia semantic retrieval. The proposed framework explores the mutual information from multiple modalities by performing correlation analysis for each feature pair and reducing the original feature space. Consequently, the original feature set was separated into HCFGs by using the maximum spanning tree-based feature graph cut algorithm at the feature level. Then a refined multi-modality strategy is employed to combine the testing scores from different training model to obtain optimal performance. The experimental analysis and results demonstrate the effectiveness of the proposed framework. In the future,
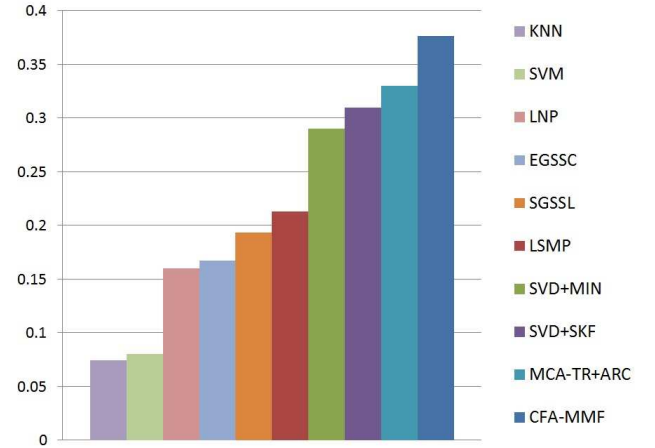
Fig. 4. MAP values of all 79 concepts of the proposed framework and other works on the NUS-WIDE-LITE dataset.

we would explore more sophisticated correlation analysis method for analyzing versatile feature types and design an adaptive framework to separate the features into multiple feature groups instead of two.

Acknowledgment

6. REFERENCES

[1] P. Smaragdis and M. Casey, "Audio/visual independent components," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA)*, 2003, pp. 709–714.

[2] F. Xiao, M. Zhou, and G. Geng, "Linear transformation technology for image feature drop dimension," in *2011 Fourth International Symposium on Knowledge Acquisition and Modeling (KAM)*. IEEE, 2011, pp. 331–333.

[3] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowledge-Based Systems*, vol. 24, no. 7, pp. 1024–1032, 2011.

[4] I. Ahmad, A. Abdullah, A. Alghamdi, M. Hussain, and K. Nafjan, "Features subset selection for network intrusion detection mechanism using genetic eigen vectors," in *Proceedings of 2011 International Conference on Telecommunication Technology and Applications (ICTTA 2011)*, 2011, pp. 75–79.

[5] H. Fu, A. Pujol, E. Dellandrea, and L. Chen, "Visual object categorization based on the fusion of region and local features," 2010.

[6] G. Kusuma, C.-S. Chua, and H. Toh, "Recombination of 2d and 3d images for multimodal 2d+ 3d face recognition," in *2010 Fourth Pacific-Rim Symposium on Image and Video Technology (PSIVT)*. IEEE, 2010, pp. 76–81.

[7] N. Liu, E. Dellandrea, C. Zhu, C.-E Bichot, and L. Chen, "A selective weighted late fusion for visual concept recognition," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 426–435.

[8] C. Chen, Q. Zhu, L. Lin, and M.-L. Shyu, "Web media semantic concept retrieval via tag removal and model fusion," *ACM Transactions on Intelligent Systems and Technology (TIST)*.

[9] M. Hofmann, M. Kiechle, and G. Rigoll, "Late fusion for person detection in camera networks," in *2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2011, pp. 41–46.

[10] R.C. Prim, "Shortest connection networks and some generalizations," *Bell system technical journal*, vol. 36, no. 6, pp. 1389–1401, 1957.

[11] P.K. Agarwal, J. Matoušek, and S. Suri, "Farthest neighbors, maximum spanning trees and related problems in higher dimensions," *Computational Geometry*, vol. 1, no. 4, pp. 189–201, 1992.

[12] T.S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "Nus-wide: a real-world web image database from national university of singapore," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. ACM, 2009, p. 48.

[13] C.C. Chang and C.J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

[14] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern classification*, Pattern Classification and Scene Analysis: Pattern Classification. Wiley, 2001.

[15] I.H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2005.

[16] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 985–992.

[17] A. Subramanya and J. Bilmes, "Entropic graph regularization in non-parametric semi-supervised classification," in *Proceedings of Neural Information Processing Society (NIPS)*, 2009, pp. 1803–1811.

[18] J. Tang, S. Yan, R. Hong, G.J. Qi, and T.S. Chua, "Inferring semantic concepts from community-contributed images and noisy tags," in *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 2009, pp. 223–232.

[19] X. Chen, Y. Mu, S. Yan, and T.S. Chua, "Efficient large-scale image annotation by probabilistic collaborative multi-label propagation," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 35–44.

Data Mining Meets the Needs of Disaster Information Management

Li Zheng, Chao Shen, Liang Tang, Chunqiu Zeng, Tao Li, Steve Luis, Shu-Ching Chen

School of Computing and Information Sciences, Florida International University

11200 S.W. 8th Street, Miami, Florida, 33199, U.S.A.

{lzhen001, cshen001, ltang002, czeng001, taoli, luiss, chens}@cs.fiu.edu

ABSTRACT

Techniques to efficiently discover, collect, organize, search and disseminate real-time disaster information have become national priorities for efficient crisis management and disaster recovery tasks. We develop techniques to facilitate information sharing and collaboration between both private and public sector participants for major disaster recovery planning and management. We design and implement two parallel systems: a web based prototype of a **Business Continuity Information Network (BCiN)** system and an **All-Hazard Disaster Situation Browser (ADSB)** system that run on mobile devices. Data mining and information retrieval techniques help impacted communities better understand the current disaster situation and how the community is recovering. Specifically, information extraction integrates the input data from different sources, report summarization techniques generate brief reviews from a large collection of reports at different granularities; probabilistic models support dynamically generating query forms and information dashboard based on user's feedback; and community generation and user recommendation techniques are adapted to help users identify potential contacts for report sharing and community organization. User studies with over 200 participants from EOC personnel and company participants demonstrate that our systems are very useful for them to gain insights of the disaster situation and make responses.

Index terms

Data Mining, Disaster Information Management, Hierarchical Summarization, Dynamic Query Form, User Recommendation

1. INTRODUCTION

Business closures caused by disasters can cause millions of dollars in lost productivity and revenue. A study in Contingency Planning and Management shows that 40% of companies that were shut down by a disaster for three days failed within 36 months. Thin margins and lack of a well-designed and regularly tested disaster plan make companies, particularly small businesses, especially vulnerable [1]. We believe that the solution to better disaster planning and recovery is one where the public and private sectors work together to apply world-class computing tools to deliver the right information to the right people at the right time facilitating the work of those feverishly restoring a community's sense of normalcy. While the improved predictive atmospheric and hydrological models and higher quality of building materials and building codes are being developed, more research is also necessary for how to collect, manage, find, and present disaster information in the context of disaster management phases: Preparation, Response, Recovery, and Mitigation [4,33].

In the United States, Federal Emergency Management Agency (FEMA) has recognized the importance of the private sector as a partner in addressing regional disasters. The State of Florida Division of Emergency Management has created a Business and Industry Emergency Support Function designed to facilitate logistical and relief missions in affected areas. Four counties, Palm Beach, Broward, Miami-Dade and Monroe, which and include over 200,000 business interests, are developing Business Recovery Programs to help facilitate quicker business community recovery through information sharing and collaboration.

Disaster management researchers at Florida International University have collaborated with the Miami-Dade Emergency Operations Center (EOC), South Florida Emergency Management and industry partners including Wal-Mart, Office Depot, Wachovia, T-Mobile, Ryder Systems and IBM to understand how South Florida public and private sector entities manage and exchange information in a disaster situation. The efficiency of sharing and management of information plays an important role in the business recovery in a disaster [3]. Users are eager to find valuable information to help them understand the current disaster situation and recovery status. The community participants (the disaster management officials, industry representatives, and utility agents) are trying to collaborate to exchange critical information, evaluate the damage, and make a sound recovery plan. For instance, it is critical that companies receive information about their facilities, supply chain, and city infrastructure. They seek this information from media outlets like television/radio newscasts, employee reports, and speak with other companies they have a relationship with. With so many sources of information, with different levels of redundancy and accuracy, possibly generated by different varieties of reports (structured and unstructured), it is difficult for companies to quickly assimilate such data and understand their situation.

We have learned that large-scale regional disaster causes a disruption in the normal information flow and channels, and the relationships between information producer and consumer. Effective communication is critical in a crisis situation. What is not very well known is how to effectively discover, collect, organize, search and disseminate real-time disaster information.

Our study of the hurricane disaster information management domain has revealed two interesting yet crucial information management issues that present similar challenges in other disaster management domains. The first issue is that reconstructing or creating information flow becomes intractable in domains where the stability of information networks is fragile and can change frequently. On the other hand, important information networks often carry and store critical information between parties, which dominates the flow of resources and information exchanges. The consequence communication degrades once critical networks are disrupted by the disaster and people may not have alternative path for transferring information. For instance,

once power is disabled and uninterruptable power supplies fail after a hurricane, computing and networking equipment will fail unless preventative measures are taken. However, maintaining a fuel-consuming generator is not always possible.

Another issue is the large volume of disaster situational information. Reading and assimilating situational information are very time-consuming and may involve redundant information. Therefore, to quickly re-assemble or create information flow for multi-party coordination activities during disaster situations, technologies that are able to extract information of recent updates, deliver information without conflicts or irrelevance and represent information of preference are needed.

This research is mainly focused on the second issue. Research in disaster management addresses the needs and challenges of information management and decision-making in disaster situations [40,41,42]. We have developed an understanding of those needs for hurricane scenarios. The information delivery should support users' complex information needs tailored to the situation and the tasks; and the information should be synthesized from heterogeneous sources and tailored to specific contexts or tasks at hand. It should be summarized for effective delivery and immediately useful for making decision.

1.1 Related Work

The approaches and the tools used for information sharing vary based on the task and scale of the participating agencies or the types of information exploration platforms.

Commercial systems, such as WebEOC [43] and E-Teams [44] used by Emergency Management departments located in urban areas, can access multiple resources. A Disaster Management Information System developed by Department of Homeland Security (DHS) is available to county emergency management offices and participating agencies to provide an effective reports/document sharing software system. The National Emergency Management Network [45] allows local government share resources and information about disaster needs; The RESCUE Disaster Portal is a web portal for emergency management and disseminating disaster information to the public [4]; The Puerto Rico Disaster Decision Support Tool (DDST) is an Internet based tool for disaster planners, responders, and related officials at the municipal, zone, and state level for access to a variety of geo-referenced information [39].

Efforts, such as GeoVISTA [34], facilitate the information distribution process in disasters. GeoVISTA monitors tweets to form situation alerts on a map based user interface according to the geo-location associated with the tweets. Such system applies Geographic Information Sciences to scientific, social and environmental problems by analyzing geospatial data [34].

These useful situation-specific tools provide query interfaces, and GIS and visualization capabilities to simplify the users' interaction and convey relevant information. The primary goal of these systems are message routing, resource tracking and document management for the purpose of supportng situational awareness, demonstrating limited capabilities for automated aggregation, data analysis and mining [4].

However these tools do not consider how different communities interact with other businesses and county organizations. Further, these tools do not allow for the integration of real-time information. They do not provide Information Extraction (IE), Information Retrieval (IR), Information Filtering (IF) and Data

Mining (DM) techniques needed when delivering personalized situation information to different types of users.

1.2 Design Challenges

We have identified four key design challenges for disaster information sharing platforms and tools.

1. Effective techniques to capture the status information. Participants need to communicate status through many channels, including email, mailing lists, web pages, press releases, and conference calls. It is desirable to capture such status information when it is available and prevent redundant reporting. To facilitate the reuse of such materials, users should be able to update status information via unstructured documents such as plain text, Adobe PDFs, and documents. It is necessary to identify the useful information in the documents.

2. Effective and interactive information summarization methods. It is important to build a summarized view to support understanding the situation from reports. Multi-document summarization provides users with a tool to effectively extract important and related ideas of current situations. Previous text summarization techniques gave users a fixed set of sentences based on the user query. An interactive summarization interface is needed to help users navigate collected information at different granularities, and locate their target information more efficiently.

3. Intelligent information delivery techniques. Data can be collected through different channels and may belong to different categories. During disaster preparation and recovery, users do not have the time to go through the system to find the information they want. Structured information can help people make decisions by providing them with actionable and concrete information representation and exploration. However, navigating large data sets on a mobile device is particularly inefficient. An interactive tabular interface can help users filter useful information by adaptively changing query conditions and user feedback.

4. Dynamic community generation techniques. In information sharing tasks, identifying a group of recipients to which a certain type of information is conveyed can improve the efficiency of communication. Also identifying how participants interact with these communities in a disaster situation may reveal information helpful in a recovery scenario. User recommendation techniques can automatically and interactively generate potential recipients for different pieces of information. In addition, user recommendation techniques can help to dynamically organize user groups according to various information sharing tasks.

We created an information-rich service on both web-based and mobile platforms in disaster management domain to address the design challenges. In particular, to address the first challenge, we apply information extraction to automatically extract the status information from documents. To address the 2nd challenge, we apply hierarchical summarization to automatically extract the status information from a large document set and also provide a hierarchical view to help users browse information at different granularities. To address the third, we create a user interface capability called the Dynamic Dashboard to improve information quality to match user's interests, and use document summarization techniques to give users a quick access to multiple reports. Also, a Dynamic Query Form is designed to improve information exploration quality on mobile platforms. It captures users' interests by interactively allowing them to refine and update their queries. To address the fourth challenge, for community

discovery, we adopt spatial clustering techniques to track assets like facilities, or equipment, which are important to participants. The geo-location of such participants can be organized into dynamic communities, and these communities can be informed about events or activities relevant to their spatial footprints. For user recommendation, we use transactional recommendation history combined with textual content to explore the implicit relationship among users.

Thus we designed and implemented a web-based prototype of a Business Continuity Information Network (BCiN) that is able to link participating companies into a community network, provide businesses with effective and timely disaster recovery information, and facilitate collaboration and information exchange with other businesses and government agencies. We also designed and implemented an All-Hazard Disaster Situation Browser (ADSB) system that runs on Apple's mobile operating system (iOS) and iPhone and iPad mobile devices. Both systems are utilizing the data processing power of advanced information technologies for disaster planning and recovery under hurricane scenarios. They can help people discover, collect, organize, search and disseminate real-time disaster information [4,5].

This current work introduces a unified framework that systematically integrates the different techniques developed in [5] and [32]. Such a framework can be utilized when dealing with different systems or applications separately (e.g., BCiN and ADSB), and can be easily applied to other scenarios having critical information sharing and management needs

The rest of the paper is organized as follows. Section 2 describes the system architecture: information extraction techniques to create structured records, the hierarchical summarization module, the dynamic dashboard and the dynamic query form modules, and community identification and user recommendation modules. Section 3 presents two case studies of the BCiN and ADSB systems. Section 4 describes the system evaluation and data crawling strategies. The conclusion is in Section 5.

2. System Architecture

2.1 STRUCTURED INFORMATION EXTRACTION FROM REPORTS

A user interface is provided to allow for information sharing among companies and government agencies. We do not request a unified format for them to submit the reports. Instead, we use information extraction methods to integrate reports from different sources. For example, Table 1 shows an example of EOC reports.

The key information is "What was/is/will be the status of Facilities/Services/… at the time of …". From the EOC reports, we need to extract such information in the form of a triple: (entity, time, status), which reveals the status information of the entity at a certain time. In EOC reports, the entity may be a facility or public service like "Miami International Airport", "schools", "bus", and an order like "curfew". If the entity represents an order, the triple means whether the order is in effect or not at that specific time. We extract these triples through two steps: first, we extract entities and time expressions, then, we classify a pair of (service, time) to a proper category, "no relation"/ "open" / "close" / "unclear". We assume that the information of one event is described in one sentence, so we process every sentence individually to extract an event. To extract those triples, both entity and relation extraction will be performed. Sometimes, two different reports generate the same events, which have the same

extracted information, such as the same hurricane name, the same date and the same status of traffic. The repeated events will be deleted. Note that the date/time is an important attribute for every event. Two events with different date/time (at the hourly level) are treated as two different events.

Table 1. An Example of EOC Report.

Time: October 21, 2005 12:30 p.m.
Miami-Dade Emergency Operations Center is currently activated at a level II and officials and emergency managers are carefully monitoring Hurricane Wilma. Residents are urged to finalize their personal hurricane preparations. On Monday, October 24, Miami-Dade County offices, public schools, and courts will be closed. Currently, transit bus and rail service continues, including Metrobus, Metrorail and Metromover. Miami International Airport is open. However, if you have travel plans please check with your airline for flight information. Tomorrow afternoon, the American Red Cross will open hurricane evacuation centers for residents who do not feel safe in their homes or live in low-lying areas.

2.1.1 Entity Extraction

For each report, sentence segmentation is conducted, and each sentence is Part-Of-Speech-tagged [38]. To extract entities and time expressions, we manually label some news and train a linear chain Conditional Random Fields (CRF) model to tag all words, using "BIO" annotation [6,7]. A word tagged as [TYPE-B]/[TYPE-I] means it is the beginning/continuing word of the phrase of the TYPE, and the word tagged by O means it is not in any phrase. TYPE can be E for entity or T for time expression. Given sentence X, the probability that its tags are Y is:

$$p(Y|X) = \frac{1}{Z_X}\exp\left(\sum_{i,k}\lambda_k f_k(y_{i-1}, y_i, X) + \sum_{i,l}\mu_l g_l(y_{i-1}, y_i, X)\right) \quad (1)$$

where Z_X is the normalization constant that makes the probability of all state sequences sum to one; $f_k(y_{i-1}, y_i, X)$ is an arbitrary feature function over the entire observation sequence and the states at positions i and $i-1$ while $g_l(y_{i-1}, y_i, X)$ is a feature function of the states at position i and the observation sequence; λ_k and μ_l are the weights learned for the feature functions f_k and g_l, reflecting the confidenceof feature functions by maximum likelihood procedure. The most probable labels can be obtained as

$$Y^* = argmax_Y P(Y|X) \quad (2)$$

by Viterbi-like dynamic programming algorithm[6]. Features we use are the local lexicons and POS tags, and the dictionary composed of the existent entity names in the database. Table 2 shows the entity extraction results of the report in Table 1.

Table 2. Entity Extraction Result of The Report in Table 1.

Table 3.Features Used to Classify Whether The Entity e Is Associated with The Time Expression t.

DistanceBetween (e, t)
WordBetween(e,t)
TenseOf Sentence(e,t)
NegativeVerbsInSentence(e,t)
PositiveVerbsInsentence(e,t)
ContainDate(t)
PrepositionBefore(t)
FromDocument(t)

Table 4. Information Extracted from The EOC Report in Table 1.

Service	Time	Status
Miami-Dade County offices	October 24, 2005	close
public schools	October 24, 2005	close
courts	October 24, 2005	close
transit bus	October 22, 2005 6:30 p.m.	open
Rail service	October 22, 2005 6:30 p.m.	open
…		
Miami International Airport	October 22, 2005 6:30 p.m.	open
hurricane evacuation centers	October 23, 2005 afternoon	open

2.1.2 Relation Extraction

If a sentence contains an entity but no time expression, the time of the report will be associated with the sentence. To generate the triple by connecting the entity with the time expression with a proper status label, we train a multi-category Support Vector Machine (SVM) [8] to classify each pair of (entity, time) to a proper category, defined as "no relation"/"open"/"close"/"unclear". Table 3 shows the features we used for classification, from which the TenseOfSentence(e,t), NegativeVerbsInSentence(e,t), and PositiveVerbsInsentence(e,t) are extracted as the heuristic rules to indicate the tense of the sentence, the verbs with negative modifier, and the verbs without negative modifier semantically in the sentence. Note that FromDocument(t) indicates whether the time is the time associated with document.

We extract those pairs of entity and time expressions in "open" and "close" categories to form the triple. The time expressions are formatted into an absolute form of expression from relative time expressions such as "next Monday", "this afternoon". using the time of the report as a benchmark. The structured information extracted from the report in Table 1 is shown in Table 4.

2.2 REPORT SUMMARIZATION

The hierarchical multi-document summarization method generates the hierarchical summaries of reports. We use the Affinity Propagation (AP) [11] clustering method to build a hierarchical structure for sentences of related reports.

2.2.1 Affinity Propagation

The input of the Affinity Propagation algorithm is the sentence similarity graph defined as $G<V, E>$:V is the set of vertices with each vertex, called data point, representing a sentence. E is the set of edges. Let $s(i,k)$ be the similarity between two distinct points i and k, indicating how well data point k is suitable to be the exemplar of point i. Specially, $s(i,i)$ is the preference of a sentence i to be chosen as the exemplar. There are two kinds of messages passing between data points: responsibility and availability.

The responsibility $r(i,k)$ is computed as follows,

$$r(i,k) \quad s(i,k) \quad _{\{k \quad \}}\{a(i,k') \quad s(i \quad)\} \quad (3)$$

The responsibility $r(i,k)$is passing from i to candidate exemplar k. It reflects the accumulated evidence of how well point $k$is selected as the exemplar for point i against other candidate exemplars.

The availability $a(i,k)$ is computed as follows,

$$a(i,k) \quad n\{0, r(k \quad k) + \sum\nolimits_{\notin\{i,k\}} \quad x\{0, r(i \quad k)\} \quad (4)$$

The availability$a(i,k)$ is passing from the candidate exemplar k to point i, reflecting the accumulated evidence of how appropriate point i to choose point k as its exemplar, considering the support from other points which share point k as exemplar. Whereas the responsibility updating lets all candidate exemplars compete for the ownership of a data point, the availability updating gathers evidence from data points to measure the goodness of each candidate exemplar.

The self-availability $a(k \quad k)$ is updated as follows:

$$a(k \quad k) \quad \sum \quad \max\{0, r(i', k)\} \quad (5)$$

This message reflects accumulated evidence of point $k$being an exemplar based on the received positive responsibilities from other points.

All availabilities are initialized to zero:$a(i,k) = 0$. After the updating converges, availabilities and responsibilities are combined to identify exemplars. For point i, its corresponding exemplar is obtained by maximizing the following expression:

$$\{a(i,k) \quad r(i,k)\} \quad (6)$$

We choose AP for the following reasons:

- AP can find clusters with much lower error than other clustering methods, such as K-Means [36].
- AP performs efficiently on sparse similarity graphs, which is the case of document space. The run time for iterations is linear with the number of edges in the graph.
- AP takes a real number as input, called preference for each data point. The preference quantifies the likelihood of it being chosen as an exemplar. Thus, prior and heuristic knowledge can be used to associate different sentences with different preferences.
- AP identifies exemplars for each cluster or group which can be naturally used as the summary sentences for the cluster.

2.2.2 Hierarchical Summarization on Affinity Propagation

For the sentences in related reports, $\{s_1, s_2, \cdots, s_n\}$, we want to build a hierarchical clustering structure and use exemplars of clusters as the summary. Starting from all sentences, we recursively apply affinity propagation in an agglomerative way to find proper exemplars until the number of exemplars is small enough. We pick 20 as the number of exemplars which means 20 sentences will be selected from the document set as the summary. The preference for each sentence and similarity between sentences are used as the input of the affinity propagation algorithm.

2.2.3 Sentence Preference

We define the preference of sentence i to be chosen as an exemplar using the following scores:

LanguageModelScore L: For sentence i, is calculated as the logarithmic probability of sentence using unigram model training on the reports $\{s$ }. Generally, a shorter sentence which has more frequent words in the reports has a higher score.

LexPageRankscore P: LexPageRank proposed by [12] calculates the Page Rank score of sentences on the sentence similarity matrix. The score measures the prestige in sentence networks assuming that the sentences similar to many of the other sentences in a cluster are more prestigious to the topic. Since the original LexPageRank can be interpreted as the probability in random walk theory, we use the logarithmic version to make it at the same scale with the Language Model score.

FreshnessScore F: Users are generally more interested in the latest information; we calculate the freshness score of sentence i based on the age of the document containing i as

$$(7)$$

where a_i is the *age* in term of the number of days the document contains the sentence i. Clearly, $F_i \in [0,1]$ decreases as the document age increases. Another property is that for two sentences from two documents with some age difference (e.g. 1 day), the difference of their freshness scores is large when both sentences are relatively new. Thus it can better differentiate freshness for latest information.

Finally, the preference of s_i is the sum of the three feature scores with a scaling parameter:

$$s(i,i) = (L \qquad F) \qquad (8)$$

The parameter e is obtained by experimentally testing the clustering results and choosing the value that achieves the best clustering performance.

2.2.4 Sentence Similarity

Sentence similarity $s(i, j)$ indicates how well the data point with index j is suited to be the exemplar for data point i. In our case, it means how likely sentence i can be summarized by sentence j. If sentence i and sentence j have non-stop word overlaps, we calculate $s(i, j)$ by the log-likelihood of sentence i given that its exemplar is sentence j.

$$s(i,j) \qquad P(i|j) \quad (9)$$

To calculate the conditional probability, a unigram language model is trained on sentence j by using the Dirichlet smoothing [35]. Then the probability of sentence i is calculated by using the language model.

2.3 DYNAMIC DASHBOARDS AND DYNAMIC QUERY FORM

2.3.1 Dynamic Dashboard

2.3.1.1 The Challenges for Dashboards

When a disaster happens, the system will receive a lot of information at once. It is necessary for the system to select a small portion of entities that a user really cares about to display in the dashboards. The dashboards provide condensed views for users to quickly explore the recent news and reports. It cannot display all the information in such a small area.

Another problem in practice is the information sent from company users may have a lot of redundancies. For instance, when a hurricane arrives in South Florida, almost all the company users in that area will report the same hurricane information: "*The storm has arrived South Florida*". Thus, different users may report the same information hundreds of times. Therefore, the system has to identify which information is redundant and the redundant information should not appear in the dashboards.

We address these problems by introducing the dynamic dashboard supported by the content recommendation engine. The engine's main task is to extract the most important, relevant and non-redundant information about entities from news and reports.

2.3.1.2 The Content Recommendation Engine

Figure 1 shows the data flow related to the content recommendation engine. There are four main data sources: EOC reports, news, company reports, and company messages. Since reports and news may contain information about multiple entities, in content recommendation engine, each report or news is divided into several documents. Each document consists of a sentence containing entity status information plus a context window (one previous and next sentence).

The content recommendation consists of two steps. The first step is text clustering, which is to cluster the same description of entities into one cluster. The second step is ranking the text by the relevance and presenting the top k items to the dashboards.

The content recommendation engine is based on unstructured text, while the situation dashboard, thread dashboard, and company dashboard display structured information. The four dashboards are denoted as Db_S(Situation Dashboard), Db_T(Threat Dashboard), Db_E(Event Dashboard), Db_C(Company Dashboard). The maximum

numbers of items allowed to show in the dashboards Db_S, Db_T, Db_E, Db_C are denoted as $size_S$, $size_T$, $size_E$, $size_C$ respectively.
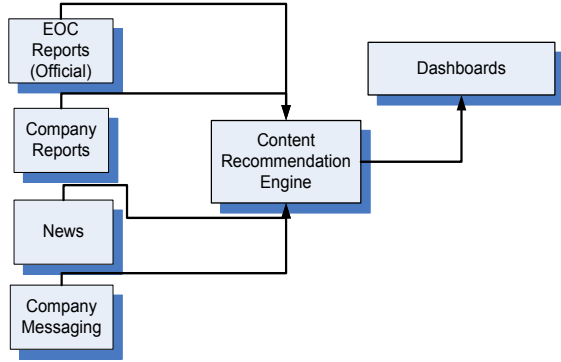

Figure 1. The Content Recommendation Engine.

Table 5.The Data Sources of Different Dashboard.

Dash-board	EOC Reports	Company Reports	News	Company Messaging
Db_S	√		√	
Db_T	√			
Db_F	√	√	√	√
Db_C		√		√

The content recommendation engine recommends information from different data sources to the four dashboards. Table 5 shows the relationship between the data sources and the four dashboards. Since the dashboards show the latest information, we use the last 48 hours records and news as the input of the engine.

For any user u, the set of information submitted by u is denoted by $I(u)$ and the set of reports/news of which the details are viewed by u is denoted by $J(u)$.u's profile is composed of $I(u)$ and $J(u)$.

2.3.1.3 Document Clustering

Before performing clustering, we use Term Frequency - Inverse Document Frequency (TF-IDF) transformation [13] to transform the text data (report, news and so on) to the vectors. The similarity between two documents can be calculated by the cosine similarity [14].

We apply the K-Medoids [15] algorithm to cluster the documents. Note k is a user-defined parameter, which is determined by the managers of the system. It is also relevant to the number of items allowed to be displayed on the dashboards. We present the top 5 (k=5) items in the dashboards.

After clustering, each cluster contains the duplicated information about an entity and one document can be selected from a cluster to show the status of the entity. But before that, we have to decide which cluster and which document should be selected.

2.3.1.4 Content Ranking

For a specific user u, there are three priorities of the information. The three priorities from highest to lowest are EOC reports,

company partner's information (messages received) and other users' information (company reports). The three priorities are denoted by user-defined parameters pr_1, pr_2 and pr_3 respectively, and $pr_1 > pr_2 > pr_3 > 0$. For a given document $d_i \in D$, we use $pr(d_i)$ to indicate the priority of this document, and $pr(d_i) \in \{pr_1, pr_2, pr_3\}$.

Suppose the current user is u, $t(u)$ represents the term vector representation of the documents submitted or read by u .We can obtain the u's feature f_u by users' profiles:

$$\frac{\Sigma_{(u)} t(u)}{|\Sigma_{(u)} t(u)|} \quad (1 \quad) \frac{\Sigma_{(u)} t(u)}{|\Sigma_{(u)} t(u)|} \qquad (10)$$

The parameter α is used to tune the importance weights of the reports submitted and viewed as the profile. α is set to 0.8 in our work.

The importance score of each document $d_i \in D$ is calculated as follows where $t(d_i)$ represents the term vector representation of document d_i:

$$e(d_i) \quad \text{im}(f \quad t(d_i)) \ pr(d_i) \qquad (11)$$

For each dashboard, we use a top-K query to greedily search the K highest scores' documents from its corresponding data sources, where $K \in \{size_S, size_T, size_E, size_C\}$ and no two documents are selected from the same cluster. The set of K highest scores' documents is Sjust the result of the content recommendation engine. The EOC official reports have the highest priority. ome of them are not very relevant to the current user; however, information from these reports is still likely to appear on the Event Dashboard.

2.3.2 Dynamic Query Form

Each report is associated with a set of attributes, such as the report location, date, or annotations added by the creator. Such structural information allows users to execute relational queries on reports. For example, we want to find those reports which are about hurricanes from 1990 to 2010 and the latitude of the hurricane center is above 30 degrees. Hence, our system applies query forms for users to support relational queries.

Traditional query forms are statically embedded by developers or database administrators. Those static query forms are used for the static database schema. However, different reports have different sets of attributes. For example, the hurricane report and the earthquake report use two very distinct sets of attributes. Furthermore, the associated values of annotation attributes created by the user at runtime are inconsistent. Therefore, it is impossible to design a static and fixed query form to cover all those attributes. So, we implement the dynamic query form to satisfy those dynamic and heterogeneous query desires.

Previous research on database query forms focuses on how to automatically generate the query form from the data distribution or query history [16-19]. However, different users can have different query desires. How to capture the current user's interests and construct appropriate query forms are the key challenges for query form generation which has not been solved.

2.3.2.1 Problem Formulation

Query forms are designed to return the user's desired results. The metric of the goodness of a query form is based on two traditional measures of evaluating the quality of the query results: *precision* and *recall*.

(a) Generated Communities   (b) Interactive clustering of large cluster
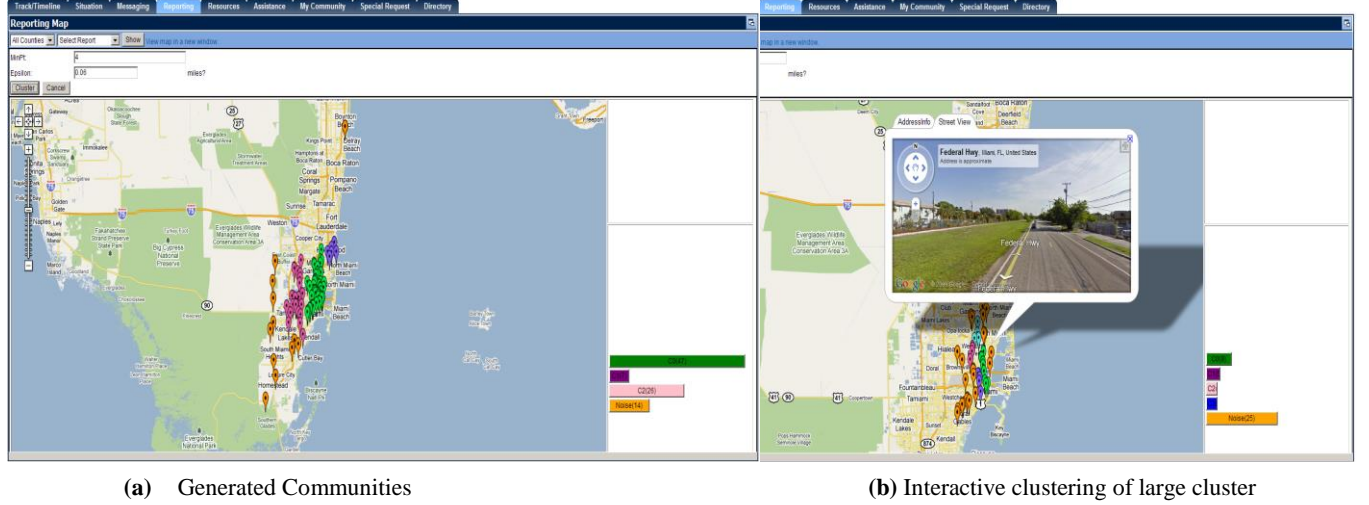
Figure 3. Dynamic Community Generation Result

Let $F=(\mathbf{A}_F, \sigma_F)$ be a query form with a set of query conditions σ_F and a set of displaying attributes $\mathbf{A}_F$. Let D be the set of all reports in the database. $|D|$ is the total number of reports. $P_u(.)$ is the distribution function of user interests. $P_u(d)$ is the user interest for a report d, and $P_u(\mathbf{A}_F)$ is the user interest for an attribute subset $\mathbf{A}_F$. $P(\sigma_F|d)$ is the probability of query condition σ_F being satisfied by d, i.e., $P(\sigma_F|d) = 1$ if d is returned by F and $P(\sigma_F|d) = 0$ otherwise. Then, given a query form $F=(\mathbf{A}_F, \sigma_F)$, the *expected precision*, *expected recall* and *expected fscore* of F are defined as follows:

$$Pr(F) = \frac{\sum (d)P_u(A_F)P(\sigma\ d)}{\sum P(\sigma\ d)} \tag{12}$$

$$(F) = \frac{\sum (d)P_u(A_F)P(\sigma\ d)}{\sum (d)P_u(A)} \tag{13}$$

$$(F) = \frac{(1\ )\cdot Pr\ (F)\cdot R\ (F)}{(F)\cdot Recall_E(F)} \tag{14}$$

where $A_F \subseteq A, \sigma\ \in \sigma$, β is a parameter defined by the user and is usually set to 2.

$FScore_E(.)$ is the metric to evaluate the overall goodness of a query form. The problem of our dynamic query form [47] is how to construct a query form $\hat{F}$ that maximizes the goodness metric $FScore_E(.)$, i.e.,

$$\hat{F} = \qquad e(F) \tag{15}$$

2.3.2.2 Method Description

It is impractical to construct an optimal query form $\hat{F}$ at the very beginning, since we do not know which reports and attributes are desired by the user. In other words, estimating $P_u(d)$ and $P_u(\mathbf{A}_F)$ is difficult.

ADSB system provides an iterative way for the user to interactively enrich the query form. Figure 2 shows the work-flow of our dynamic query form system. At each iteration, ADSB computes a ranked list of query form components for users, and then lets users make the choice for their query form. Those query form components are ranked by the metric $FScore_E(F)$.

There are two types of query form components: attribute display and query condition.
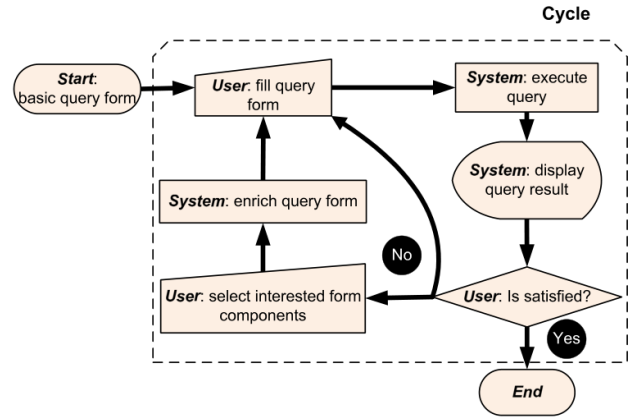
Figure 2. Flowchart of Dynamic Query Form

Assuming the current query form is F_i in the flowchart, and the next query form is F_{i+1}. We need to estimate $P_u(d)$, $P_u(\mathbf{A}_{F+1})$ and $P(\sigma_{F+1}|d)$ to compute $FScore_E(F_{i+1})$. The estimation is based on user behaviors when interacting with the ADSB system. Let $D_{u,f}$ be the set of reports viewed by the users and d' be one of the document in $D_{u,f}$. We assume those reports are interesting to the current user, then

$$(d) = \sum (d\ ')P_u(d') \tag{16}$$

We use the random walk model to compute the relevance score between reports as the value of $P_u(d|d')$ [20].

Suppose A is displaying an attribute we suggest for query form F_{i+1} and $A_{f_{i+1}} = A \cup A_{F_i}$, where $A \in \mathbf{A}$, $A \notin A_{F_i}$. So, A_{F_i} can be obtained in the current query form F_i.

$$(A\ ) = P_u(A|A_{F_i})P_u(A_{F_i}) \tag{17}$$

We also estimate $P_u(A\ )$ by using a random walk model on the *attribute graph*. The nodes of the attribute graph are report attributes, and the edges are common reports. So the weight of edge ij is computed by how many reports use both the two attributes i and j.

Suppose s is a query condition we suggest for query form F_{i+1}. So σ σ_{F_i}, where s is a single query condition for attribute A_s, $A_s \in \mathbf{A}$. σ_{F_i} can be obtained in the current query form F_i. For each report $d \in D$, $P(\sigma_{F_{i+1}}|d) = P(s|d) P(\sigma_{F_i}|d)$. It is very time-consuming to find the best s by brute-force search on all $P(s|d)$. So we pre-compute the $P(s|d)$ and store it in the database.

2.4 COMMUNITY GENERATION AND USER RECOMMENDATION

2.4.1 Community Generation

Two characteristics in disaster recovery scenarios motivate us to consider geo-location information. The first characteristic is that any event extracted from a report is associated with a/several location(s) indicating the place(s) where the announced event takes place. The second characteristic is that spatially co-located entities are more likely sharing similar disaster damage situations.

These two characteristics motivate the concept of community: a community is a certain geographical region in which entities tend to share more recovery status or interests in common. So geographically identifying those communities is important to help companies understand the current disaster situation and any interested resources nearby. Our system addresses community generation by adapting existing spatial clustering algorithms. In practice, we provide an interactive spatial clustering interface for users to access multi-level communities in a top-down manner and consider physical or non-physical obstacles when generating spatial clusters to form more practical communities.

2.4.1.1 Spatial Clustering

Spatial data clustering identifies clusters, or densely populated regions, according to some distance measurement in a large, multi-dimensional data set [14,21]. Many spatial clustering techniques [22,23,29] have been developed for identifying clusters with arbitrary shapes of various densities and with different physical constraints.

In practice, communities formed by geographically related entities can be of various shapes. So we extend DBSCAN [22], a well-known density-based clustering algorithm, which is capable of identifying arbitrary shape of clusters, to generate dynamic communities.

2.4.1.2 Spatial clustering with constraints

We consider the method of spatial clustering with constraints. Generally, there are three types of constraints [21]: 1). Constraints on individual objects: such constraints are non-spatial instance-level constraints that can be preprocessed before performing clustering algorithms. 2). Constraints as clustering parameters: such constraints are usually confined to the algorithm itself. Usually, user-specified parameters are given through empirical studies. 3). Constraints as physical obstacles: such constraints are tightly intertwined with clustering process. It is clear that physical obstacles are such constraints which prevent two geographically close entities from being clustered together. For example, the bridge, highway and rivers are of this type.

In our BCIN system, we focus on object constraints and physical constraints.

Object constraints: We have two ways to obtain object constraints: 1) users submit formatted reports through report interface. Those reports are immediately recorded in the database;

2) our system extracts entity status from reports. For example, Table4 can be used as object constraints.

Obstacle constraints: Polygon is a typical structure in spatial analysis to model objects. Obstacles modeled by a polygon can be represented as a set of line segments after performing polygon reduction [23].

Figure 3a shows the communities generated by clustering all open facilities and companies in Miami with the constraint: "I75 closed".

2.4.1.3 Interactive Spatial Clusters

In order to deal with unbalanced size of clusters, we provide users with an interactive mechanism to track the sub-community information within a large size community. Further clustering process will be triggered in the runtime when a user selects a larger community and wants to see the cluster information within such a community at a finer granularity. By using this mechanism, users can obtain clusters with different granularities and more meaningful results. The Figure 3b shows the interactive clustering results within the largest cluster in Figure 3a.

2.4.2 User Recommendation

The user recommendation component provides an interface to explore other users' recommendations or share reports with other people. It also helps the user quickly identify sets of users with shared interests. It is designed by considering each individual's transactional sharing history, textual content of each transaction and timeliness of interaction to provide each user with a personalized information sharing experience.

Related work has been applied to email communication networks analysis to find important persons, identifying frequent communication pattern and detecting communities based on transactional user relationships [24-28]. Those techniques can prevent a user from forgetting to add important recipients, avoid costly misunderstandings, and communication delays. Carvalho et al. [27] introduced several supervised learning models to predict the score of each user associated with a given email content. By aggregating TF-IDF vector of each email addressed to a user (by To, CC or BCC), it can predict the score of a new email to such user. However, it was not aware of the different importance of emails for senders and recipients. Horn et al. [28] explicitly associated higher weights to senders, and also consider user-interaction graph as a directed hyper-graph. It focused on the time and frequency of interactions but ignored the content information involved in each email, which could be an important indication of potential related users.

There are three practical considerations motivating the user recommendation: 1.To share information to the right/related people, users need an intelligent tool to auto-generate a recipient list which covers active users interested in specific information; 2. Manually identifying meaningful groups of users is time-consuming, so users prefer efficient ways to organize contacts instead of navigating the contact list repeatedly; 3. It could be more effective for a user to access information that others think is important.

So, our system addresses the above-mentioned issues by considering both user interactions and textual information. In practice, we provide dynamic user suggestions for the news recommendation and community recommendation interface to help our system users organize their critical partnerships.

2.4.2.1 Transactional Interactions

An interaction or transaction is defined as the process of a user sharing a report with one or more other users. So, the reports sharing transaction database can be treated as a hyper-graph with each node representing a registered user and a set of edges created at the same time from one node to a set of nodes representing an occurred transaction. There are three important factors associated with each edge:

Time: The time that the transaction happened. It indicates the importance of recency. In general, the more recently a transaction happens, the more important the report is to those users involved.

Direction: The relation of an interaction. An edge pointed from node A to node B indicating that A shares some information with a set of users including B. The direction indicates that the shared information is more important to the sender than to receivers.

Textual Content: Each transaction is associated with some certain textual content, so the content of an edge means that someone thinks such content is important or related to some group of users.

In practice, a personalized user recommendation requires the algorithm to identify potential users who have frequent and active interactions with the sender and are also interested in some certain topics. In completion of two recommendation tasks, we extend both [27] and [28] by taking the direction, timeliness and textual content of the interaction into consideration to generate 1) a suggested user list for specific report and 2) a suggested user list for specified seeds (users).

2.4.2.2 User Groups

There could be multiple transactions associated with a specified user and each transaction involves a group of users (Figure 4).

Even though transactions may include the same sender and receivers, they are treated as unique in the transactional hyper graph since they are associated with unique timestamps. Despite the textual content of each transaction, the contribution of each group to current user seeds can be evaluated by Interaction Rank proposed in [28].

Transaction Hyper-graph

Legal groups for u:
U as sender
{1, 2}: solid black.
{1, 3}: mixed purple
{2, 3, 4}: dotted blue.
U as receiver
{3, 4}: dashed red.
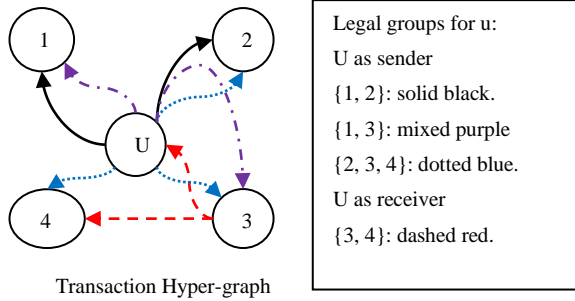
Figure 4. Transactional User Groups

2.4.2.3 User Profile

To build the user profile, we consider textual content in all transactions related to the user. Carvalho [27] introduced a centroid vector-based representation which aggregates all related documents to build a user profile. In our method, we consider transaction directions and assign document sending weight $\mathcal{W}_s$ or receiving weight $\mathcal{W}_r$ respectively. The values of $\mathcal{W}_s$ and $\mathcal{W}_r$ are manually decided based on different scenarios. For example, if the relevance of an email to the sender is higher than to the receiver

(for example, junk mail or ads), then we can assign a much larger value to $\mathcal{W}_s$ than to $\mathcal{W}_r$. We use TF-IDF transformation to represent textual content as a vector. So the user profile can be represented as:

$$profile(u) = \mathcal{W}_s \cdot \sum_{d \in S(u)} tfidf(d) + \mathcal{W}_r \cdot \sum_{d \in R(u)} tfidf(d)$$

(18)

where $tfidf(d)$ is defined as

$$tfidf(d)_i = TFIDF(d)_i{}^t$$ (19)

where $t = \frac{time(now) - time(n)}{\lambda}$ indicates an over-time exponential decay of each document's contribution. $S(u)$, $R(u)$ are sets of documents sent and received by u respectively. So user u's preference to report d can be generated by computing the cosine similarity between the user's profile and the TF-IDF vector of d:

$$preference(u, d) = cos\left(profile(u), ts_{tfidf}(d) \right)$$ (20)

Practically, the user profile is stored separately and will not be updated in each calculation. Typically, it will be updated every few days or when new events are announced.

2.4.2.4 Algorithm

We extended the friend-finding algorithm proposed in [28] to generate a list of user recommendations by aggregating the groups' contribution to a user and considering the relevance between users and reports. Our algorithm is described in Figure 5. The score of each user in the list represents the interaction preference with respect to the given user and report.

Input: *u*, the user; *d*, the report, and *𝓢*, the seeds
Output: *𝓡*, recommended user list
1. *𝓖* ← *GetTransactionalGroups(u)*
2. *𝓡* ← ∅
3. for each group *g* ∈ *𝓖*
4.   for each user *c* ∈ g, *c* ∉ *𝓢*
5.   if *c* ∉ *𝓡*
6.   *𝓡* [*c*] ← 0
7.   *𝓡* [*c*] ← *𝓡* [*c*] + *GroupScore(c, 𝓢, g, d)*
or   [*c*] ← *𝓡* [*c*] + *CommunityScore(c, 𝓢, g)*

Figure 5. Suggesting User Routine

2.4.2.5 Group Contribution

From the algorithm described in Figure 5, the interaction preference of a user is the aggregated value of the contribution that each transaction made to the user. There are two types of contribution measurements with respect to different tasks. We use group score and community score to represent contributions for report sharing and community user recommendation respectively.

2.4.2.6 Group Score

The group contribution $\mathcal{GC}$ described below represents the contribution that a user group contributes to a user. There are two situations considered, 1) suggesting users related to a document

based on the preference (similarity) between the document and a user; 2) suggesting a user group based on the similarity between users. We defined $\mathcal{GC}$ as an aggregated score of users' preferences to a specific document considering the direction and timeliness of each interaction.

For the first situation, we use similarities between each user in a group and report d:

$$\mathcal{GC}(d,g) = \mathcal{W}_s \cdot \sum_{i \in O(u,g)} s(i,d)^t + \mathcal{W}_r \cdot \sum_{i \in I(u,g)} s(i,d)^t \quad (21)$$

where $s(i,d) = \sum_{u \in i} preference(u,d)$.

For the second situation, we simply modified the $\mathcal{GC}(d,g)$ as $\mathcal{GC}(c,g)$ and $s(i,d)$ as

$$s(i,c) = \sum_{u \in i} cos\big(\, profile(u)\,, profile(c)\big) \quad (22)$$

to calculate the similarity without document information.

In both situations, $O(u,g)$ and $I(u,g)$ are sets of sending and receiving interactions/transactions respectively in which user u was involved.

2.4.2.7 Recommend Users with Report

To recommend a report to a group of users, one should consider historic recommendation transactions and the report's textual content. The score that a transaction contributes to a user is the aggregation of preferences of a group of users to the given report:

$$\boldsymbol{GroupScore}(c,\mathcal{S},g,d) = \begin{cases} \mathcal{GC}(d,g), \text{if } \mathcal{S} \cap g \neq \emptyset; \\ 0, \ otherwise. \end{cases} \quad (23)$$

2.4.2.8 Recommend Users for Communities

Recommending users to form communities involves historic transactions without textual information. The score that a transaction contributes to a user is the aggregation of similarities between the user and users in the group:

$$\boldsymbol{CommunityScore}(c,\mathcal{S},g) = \begin{cases} \mathcal{GC}(c,g), \text{if } \mathcal{S} \cap g \neq \emptyset; \\ 0, \ otherwise. \end{cases} \quad (24)$$

A user can arbitrarily choose target users at runtime. Starting from those chosen users as seeds, our recommendation components can dynamically generate more users related to the given textual content and list of users with high concurrence.

3. Case Study

3.1 BCiN

BCiN (Figure 6) is a web-based prototype implementation of a Business Continuity Information Network that is able to link participating companies into a community network, provide businesses with effective and timely disaster recovery information, and facilitate collaboration and information exchange with other businesses and government agencies. The system allows company users to submit reports related to their own business, and government users to make announcements on the public issues. To collect more information during the disaster, BCiN can monitor the news published on the websites and takes the news as its input. Like traditional information systems, these reports and news, and the status information of entities they contain can be retrieved and accessed by queries. For example, reports can be viewed according to alert categories or geo-locations, and resources can be viewed according to status or usages. Furthermore, BCiN not only displays users-submitted information but also conducts necessary and meaningful data processing work. BCiN makes recommendations based on the current focus and dynamically adapts based on users' interests. BCiN summarizes reports and news to provide users with brief

and content-oriented stories, preventing users from being troubled when searching in huge amount of information. By introducing the concept of Community, BCIN offers users a hierarchical view of important reports or events around them.
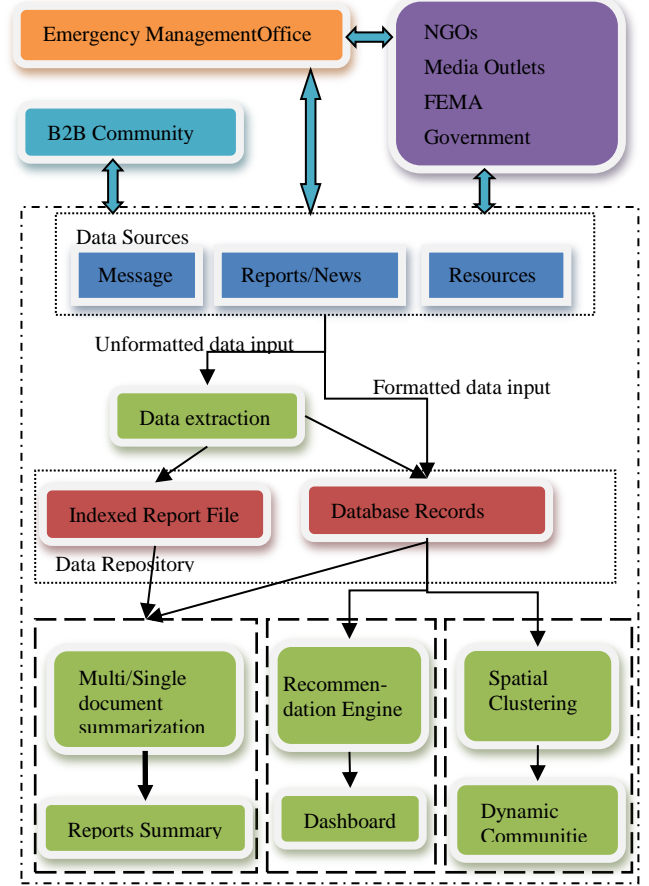
Figure 6. BCIN System Architecture

Four main information processing and representation components are implemented in BCiN: Information Extraction (See Section 2), Report Summarization (See Section 3), Dynamic Dashboard (See Section 4.1), and Dynamic Community Generation (See Section 5.1). These four different components are tightly integrated to provide a cohesive set of services and constitute a holistic effort on developing a data-driven solution for disaster management and recovery.

3.2 ADSB

Professionals who have an operational responsibility in disaster situations are relying on mobile phones to maintain communications, update status and share situational information. Consumers, too, are finding mobile devices convenient for sharing information about themselves and what is going on in their lives. By using a mobile platform we can build native applications which utilize onboard sensors, rich media, and simplified user interface to engage users in a way they feel most comfortable to share such information in a disaster situation.
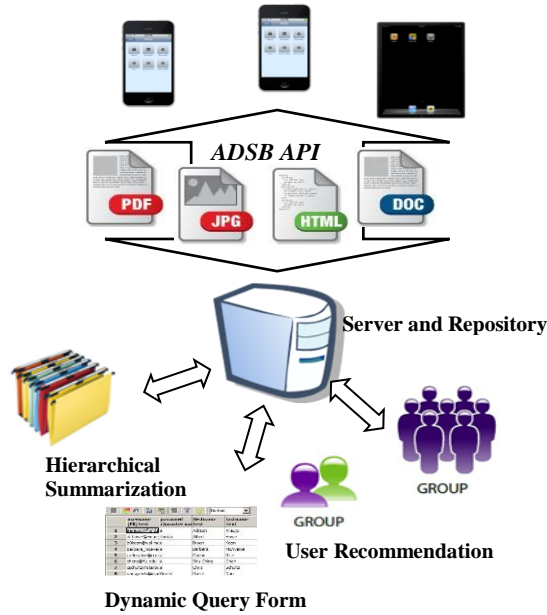
Figure 7. ADSB System Architecture

ADSB is an **All-Hazard Disaster Situation Browser (ADSB)** system that runs on Apple's mobile operating system (iOS) and iPhone and iPad mobile devices. Figure 7 illustrates the system architecture, and Figure 8 illustrates the system screenshot. Four major components are implemented in ADSB: Information Extraction (See Section 2), Hierarchical Summarization (See Section 3), Dynamic Query Form (See Section 4.2), and User Recommendation (See Section 5.2). A video demonstration is available at http://users.cis.fiu.edu/~taoli/ADSB-Demo/demo.htm.

4. SYSTEM EVALUATION

The data sources used in our project can be broadly divided into two categories based on the temporal characteristics: static data sources and dynamic data sources. Static data sources include historical data from Miami-Dade EOC. Dynamic data sources include: (a) situation reports from Miami-Dade EOC and participating companies illustrating the current status of threat, ongoing operations and goals/objectives for preparation and recovery efforts;(b) open/closure status about roadways/highways/bridges and other infrastructure such as Fuel, Power, Transportation, Emergency Services (Fire Stations, Police Stations), Schools and Hospitals; (c) reports crawled from FEMA [31] web site with information about twenty major disasters since 2000; and (d) tweets posted in August 2010 by using Twitter API [30] from dozens of active accounts.

Evaluation is conducted on two levels: algorithm evaluation and system evaluation. To evaluate the algorithms, we use standard performance metrics and compared our algorithms with existing work when applicable. Using report summarization as an example, we conducted experiments on a dataset of press releases collected from Miami-Dade EOC and Homeland Security during Hurricane Wilma from Oct. 19, 2005 to Nov. 4 2005. The dataset contains 1700 documents in total, concerning all the related events before Hurricane Wilma came, during Hurricane Wilma, and after

Hurricane Wilma passed [46]. The documents report various types of information such as the movement of Hurricane Wilma, the location of evacuation zones, and the cancellation of social activities. In order to evaluate the summarization performance, human generated summaries are used as references. The summarization results are evaluated by ROUGE [37].

Report List

Report Detail

Report Sharing

Community Management

Hierarchical Summarization

Dynamic Query Form

Figure 8. ADSB Screen Shots of Important Components

* iPhone implementation has the same style with iPad but without rich visual abilities, such as the split view.

Table 6 shows the experiment results and demonstrates the efficiency of using affinity propagation to generate hierarchical document summarization (Centroid means picking the cluster centroid as the representative sentence).

Table 6 Summarization results comparison

Measure		Centroid	Affinity Propagation
ROUGE-1	Recall	0.3409	**0.3788**
	Precision	0.1991	**0.3311**
	F-Score	0.2514	**0.3534**
ROUGE-2	Recall	0.0916	**0.1069**
	Precision	0.0533	**0.0933**
	F-Score	0.0674	**0.0996**
ROUGE-SU4	Recall	0.1121	**0.1173**
	Precision	0.0649	**0.1023**
	F-Score	0.0822	**0.1093**

Our system evaluation process consists of presenting the system to emergency managers, business continuity professionals and other stakeholders for feedback and performing community exercises. The exercises involve a real time simulation of a disaster event integrated into an existing readiness exercise conducted each year. This evaluation exposes information at different time intervals and asks the community to resolve different scenarios by using the tool. The evaluation is a form of a "table-top" exercise in which injected information provides details about the current disaster situation and specifies potential goals and courses of action. Participant use the system to gather information to assess the situation and provide details about the actions they will take. We gather information about what information they found to derive their conclusions (or lack thereof). This information allows us to better understand how those techniques improve the information effectiveness.

Table 7. Evaluation Exercises.

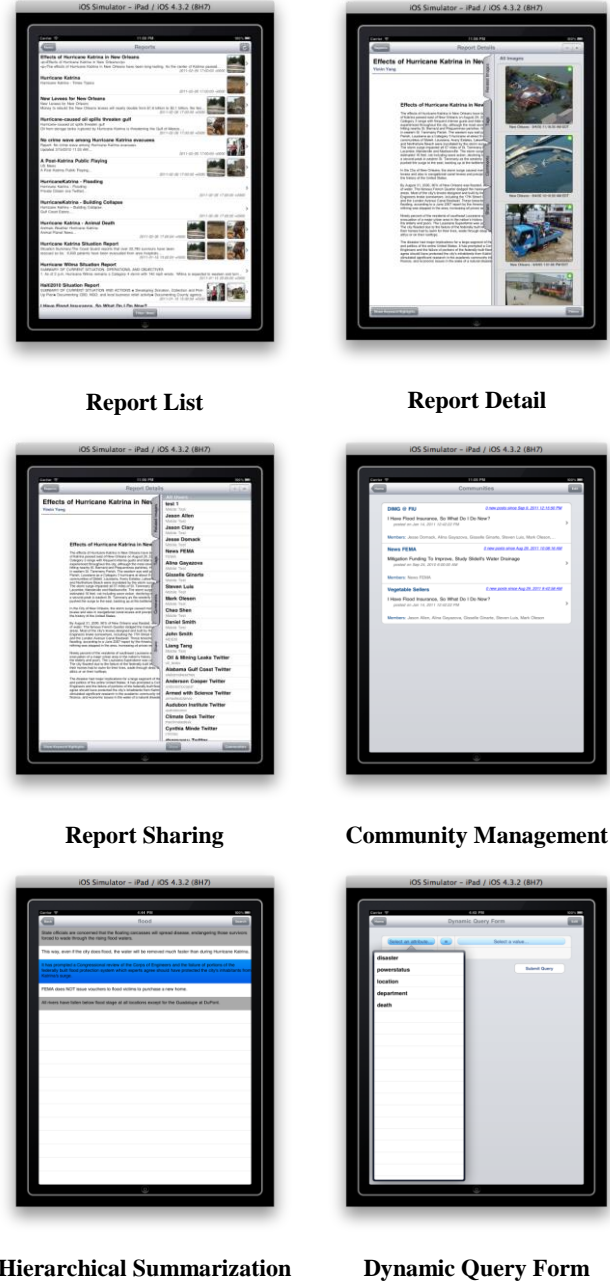| Date	Description of the Exercise
Jun. 01 2009	In Florida Dept. of Emergency Management's Statewide Hurricane Exercises, BCiN was utilized in a scenario where Miami-Dade County Emergency Management Business Recovery Desk facilitated the logistics to deploy portable ATMs at Shelters and PODs in Miami-Dade County.
Jun. 29 2009	In Miami-Dade UASI exercise, BCiN supported communicating and collaborating with several companies that participated in the event as observers.
Aug. 20 2009	In a full scale company BCiN training, about 30 companies were given injects to provide information to resolve different information requests.
May 10 2010	In Miami-Dade Dept. of Emergency Management's Statewide Hurricane Exercise, our systems were responsible for disseminating and responding to injects during the course of the exercise for both government and company users.
Jul. 29 2010	In Miami-Dade company exercises, over 50 company attendees used our systems for a training exercise.
May 12 2011	In the county of West Palm Beach exercise, we demonstrated the system to WPB Dept. of Emergency Management and companies.

Table 7 describes the exercises. In a regional disaster such as a hurricane, business continuity professionals are under extreme pressure to execute their continuity of operation plans because many of the usual sources of information and services about the community and supply chain are completely disconnected, sporadic, redundant, and many times lack actionable value. The system focuses user input and collaboration around actionable information that both public and private sector can use.

To validate the usability and performance of our system, the participants and the EOC personnel at Miami-Dade participated in the questionnaire session after the exercise. A set of ten questions was designed to evaluate our system where nine of them are multiple choice questions with a 5-level scale (Strongly Agree, Agree, Not Sure, Disagree, and Strongly Disagree) and the last one is an open-ended question. Some of the multiple choice questions are: Are you able to identify related reports that you are interested in? Are you able to identify the correct modules for your tasks? Are you able to switch between different modules? Are the system generated summaries useful? The open-ended question is about the user feedback and suggestion. On average, about 4 EOC personnel and 30 participants attended each exercise. The evaluation demonstrated that most of participants are satisfied with the performance of the tools. Specifically, seven out of nine multiple choice questions received "Strongly Agree" or "Agree" from over 90% of the participants, implying a high level of satisfaction with our system.

The feedback from our users are positive and suggest that our system can be used not only to share the valuable actionable information but to pursue more complex tasks like business planning and decision making. There are also many collaborative missions that can be undertaken on our system, which allows public and private sector entities to leverage their local capacity to serve the recovery of the community. We summarized the feedback as follows:

- Positive feedback: 1) the system is easy to use; 2) related reports are well organized based on personalized user groups; 3) reports summarization is representative and interesting.

- Some suggestions: 1) related multimedia information, including images and video, could be shown during navigation; 2) report summaries could be organized based on some points of interests.

5. CONCLUSION

We identified 4 key design challenges to support multi-party coordination during disaster situations. We proposed a unified framework that systematically integrates the different techniques developed in our previous work [5, 32]. Such a framework can be utilized when dealing with different systems or applications separately (e.g., BCiN and ADSB), and they are essentially collaborative platforms for preparedness and recovery that helps disaster impacted communities to better understand what the current disaster situation is and how the community is recovering. The system evaluation results demonstrate the effectiveness and efficiency of our proposed approaches.

During the system implementation and assessment process, the users provided suggestions, limitations and possible enhancements. Our future efforts will be focusing on the

following tasks: developing efficient tools to automatically crawl related information from public resources including news portals, blogs, and social Medias; capturing the current user's interests and construct appropriate query form; and understanding users' intends to provide them with actionable answers to their information inquiries.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] The Conference Board. Preparing for the worst: A guide to business continuity planning for mid-markets. Executive Action Series, February 2006.

[2] FEMA public Private Partnership Models. http://www.fema.gov/privatesector/ppp_models.shtmunder Miami-Dade County.

[3] K. Saleem, S. Luis, Y. Deng, S.-C. Chen, V. Hristidis, and T. Li. Towards a business continuity information network for rapid disaster recovery. International Digital Government Research Conference. 2008: 107-116.

[4] V. Hristidis, S. Chen, T. Li, S. Luis, and Y. Deng. Survey of data management and analysis in disaster situations. The Journal of Systems and Software, 83:1701–1714, 2010.

[5] L. Zheng, C. Shen, L. Tang, T. Li, S. Luis, S. Chen, and V. Hristidis. Using data mining techniques to address critical information exchange needs in disaster affected public-private networks, KDD '10, pages 125–134, 2010.

[6] J.Lafferty,A. McCallum, and F.Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proc. of ICML. 2001.

[7] F. Sha, and F.Pereira. Shallow parsing with conditional random fields. In Proc. of HLT-NAACL. 2003.

[8] C.W. Hsu, and C.J. Lin. A comparison of methods for multiclass support vector machines. IEEE Transaction on Neural Networks, 13(2):415-425, 2002.

[9] D. Wang, L. Zheng, T. Li, and Y. Deng. Evolutionary document summarization for disaster management. In Proc. of SIGIR. 2009.

[10] F.L. Han, C. Peng, and C.Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transaction on Pattern Analysis and Machine Intelligence, 27:1226–1238, 2005.

[11] B.J. Frey and D. Dueck, Clustering by passing messages between data points. Science. 2007.

[12] G. Erkan and D.R. Radev, Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP, 2004.

[13] G. Salton and M. J. McGill. Introduction to modern information retrieval. McGraw-Hill.

[14] P.N.Tan, M. Steinbach, and V. Kumar, 2005. Introduction to Data Mining, Addison-Wesley.

[15] J.Han,and M.Kamber. Data Mining Concepts and Techniques 2nd. Morgan Kaufmann.

[16] M. Jayapandian and H.V. Jagadish. Automated creation of a forms-based database query interface. In Proceedings of VLDB 2008, pages 695-709.

[17] M. Jayapandian and H.V. Jagadish. Expressive query specification through form customization. In Proceedings of EDBT 2008, pages 416-427.

[18] M. Jayapandian and H.V. Jagadish. Automating the design and construction of query forms. IEEE TKDE 21(10): 1389-1402, 2009.

[19] P.P. Talukdar, M. Jacob, M.S. Mehmood, K. Crammer, Z. G. Ives, F. Pereira, and S. Guha. Learning to create data-integrating queries. In Proceedings of VLDB 2008, pages 785-796.

[20] H. Tong, C. Faloutsos, and J. Pan. Fast random walk with restart and its application. In Proceedings of ICDM 2006, pages 613-622.

[21] J. Han and M.Kamber. Data Mining Concepts and Techniques 2nd. Morgan Kaufmann.

[22] M. Ester, H.P.Kriegel, J.Sander, and X.Xu. A density-based algorithm for discoverying clusters in large databases with noise. In Proc. of KDD. 1996.

[23] C.H.Lee. Density-based clustering of spatial data in the presence of physical constraints. Master's thesis, University of Alberta, Edmonton, AB, Canada, July 2002.

[24] M.D. Choudhury, W. A. Mason, Jake M. Hofman, Duncan J. Watts, Inferring relevant social networks from interpersonal communication, Proceedings of the 19th international conference on World wide web, April 26-30, 2010.

[25] I. Kahanda and J. Neville. Using transactional information to predict link strength in online social networks. In Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM), June 2009.

[26] S. Yoo, Y. Yang, F. Lin, I. Moon, Mining social networks for personalized email prioritization, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, June 28-July 01, 2009, Paris, France.

[27] V. R. Carvalho , W.W. Cohen, Ranking users for intelligent message addressing, Proceedings of the IR research, 30th European conference on Advances in information retrieval, March 30-April 03, 2008, Glasgow, UK.

[28] I. Horn, A. Leichtberg, N. Leiser, Y. Matias, and R. Merom. Suggesting friends using the implicit social graph. In Proceedings of KDD-2010, pages 233–242.

[29] O.R.Aaiane, A. Foss, C.H.Lee, and W.Wang. On data clustering analysis: Scalability, constraints and validation. In Proc. of PAKDD 2002.

[30] Twitter API, http://apiwiki.twitter.com.

[31] FEMA, http://www.fema.gov.

[32] L. Zheng, C. Shen, L. Tang, T. Li, S. Luis, and S. Chen. Applying data mining techniques to address disaster information management challenges on mobile devices. KDD '11, pages 283–291, 2011.

[33] D. McEntire. The Status of Emergency Management Theory: Issues, Barriers and Recommendations for Improved Scholarship. Paper presented at FEMA Higher Education Conference, Emmitsburg, MO.

[34] GeoVISTA, http://www.geovista.psu.edu.

[35] C.X. Zhai and J. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. SIGIR'01, pages 334-342, 2001.

[36] B.J. Frey and D. Dueck. Clustering by passing messages between data points. Science, 315(5814), 972-976.

[37] C. Lin. Rouge: A package for automatic evaluation of summaries. Post-Conference Workshop of ACL, 2004.

[38] E. Brill. Part-of-speech tagging. Handbook of Natural Language Processing(2000): 403-414.

[39] The Puerto Rico Disaster Decision Support Tool (DDST), http://www.udel.edu/DRC/DDST/.

[40] E.J.Bass, L.A.Baumgart, B.Philips, K.Kloesel, K.Dougherty, H.Rodríguez, W.Díaz, W.Donner, J.Santos, &M.Zink. Incorporating

emergency management needs in the development of weather radar networks. Journal of Emergency Management(2009), 7(1), 45-52.

[41] L.A.Baumgart, E.J.Bass, B.Philips & K.Kloesel. Emergency management decision-making during severe weather. Weather and Forecasting(2008), 23(6), 1268–1279.

[42] C.E.League, W.Díaz, B.Philips, E.J.Bass, K.A.Kloesel, E.C.Gruntfest, & A. Gessner. Emergency manager decision-making and tornado warning communication. Meteorological Applications(2010), 17(2), 163-172.

[43] WebEOC, Manufactured by ESi Acquisition, Inc. http://www.esi911.com/home

[44] E-Teams, by NC4. http://www.nc4.us/ETeam.php

[45] National Emergency Management Network. http://www.nemn.net/.

[46] L. Li, and T. Li. An Empirical Study of Ontology-based Multi-document Summarization in Disaster Management. IEEE Transactions SMC: Systems, in press, 2013.

[47] L. Tang, T. Li, Y. Jiang, Z. Chen, Dynamic Query Forms for Database Queries, IEEE Transactions on Knowledge and Data Engineering(TKDE), 2013.

Disaster SitRep - A Vertical Search Engine and Information Analysis Tool in Disaster Management Domain

Li Zheng, Chao Shen, Liang Tang, Chunqiu Zeng, Tao Li,
Steve Luis, Shu-Ching Chen, Jainendra K. Navlakha
School of Computing and Information Sciences, Florida International University
{lzhen001, cshen001, ltang002, czeng001, taoli, luiss, chens, navlakha}@cs.fiu.edu

Abstract

*With the rise of heterogeneous information delivering platform, the process of collecting, integrating, and analyzing disaster related information from diverse channels becomes more difficult and challenging. Further, information from multiple sources brings up new challenges for information presentation. In this paper, we design and implement a **Disaster Situation Reporting System (Disaster SitRep)** that is essentially a disaster information collecting, integration, and presentation platform to address three critical tasks that can facilitate information acquisition, integration and presentation by utilizing domain knowledge as well as public and private web resources for major disaster recovery planning and management. Our proposed techniques create a disaster domain-specific search engine and a geographical information presentation and navigation platform using advanced data mining and information retrieval techniques for disaster preparedness and recovery that helps impacted communities better understand the current disaster situation. Specifically, hierarchical clustering with constraints are used to automatically update existing disaster concept hierarchy; taxonomy-based focused crawling component is developed to automatically detect, parse and filter those relevant web resources; a domain-oriented skeleton for each type of disasters is used to extract disaster events from disaster documents by defining the set of structural attributes. Furthermore, the platform can perform not only as a domain-specific search engine but also as an information monitoring and analysis tool for decision support during recovery phase of disasters.*

Keywords: Data Mining, Disaster Information Management, Vertical Search Engine, Concept Hierarchy, Focused Crawler, Disaster Event Extraction

1. Introduction

Natural or man-made hazardous disasters cause huge impact in business continuity activities. Thin margins and lack of a well-designed and regularly tested disaster plan make companies, particularly small businesses, especially vulnerable [1-2]. Our previous work [4-5] also demonstrates that building robust and intelligent disaster information extraction and analysis platform can help the public and private sectors work together to apply world class computing tools to deliver the right information to the right people at the right time.

Needs for heterogeneous information integration in disaster management domain: People have been firmly convinced that the use of timely, accurate and effective disaster information can significantly facilitate the disaster recovery process. Typical data resources include news/articles/blogs from web, announcements from governments, business reports from company participants, social media snippets and multimedia data like images and videos. However, information management and processing in disaster management are particularly challenging because of miscellaneous information resources that are publicly available and the unique combination of characteristics of those data, including: a great amount of information production and consumption; time sensitivity of the exchanged information; level of trustworthiness of the information sources; lack of common terminology; and heterogeneous formats [3]. However, very few information integration tools have been developed in disaster management tasks.

Growth of vertical search engine in various domains: General-purpose search engines, such as Google, Yahoo, or Bing have shown their efforts to exhaustively grasp all possible information from the giant web. However, the drawback of the general search strategy is the obviously overwhelming ambiguous and irrelevant information when digging in for a specific topic. Vertical search engine, also called domain-specific search engine, has been deemed as a powerful and necessary complementary tool to overcome those shortcomings. A well-established vertical search engine can substantially improve the efficiency of users getting more insights about a certain topic (in both coverage and relevance) and it also can

significantly save the cost for web crawlers in terms of time and storage.

Successful vertical search engines: There are quite a few successful vertical search engines currently serving various communities, such as Flight/Travel (SkyScanner), Law/Legal (FindLaw), BioInformatics (BioMed), and Academic Search (RefSeek). These search solutions focus on one area of knowledge creating customized search experiences and utilize existing knowledge from domain expertise. In disaster management domain, information collection and presentation platforms have also been implemented to gather disaster related information. For example, GeoVISTA [6] from Penn State created GeoTwitter [7] component to plot new tweets in real-time to support for situational awareness; OilReport [8] from Colorado University collected tweets related to oil spill and categorize those tweets by types of events. However, there is no previous work that can simultaneously handle information from heterogeneous resources (web, social media and government official reports, etc.) and systematically integrate resources together in disaster management domain.

1.1 Motivation for Integrated Disaster Information Analysis Tools

Our disaster management team at Florida International University has cooperated closely with experts and participants from South Florida Emergency Management and industry partnerships for over four years. We have designed and implemented a web-based prototype of disaster information sharing platform and a disaster situation awareness and community organization application running on iOS-based mobile devices utilizing the data processing power of advanced information technologies for disaster planning and recovery under hurricane scenarios [3-5]. They can largely help people discover, collect, organize, search and disseminate real-time disaster information.

This collaboration provides us with the opportunity to gain insight into the manner that South Florida public and private sector entities manage and exchange information in a disaster situation. The emergency managers and business continuity professionals eagerly desire that a powerful and intelligent data analysis system be specifically designed for disaster management domain and satisfy the information acquisition needs for different types of users including emergency management officers, business continuity participants, and other users without business recovery capabilities. They agreed that such a system can significantly help them facilitate their disaster management and recovery efforts. In order to efficiently and effectively deliver high-quality disaster related information, several interesting yet crucial information management issues have been brought up.

1. **Real-time disaster information.** People always prefer the latest news and situation reports related to their search interests when being affected by disasters. Timeliness is the most important requirement for information collecting and sharing system.

2. **Heterogeneous information resources.** News portal is no longer the only information sources in disaster situation. Some micro blog or social media applications make faster response when emergency happens. Information from various channels can largely accelerate the information discovery process.

3. **Diverse information presentation.** Textual results are no longer the standard information representing approach. News visualization methods focusing on combining different aspects or dimensions become more popular. For disaster events, geo-location is considered as one of the most important features for disaster preparedness and recovery, which can greatly improve information monitoring and organizing capabilities.

4. **Integrated information portal.** Users prefer information portal that offers a multitude of services to meet their information needs.

In summary, user requirements from professionals who have an operational responsibility in disaster situations have been converging gradually to a disaster information integration and analysis platform that is able to assimilate massive information and provide actionable information for decision support. Non-professional users also like a domain specific search tool that provides them with insight of disaster situations.

1.2 Research Challenges and Proposed Solutions

The following three key tasks have been identified to fully utilize the advantages and overcome the shortcomings of traditional general search and information management platform that have never been applied to disaster management domain.

1. Design and develop effective and dynamic concept hierarchy generation and reuse methods in disaster management domain to help the domain experts, the crawler and search engine behave efficiently in situation. Concept hierarchy, as means of formalizing and sharing knowledge, provides domain experts and knowledge engineers support for modeling specific domain of the world and can be applied in various areas to implement intelligent knowledge and information management system. However, building the hierarchy from scratch is a costly process that requires massive human labor, so automatically improving concept hierarchy generation and reuse becomes a challenging but critical task. Combining existing hierarchy with concepts extracted from Semantic Web contents largely helps to

extend and enrich existing structural concepts in a given domain.

2. Design and develop intelligent focused web crawling techniques to manage the data acquiring process and to increase the information coverage and relevance in disaster domain. Heterogeneous data collected from various sources bring difficulties to assimilate information at different levels. The strategies for general-purpose search engine will lead to many irrelevant web pages being indexed and also the seeds set will be expanded unexpectedly. Intelligent crawling strategies are needed to systematically control the crawling process to guarantee the indexed web contents with high quality and relevance. Also the given seeds can be expanded to a certain level and finally converge to a good seeds list. On the other hand, the query results are required to be personalized to remove duplicity and increase diversity.

3. Design and develop data integration techniques for disaster events identification and extraction. In disaster situation, many recovery processes are running in a confused mass. Undergoing activities and important situations are hard to detect from many information channels in unformatted patterns. How to understand the information and organize useful knowledge in a unified manner becomes especially helpful for government officials, disaster management agents, business continuity staff, and even public users suffer from disorders during disaster recovery phases. After getting related information from the web, particular techniques need to be designed to integrate the raw data into certain format that are ready to be used by the search engine and topic visualization modules.

Generally, to accomplish the above three tasks, Disaster SitRep utilizes the latest advances in database, data mining, and information extraction technologies, to create a user friendly, information-rich service web application in disaster management domain. In particular, to address dynamic concept generation task, we apply document hierarchical clustering with constraints to automatically expand and enrich the existing disaster concept hierarchy. To address focused crawling task, we create a focused crawler that is able to classify newly discovered web resources into different disaster concepts and also discover new concepts simultaneously. To address disaster data integration task, we use dictionary-based and rule-based named entity recognitions to extract disaster related events from massive document in heterogeneous formats.

In this paper, we design and implement a **Disaster Situation Reporting System (Disaster SitRep)** system. The rest of the paper is organized as follows. Section 2 presents the overview of Disaster SitRep system; Section 3 describes the disaster taxonomy generation in detail. Existing taxonomy is utilized as partially known concept hierarchy that can be used as constraints to update

taxonomy; Section 4 discusses the focused crawler. We propose a taxonomy-based focused crawling component to automatically detect, parse and filter those relevant web resources; New concepts can be extracted from crawled documents to update exiting taxonomy; Section 5 describes the disaster data integration module for assimilating miscellaneous resources; Section 6 describes the system evaluation.

2. Disaster SitRep Overview

Disaster SitRep is an integrated platform specializing in disaster management domain. It provides a collection of disaster related search, integration, and visualization tools to deliver personalized search results based on specific user needs. The goal to design this system is to help the user efficiently identify important information, organize emergent resources, and understand current damage and/or recovery status.
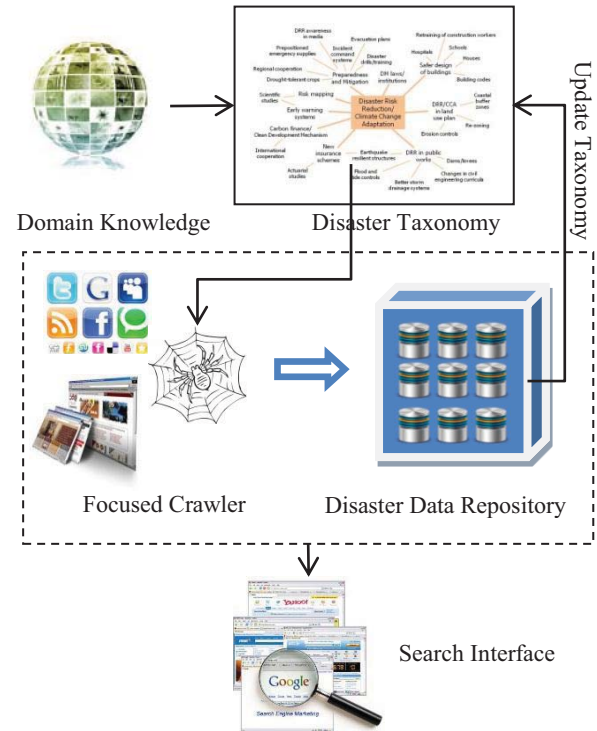
Figure 1. System architecture

Disaster SitRep has several major procedures to combine domain knowledge with disaster information from various resources and provide system users with high-quality disaster information related to their query interest. A hierarchy of disaster concepts is specified at the very beginning to help the focused crawler filter unrelated web pages. The focused crawler fetches web resources including web pages, documents and textual content from major social media. New concepts are generated after certain period of crawling and are used to

update previous concept hierarchy. At this time, heterogeneous data is integrated into a unified data repository for information search and visualization interface. Figure 1 illustrates the system architecture.

2.1 Disaster SitRep System Components

Based on the experience of developing our prototype we have designed two major components that are necessary for disaster information search and visualization:

used to give users a visual summary and corresponding textual summary [19, 20] of the disaster situation for corresponding query conditions.

We illustrate those important components in Figure 2.

2.2 Disaster SitRep System Architecture

Following our previous application framework, Disaster SitRep is designed and implemented to be a lightweight, comprehensive and fully Java implemented Web-based application. Our major information processing

| Search Results | Information Clusters – State Level | City Level -FL | City Level - NY |

System Integrated View

Figure 2. System components

Search Panel: A search component that supports keyword query provided by users returns a list of most important news crawled from various resources, including web, famous social media, and official government or company announcements and reports submitted to our previous business continuity web portal (www.bizrecovery.org). Results can be displayed based on different resource types or just in an integrated view.

News/Reports Map: Map in our system is used in two modes. Firstly, each of the search results associates with one or more points in map indicating the locations mentioned in the text. This allows users to visually know where those events happened and the geographic distributions of the events associated with a query. Secondly, the map provides a comprehensive view of all disaster information in our repositories. Also, such disaster information can be visually manipulated by utilizing different filters in map module. There are 4 filters including time, resource type, zoom level, and disaster type. Any combination of those filters can be

and representation functionalities are integrated with the following three critical modules: Taxonomy Generation, Focused Crawling and Disaster Event Extraction.

Taxonomy Generation: Based on our cooperation with domain experts, we initialize fundamental disaster taxonomy from disaster expertise. As the system keeps running, more web contents are crawled and extracted from unforeseen sources and new disaster terminologies are dynamically generated and are appended to the existing taxonomy. We propose a semi-supervised hierarchical clustering algorithm to enrich and modified previous taxonomy. Details of taxonomy generation and extension approaches are discussed in Section 3.

Focused Crawling: Our focused crawler is implemented to discover more disaster information by intelligently traversing the web contents based on their relevance to ongoing disasters. Usually, the more a web page is related to a certain topic, the higher probability it contains more resources (including hyperlinks to other web pages or possibly relevant concepts) in the same domain. The

disaster taxonomy in the previous stage can be utilized to classify web pages into various disaster categories. In general, there are two levels of judgments that help scoring the relevance of a page:

- *Web Page Classifier:* The classifier adopts hierarchical classification strategy to automatically categorize a crawled web page into different aspects according to the disaster taxonomy or simply report that current web page is irrelevant to any disaster topic.
- *Queue Prioritizer:* From the categorization results, the focused crawler adjusts the priority of each web page in the queue to guarantee that the most related web resource will be accessed earliest during the crawling process.

Combining these two functionalities, the focused crawling module attempts to assign the most relevant web page with the highest score to make sure such resource can be downloaded earliest. By properly designing those two parts, the crawler can access more related web resources by accessing fewer web pages. Also, as we crawl more disaster related content, it can largely contribute to extend our current taxonomy by including more concepts. Details are discussed in Section 4.

Disaster Event Extraction: Textual documents and situation reports crawled from the websites do not usually provide actionable information immediately, such as time, location, status, etc. The replication of information from various resources also challenges the search engine to provide highly related and diversified content to users. To gain further insight about the disaster event rather than a collection of textual documents, we need a domain-oriented skeleton for each type of disasters. The domain-oriented skeleton is the set of structural attributes that we try to extract from disaster documents. The details will be described in Section 5.

These modules are tightly integrated to provide a cohesive set of services including disaster information searching, querying, and visualization. Furthermore, they constitute a holistic effort on developing a data-driven solution for disaster management and recovery.

3. CONCEPT HIERARCHY GENERATION

Taxonomies or conceptual hierarchies play significantly important role in most knowledge-based information management systems applied in various application domains. They are designed to provide structurally organized terminologies that are formal, application-independent and with common agreement within a community of practice [9, 19]. However, generating taxonomy from the scratch suffers high-cost, low-efficiency problem. Ensembling several existing

taxonomies or incrementally integrating new concepts into existing taxonomy becomes effective and well-accepted approach for taxonomy generation and reuse. In our taxonomy generation component, we model this problem as **document hierarchical clustering with ordered constraints** in which the constraints are given as a partially known hierarchy, the disaster related concepts extracted from web documents are treated as instances, and our goal is to build a term hierarchy which satisfies the relative hierarchical structure in given partial hierarchy.

3.1 Base Concept Hierarchy Generation

Our initial disaster taxonomy is built manually from the scratch. Based on our long cooperation with Miami-Dade Emergency Operational Center (EOC), we extracted hundreds of frequent terms in its official announcements and situation reports in the past 5 years. We reasonably assume that those terms with high frequency indicate important concepts in disaster domain. Through careful filtering and organizing those terminologies from our staff and developers, our initial disaster taxonomy is obtained and then verified by our domain experts.
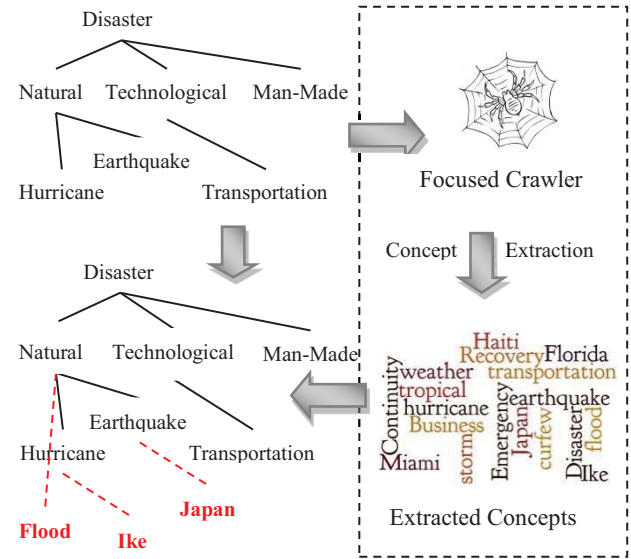
Figure 3. Iterative taxonomy generation

3.2 Iterative Taxonomy Generation

The disaster taxonomy generation process follows an interactive and iterative strategy. The focused crawler utilizes the taxonomy to classify accessed web pages and prioritizes those pages with highest relevance to disaster domain. From the repository of crawled data, high quality data will be analyzed and disaster-related concepts without being mentioned in the existing taxonomy, will

be extracted. Those extracted concepts are considered as highly popular terms that can extend and enrich the existing taxonomy. After integrating those newly-discovered concepts into disaster taxonomy, domain experts can verify the updated knowledge based and provide valuable feedback. Figure 3 shows the typical workflow of iterative taxonomy generation strategy.

3.3 Hierarchical Clustering with Constraints

Our aim is to build a hierarchical structure to model the basic human understanding of the relationships among disaster relevant concepts. A basic taxonomy/concept hierarchy is given at the very beginning of the generation process. In our work, we use agglomerative hierarchical clustering with constraint to algorithmically integrate newly-discovered terms or concepts into the existing ones.

3.3.1 Problem Definition. All concepts in existing taxonomy are denoted as $T = \{t_1, t_2, \cdots, t_n\}$ and the newly-discovered concepts are denoted as $C = \{c_1, c_2, \cdots, c_m\}$. H is the existing concept hierarchy formed by terms from T. Our goal is to generate an updated concept hierarchy H' that is formed by all terms from both T and C. The integration of T and C is non-trivial. There are three important aspects worth mentioning:
1. Each concept in *T* or *C* is represented by a set of terms extracted from the web documents repository. So, essentially there is a subset of web documents under each concept.
2. *H* is essentially a hierarchical clustering on all documents. The hierarchy of the concepts reflects the inclusion or exclusion of documents sets. There is no partial overlap between document sets under different concepts.
3. There is a merging preference/order for each pair of concepts in both H and H' which indicates the level of closeness between two document sets. The new concepts in *C* should not change the relative merging order of existing concepts in *T*. The details are given in the following section.

3.3.2 Algorithm and Partial Hierarchy Constraint. The merging preferences mentioned above are modeled as relatively ordered constraints when performing hierarchical clustering on document set. Constraints defined in hierarchical clustering are different from constraints, such as instance-level constraints [10] and prior knowledge [21] in partitional clustering. Several types of constraints that can be applied in hierarchical clustering are defined in the literature [11-13].

In our application, we use Bade's algorithm [11] to refine the given disaster concept hierarchy by considering further extracted concepts. The constraint in [11] is

named must-link-before (MLB), shown in Figure 4, which specifies the order in which objects are linked. When applied to concept hierarchy, such order indicates the merge preference between concepts (document sets). Bade's algorithm [11] can utilize the existing concept hierarchy as partially known hierarchy and update it by directly attaching newly-discovery concepts to previous hierarchy. The other two methods do not meet our needs because updated hierarchy requires to be built from the scratch.
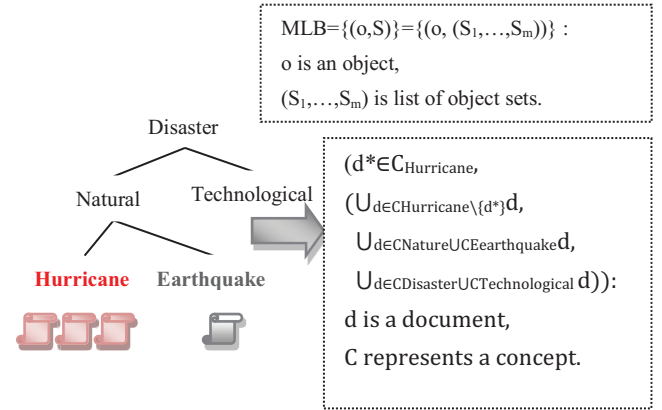
Figure 4. MLB constraints from partial hierarchy

4. FOCUSED CRAWLER

We adopt focused crawling technique to retrieve the disaster aware information in the Web. In addition, contents also come from subscription of some local news feeds and monitoring announcement from government sites. Compared with a standard focused crawler defined in [14, 15], there are some challenges in our problem.

Loose cohesion: Except large disaster events accompanied with intensive reports, most disaster information is scattered in the Web. In news websites, stories about disasters may embed in other types of news. As for government sites like county emergency management homepages, they are more likely to link to websites of the county's other departments than that of another county's emergency management homepage.

Diversity of disaster topic: Disasters we are interested in include many subtopics, from various types of disasters to four different phases of emergency management, it is difficult to evaluate a web page's relevance on a consistent scale among all these subtopics. It is very likely that the crawled data will bias towards some of the subtopics and leave some others uncovered.

To address the above issues, we utilize the concept hierarchy we developed.

4.1 Selection Strategy

Best-first approaches are widely used by focused crawlers, selecting the next page to be crawled from all currently assessed candidate page URLs by their scores as

$$l^* = \text{argmax}_{l \in queue}\ score(l)\ ,$$

where $score(l)$ is calculated based on a classifier indicating whether or not the URL l belongs to the topic. However, the "best" may bias to some of the subtopics of general disaster topic because of the unbalance of these subtopics and a limited initial training dataset. To get a set of web pages with high diversity for a specific disaster, we simultaneously crawl web pages for each disaster concept based on the concept hierarchy. Our selection strategy considers a disaster concept:

$$l_C^* = \text{argmax}_{l \in queue}\ score(l, C),$$

that is, for each disaster concept, select the next page to be crawled from all currently assessed candidate page URLs according to their scores with respect to the concept.

Figure 5. An example page of hurricane Irene.

4.2 Prioritization Based on Concept Relationship

For a web page, instead of classifying it into "Disaster" and "Non-disaster", we assigned to it a concept in our concept hierarchy, such as "weather", "government" and "environment protection". These disaster related concepts increase the coherence of the Web pages of disaster topic, playing a role of bridging between pages of different sites of disaster concepts and pages of different disaster concepts. To calculate the prioritization score of a URL, the concept of the page from which the URL is linked is utilized as follows:

$$score(l, C_d) = P(C_i^* \to C_d) * P(page_l = C_i^*),$$

where $P(page_l = C_i^*)$ is the output of our content classifier indicating the probability the page where the link l is linked from belongs to its optimal concept C_i^*, and $P(C_i \to C_d)$ is the link relationship between concepts,

the probability that a page of concept C_i links to a page of concept C_d. It can be calculated as

$$P(C_i \to C_d) = \frac{\sum_{p \in C_i} |L_{p,fetched} \cap C_d| + \lambda}{\sum_{p \in C_i} |L_{p,fetched}| + \lambda \sum_{p \in C_i} |L_{p,unfetched}|}\ ,$$

the ratio of the number of links classified as C_d from pages of C_i to the number of all fetched links from pages of C_i, with a Dirichlet smoothing using unfetched links. Note that with the process of crawling, $P(C_i \to C_d)$ is being updated, so that the scoring of links is also adaptive with more data crawled.

4.3 Link Prediction

Although a page is disaster relevant, the links of the page may not necessarily lead to other pages of disasters. Figure 5 shows an example page.

To further distinguish the links in a page, a link classifier is trained, using the prediction of the content classifier for crawled pages as training data. The rationale is that many links contain a description of the content of the linked page. Another observation we find is about link structure, that for a pair of link which are in the sibling nodes of the HTML DOM tree, e.g. in a list of the page, they tend to be of a similar topic. We follow the work of [16] and build a link classifier based on Native Bayes. To apply the link prediction:
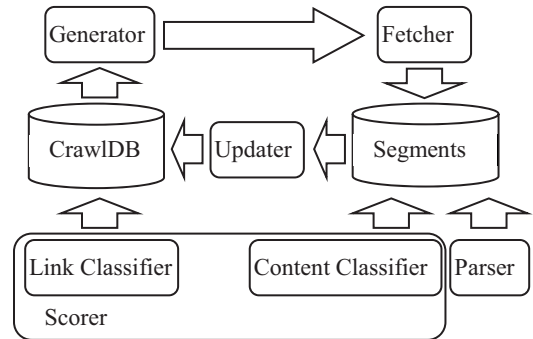
Figure 6. Architecture of the focused crawler.

1. The prioritization score can be extended as:
$$score(l, C_d) = P(C_i^* \to C_d) * P(page_l = C_i^*) * P(C_d|l)\ ,$$
where $P(C_d|l)$ is the output of the link classifier, probability that link l leads to a page under concept C_d.

2. To reduce the redundancy, we first divide the links into clusters, and constrain the crawler such that links in the same cluster are not fetched at same time. Once a link is fetched, the prediction of links in the same cluster will be updated.

4.4 Architecture of the focused crawler

We build our crawler based on Nutch[17], which is a distributed general crawling tool running on Hadoop[18]

clusters. We customize the scoring module and generator module in Nutch. The current architecture is shown in Figure 6. In each iteration, the Fetcher fetches page content of a list of URLs, and stores them as a segment. The updater updates CrawDB, where the crawled data is associated with a URL. The scoring module assigns a prioritization score to each URL indicating the importance of the URL. The generator module generates a set of URL, covering all disaster concepts in the concept hierarchy. The Fetcher fetches the web page content.

5. DISASTER EVENT EXTRACTION AND INTEGRATION

We hope to gain further insight about the disaster event rather than create a collection of textual documents. For example, the location and the date time of a storm, the status of the electrical power impacted by an earthquake and so on. These are key domain-oriented information of disasters. In our vertical search engine, the rank of the search results is mainly based on this domain-oriented information. On the other hand, when a disaster happens, a huge amount of news, situation reports, and announcements will burst in a very short time. Most of the documents have replicated content. To eliminate replicated content and integrate all related documents, we need a domain-oriented skeleton for each type of disaster. The domain-oriented skeleton is the set of structural attributes that we try to extract from disaster documents.
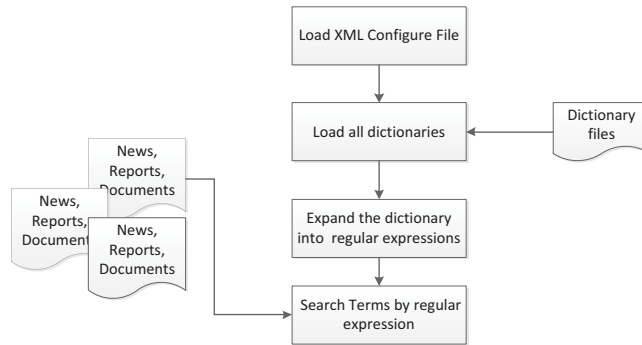
Figure 7. Workflow of event extractor

5.1 Event Extractor

Dictionary-based and Rule-based named entity recognitions are utilized in this part. We do not use probabilistic and training based name entity recognition because that approach requires people to label the web documents word by word, which is time-consuming. Each domain-oriented disaster attribute (e.g. the date time, the location, the power status) is defined as a rule in a XML configuration file. The rule is similar to the regular expression but combines more functionality. It can

include a dictionary given by the user and supports approximate word matching by edit distance.

For each concept, we have particular event attributes for extraction. The event attributes are also associated with the extraction rules in the configuration file. Each concept owns the attributes of its ancestors. The work flow for event extractor is shown in Figure 7.

5.2 Event Integration

The disaster event integration is based on the similarity table join. We consider each extracted disaster event as a database record. Each web document corresponds to a database record. The problem of event integration is how to join these database records. In our disaster event integration, we apply the similarity join, which does not require two identical attribute values. It considers the overall similarity of the values of common attributes. If common attribute values have a similarity greater than a threshold, the two records can be joined. If the threshold is 1, the similarity join becomes the traditional equal join.

Let a disaster event consist of n attributes, x and y be two extracted event records. x_i denotes the i-th attribute value of x, i=1,…n. The overall similarity for record joining is defined as follows:

$$\text{sim}_{all}(x, y) = \frac{\sum_{i=0}^{n} I(i,x)I(i,y)w_i f_i(x_i, y_i)}{\sum_{i=0}^{n} I(i,x)I(i,y)w_i}$$

where I(i,x) is a binary variable that I(i,x)=1 if x has the i-th attribute value, otherwise I(i,x)=0. w_i is the weight for the i-th attribute. f_i is the predefined similarity function for the i-th attribute of the event.

Figure 8 shows an example of the joining two records which are extracted from two different web news. W1 is a situation report from an official web site such as FEMA, and W2 is a news report from a local news website. The common attributes of W1 and W2 have identical values, so they are joined to a record W*.

W1

Disaster	Datetime	Disaster Category	Geo-Location	HasFlooding
Irene	08/21/2011	Hurricane	West Virigina	Yes

W1

Disaster	Datetime	Disaster Category	Geo-Location	Death
Irene	08/21/2011	Hurricane	West Virigina	4

W*   Join

Disaster	Datetime	Disaster Category	Geo-Location	Death	HasFlooding
Irene	08/21/2011	Hurricane	West Virigina	4	Yes

Figure 8. Event record join

Many existing studies on web documents apply textual clustering based methods. Then, the information shown to the users is the representative document of each cluster.

This approach is not quite accurate because it does not consider the content meaning. For instance, it does not distinguish the name of the disaster, the date time of the event, and the geo-locations and other key attributes of disaster reports with other document words. Therefore, our disaster event integration is based the extracted events rather than the original textual documents.

6. DATA COLLECTION AND EVALUATION

To evaluate the algorithms used in our system we use standard performance metrics used in the research literature and carefully compare our algorithms with existing work when applicable.

Our system evaluation process consists of presenting the system to our community of emergency managers, business continuity professionals and other stakeholders for feedback and performing community exercises. The community exercises involve a real time simulation of a disaster event and are integrated into an existing exercise that the community conducts for readiness each year. This evaluation exposes information at different time intervals and asks the community to resolve different scenarios by using the tool developed. The evaluation conducted takes on the form of a "table-top" exercise in which information injects provide details about the current disaster situation and specify potential goals and course of action. In return, the participant uses the system to gather information to best assess the situation and provide details about the actions to be taken. We gather information from the user about what information they found to derive their conclusions or lack thereof. This information allows us to better understand how our techniques overall improve the information effectiveness.

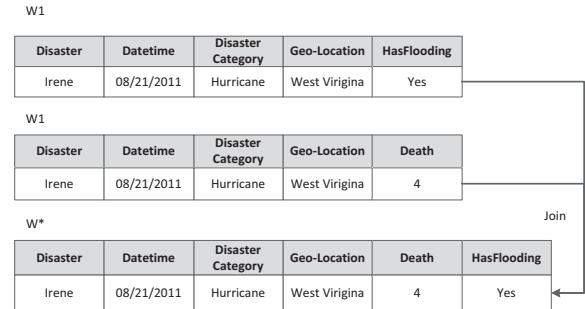Feedback from our users are overwhelmingly positive and suggest that our system can be used not only to share the valuable actionable information but to pursue more complex tasks like business planning and decision making. There are also many collaborative missions that can be undertaken on our system which allows public and private sector entities to leverage their local capacity to serve the recovery of the community. Our initial work has been recognized by FEMA (Federal Emergency Management Agency) Private Sector Office as a model in assistance of Public-Private Partnerships.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] The Conference Board. Preparing for the worst: A guide to business continuity planning for mid-markets. *Executive Action Series*, February 2006.

[2] R. Berg. Hurricane Ike Tropical Cyclone Report. NHC. Retrieved 2009-09-12.

[3] V. Hristidis, S. Chen, T. Li, S. Luis, and Y. Deng. Survey of data management and analysis in disaster situations. *The Journal of Systems and Software*, 83:1701–1714, 2010.

[4] L. Zheng, C. Shen, L. Tang, T. Li, S. Luis, S. Chen, and V. Hristidis. Using data mining techniques to address critical information exchange needs in disaster affected public-private networks, *KDD '10*, pages 125–134, 2010.

[5] L. Zheng, C. Shen, L. Tang, T. Li, S. Luis, and S. Chen. Applying Data Mining Techniques to Address Disaster Information Management Challenges on Mobile Devices. *In Proceedings of KDD 2011*, pages 283–291.

[6] GeoVista. http://www.geovista.psu.edu/.

[7] A. M. MacEachren, A. C. Robinson, A. Jaiswal, S. Pezanowski, A. Savelyev, J. Blanford. Geo-Twitter Analytics. *Applications in Crisis Management 2011*. Paris, France.

[8] OilReport. http://www.cs.colorado.edu/~starbird/oilreport_map.html.

[9] E. Simperl: Reusing ontologies on the Semantic Web: A feasibility study. *Data Knowl. Eng*. 68(10): 905-925 (2009).

[10] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl. Constrained K-means Clustering with Background Knowledge. *In Proceedings of ICML 2001*, pages 577-584.

[11] K. Bade and A. Nrnberger. Creating a cluster hierarchy under constraints of a partially known hierarchy. *In SDM 2008*, pages 13–24.

[12] L. Zheng and T. Li. Semi-supervised Hierarchical Clustering. *In ICDM 2011*, pages 982 - 991.

[13] H. Zhao and Z. Qi. Hierarchical agglomerative clustering with ordering constraints. *In WKDD 2010*, pages 195–199.

[14] S. Chakrabarti, M. van den Berg, B. Dom. Focused Crawling: A New Approach to Topic Specific Resource Discovery. *In WWW 1999*.

[15] C. C. Aggarwal, F. Al-Garawi, and P. S. Yu. Intelligent crawling on the World Wide Web with arbitrary predicates. *In WWW '01*. ACM, 2001.

[16] D. Ahlers and S. Boll. 2009. Adaptive geospatially focused crawling. *In Proceeding of the 18th ACM conference on Information and knowledge management*, pages 445–454.

[17] Nutch. http://nutch.apache.org/

[18] Hadoop. http://hadoop.apache.org/

[19] L. Li, D. Wang, C. Shen, and T. Li. Ontology-Enriched Multi-document Summarization in Disaster Management. *In Proceedings of SIGIR 2010*.

[20] D. Wang, L. Zheng, T. Li, and Y. Deng. Evolutionary Document Summarization for Disaster management. *In Proceedings of SIGIR 2009*.

[21] T. Li, Y. Zhang, and V. Sindhwani. A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge. In Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics (ACL 2009), Pages 244-252, 2009.

Ontology-enriched Multi-Document Summarization in Disaster Management

Lei Li, Dingding Wang, Chao Shen, Tao Li
School of Computing and Information Sciences
Florida International University
Miami, FL 33199
{lli003, dwang003, cshen001, taoli}@cs.fiu.edu

ABSTRACT

In this poster, we propose a novel document summarization approach named Ontology-enriched Multi-Document Summarization(OMS) for utilizing background knowledge to improve summarization results. OMS first maps the sentences of input documents onto an ontology, then links the given query to a specific node in the ontology, and finally extracts the summary from the sentences in the subtree rooted at the query node. By using the domain-related ontology, OMS can better capture the semantic relevance between the query and the sentences, and thus lead to better summarization results. As a byproduct, the final summary generated by OMS can be represented as a tree showing the hierarchical relationships of the extracted sentences. Evaluation results on the collection of press releases by Miami-Dade County Department of Emergency Management during Hurricane Wilma in 2005 demonstrate the efficacy of OMS.

Categories and Subject Descriptors: H.3.3[Information Storage and Retrieval]: Information Search and Retrieval

General Terms: Algorithms, Experimentation, Performance

Keywords: Ontology, Multi-Document Summarization, Disaster Management

1. INTRODUCTION

Ontology is a philosophy concept, dealing with questions about what entities exist or can be said to exist, and how such entities can be grouped within a hierarchy, and subdivided according to their similarities and differences. Ontology has been applied in many research areas in information retrieval, particularly, in text mining. For example, D. Sánchez et al use the ontology to compute semantic similarity [1], and I. Yoo et al utilize the ontology to improve document clustering [2]. However, relatively few research efforts have been reported on using the ontology for improving document summarization.

Generally, given a query, multi-document summarization is the process of generating a query-focused/relevant condensation (i.e., a generated summary) of the content of the entire input set. Existing summarization methods usually rank the sentences in the documents according to their scores calculated by a set of predefined features, such as term frequency-inverse sentence frequency (TF-ISF), sentence or term position, and number of keywords [3]. The above analyses are difficult to capture the hidden semantic relationships between the sentences and queries. Ontology, with abundant concise concepts and rich domain-related information, can capture the hidden semantic information.

In this poster, we develop a novel method, OMS, to generate query-relevant summary from a collection of documents by making use of the ontology. In particular, OMS first links the sentences of documents being considered onto a domain-related ontology, then maps the given query to a specific node in the ontology, and finally extracts the summary from the sentences in the subtree rooted at the corresponding query node, by using FGB, a text summarization approach described in [4]. As a byproduct, the summaries we finally acquire can be represented as a tree showing the hierarchical relationships of the extracted sentences.

We apply OMS to disaster management for evaluation. For natural calamities, such as hurricanes and earthquakes, vast amount of related news and reports are generated through time for broadcasting and recording events. Experimental results on such disaster management demonstrate the efficacy of OMS.

2. FRAMEWORK OVERVIEW

Figure 1 shows the framework of OMS. First of all, a domain specific ontology hierarchy was created by domain experts to describe concepts appearing in disaster related document sets. Given a collection of documents related to disasters, we disassemble them into a set of sentences. Then we map these sentences onto the ontology hierarchy based on their semantic correlations, discarding some sentences not relevant to any concept in the ontology, and omitting from the ontology some concepts with less importance.

Up to this point, we obtain an ontology-sentence tree representation in which each node is linked by a set of relevant sentences. Given a query q, OMS links it to a specific node i in the ontology-sentence tree according to the semantic relationship between the query and the nodes, and extracts a sub-hierarchy rooted at i. Then the FGB model [4] is applied to summarize sentences linked to each node in this subtree. Finally, a summary tree that satisfies the query q is achieved.

Ontology Refinement: In the ontology hierarchy, a myriad of concepts relevant to disaster management documents are specified by domain experts. Given a subset of documents, some concepts may not be predominant or even not appear in this documents set. We need to refine the ontology hierarchy so that the ontology-sentence tree representation can better reflect the subject of the document set.

To do so, we first rank nodes at the same level in order of sentence counts, then ignore the nodes and their subtrees with sentence counts less than the average. By iteratively running the above procedure in a top-down manner, a thematic ontology-sentence tree representation is generated.
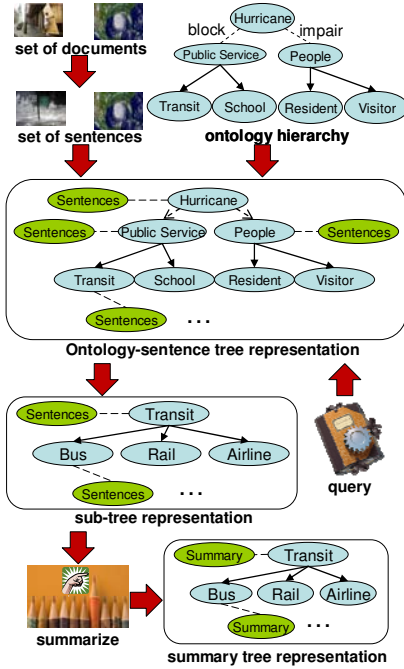
Figure 1: *OMS* Framework

3. EXPERIMENTS

3.1 Real World Data

The document set used in our experiments is a collection of press releases from Miami-Dade County Department of Emergency Management and Homeland Security during Hurricane Wilma from Oct. 19, 2005 to Nov. 4, 2005. It contains approximately 1,700 documents, about half of which contain similar contents. We randomly select 100 documents from this document set as our experiment data.

Data Preprocessing: For the sake of the theme embodiment of ontology-sentence tree representation, some sentences are removed from the whole sentences set in the procedure of sentence mapping, since they have no semantic relationship with any ontology concept node. For example, "For those outside of Miami Dade County can call (305) 468-5900 to reach the Answer Center." describes the phone number of the Answer Center; however, this kind of sentences repeatedly appears in most disaster related documents we are considering, and they have no specific meaning for summarization.

3.2 An Illustrative Case Study

In order to illustrate the interpretability of our proposed method, we provide an example of queries and the corresponding result generated by *OMS*. Figure 2 demonstrates this case study.

Given the query "get all the information related to *transit* in Miami-Dade County after Hurricane Wilma passed", the result is represented as a summary tree in Figure 2, in which
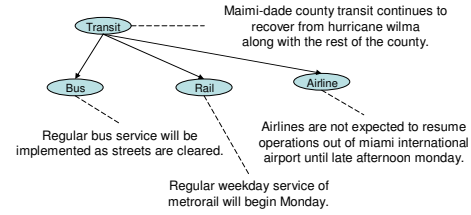
Figure 2: Summary tree related to *transit*

the topics in eclipses are domain concepts, and the sentences linked to them are summaries correlating to concepts. We observe that (1) every summary in the summary tree concretely reflects the status of relevant concept; (2) summaries generated by *OMS* exhibit apparent semantic hierarchy.

3.3 Performance Evaluation

For performance evaluation, 42 queries related to some specific concepts in the ontology are designed by domain experts for experimentation. In order to evaluate the quality of the generated summaries by *OMS* and other methods, we use human generated summaries as references. For hurricane data, we hire 5 human labelers to manually create summaries based on the selected document set and the given queries. Then we run *OMS* on such document set; meanwhile, we apply FGB [4] to automatically summarize query-relevant documents without using ontology. The summarization results are evaluated by ROUGE, a document summarization evaluation tool described in [5]. The experiment results are shown in Table 1. The results indicate that summarization efficacy is significantly improved by adopting ontology.

Measure	Using Ontology			Without Ontology		
	AVG-R	AVG-P	AVG-F	AVG-R	AVG-P	AVG-F
ROUGE-1	0.83236	0.75748	**0.78326**	0.56601	0.47562	0.50340
ROUGE-2	0.77324	0.72221	**0.74223**	0.43841	0.40197	0.41570
ROUGE-L	0.82032	0.75032	**0.77512**	0.53661	0.46032	0.48460
ROUGE-S	0.76362	0.68136	**0.70562**	0.43231	0.37311	0.39071
ROUGE-SU	0.77299	0.68715	**0.71202**	0.44875	0.38025	0.39913

Table 1: Summarization results comparison between *OMS* and FBG without using ontology. Remark: We use ROUGE-N, ROUGE-L, ROUGE-S and ROUGE-SU, and compare F-scores of the two different methods.

Acknowledgements

4. REFERENCES

[1] D. Sáchez, M. Batet, A. Valls, and K. Gibert. Ontology-driven web-based semantic similarity. *Intelligent Information Systems*, October 2009.

[2] I. Yoo and X. Hu. Clustering large collection of biomedical literature based on ontology-enriched bipartite graph representation and mutual refinement strategy. PAKDD, 2006.

[3] D. Jurafsky and J.H. Martin. *Speech and Language Processing*. Pearson, second edition, 2008.

[4] D. Wang, S. Zhu, T. Li, Y. Chi, Y. Gong. Integrating Clustering and Multi-Document Summarization to Improve Document Understanding. CIKM, 2008

[5] C. Lin. Rouge: A package for automatic evaluation of summaries. Post-Conference Workshop of ACL, 2004.

Web Multimedia Object Classification using Cross-Domain Correlation Knowledge

Wenting Lu, Jingxuan Li, Tao Li, Weidong Guo, Honggang Zhang, and Jun Guo

Abstract—Given a collection of web images with the corresponding textual descriptions, in this paper, we propose a novel cross-domain learning method to classify these web multimedia objects by transferring the correlation knowledge among different information sources. Here, the knowledge is extracted from unlabeled objects through unsupervised learning and applied to perform supervised classification tasks. To mine more meaningful correlation knowledge, instead of using commonly used visual words in the traditional bag-of-visual-words (BoW) model, we discover higher level visual components (words and phrases) to incorporate the spatial and semantic information into our image representation model, *i.e.*, bag-of-visual-phrases (BoP). By combining the enriched visual components with the textual words, we calculate the frequently co-occurring pairs among them to construct a cross-domain correlated graph in which the correlation knowledge is mined. After that, we investigate two different strategies to apply such knowledge to enrich the feature space where the supervised classification is performed. By transferring such correlation knowledge, our cross-domain transfer learning method can not only handle large scale web multimedia objects, but also deal with the situation that the textual descriptions of a small portion of web images are missing. In addition, our proposed method can apply in both specific domain (*e.g.*, disaster emergency management) and general domain (*e.g.*, social-media images organization). And empirical experiments on two different datasets of web multimedia objects are conducted to demonstrate the efficacy and effectiveness of our proposed cross-domain transfer learning method.

Index Terms—Multimedia Object Classification, Cross-Domain, Correlation Knowledge, Transfer Learning, Bag-of-Visual-Phrases Model

I. INTRODUCTION

A. Web Multimedia Object Classification

With the rapid development of the Web 2.0 technologies and the popularization of digital devices of fast access to the Internet, people tend to share images on the web, and consequently enormous amounts of images are posted online everyday. Besides the great convenience of the information access, how to efficiently retrieve images that satisfy the need of web users in multimedia databases is becoming more and more difficult and challenging. Accordingly, the multimedia object

classification, as a crucial step of multimedia information retrieval, has found many applications such as indexing and organizing web multimedia databases, browsing and searching web multimedia objects, multimedia information delivery and discovering interesting patterns from objects [1]. Given a collection of web images with the corresponding textual descriptions, as shown in Fig. 1 where the auxiliary textual information is often provided by web users to describe the general contents of images (*i.e.*, image titles, headers and tags), a challenging question is how to perform classification tasks by leveraging both the image features and the textual information.
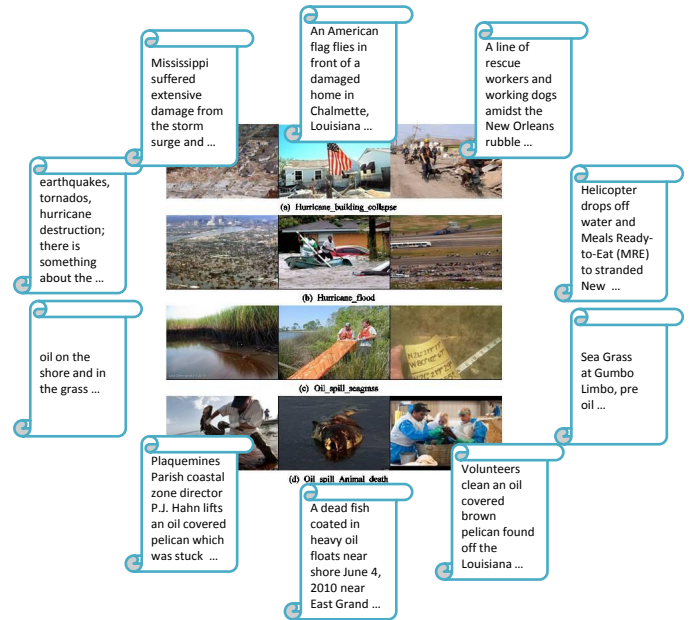
Fig. 1. Example web images and their corresponding textual descriptions. Each row represents a class, *i.e.*, a topic-related event.

As we all know, classic supervised learning algorithms for classification aim at training suitable classifiers using labeled data samples. However, for dramatically increased web data, although the above textual information is often provided by web users, true labels of web images for classification is usually difficult and expensive to obtain. Moreover, the testing data may not follow the same class labels or generative distributions as the labeled training data. The above two limitations may be addressed by taking the learning process into consideration. In the process of the human learning, besides learning the samples from new learning task, we usually apply our knowledge and experiences obtained in

W. Lu and W. Guo are with Capital University of Economics and Business, Beijing, 100070 P.R.China. E-mail:{luwt, guowd}@cueb.edu.cn, Tel:+86-13522870153.

J. Li and T. Li are with School of Computing and Information Sciences, Florida International University, Miami, Florida 33199, United States. E-mail:{jli003, taoli}@cs.fiu.edu, Tel:+1-(305)348-1218, Fax: +1-(305)348-3549.

H. Zhang and J. Guo are with School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876 P.R.China. E-mail:{zhhg, guojun}@bupt.edu.cn, Tel:+86-(010)-62283059-1003.

the previous learning processes to facilitate the new learning task. This principle is also applicable in machine learning process. For example, the robot might be easier and more quickly to recognize "Toyota sedan" if it already knows how to identify "Ford SUV" and "Benz sedan"; Previous experiences in playing violin might provide some guidance to learn playing piano and sachs. Therefore, maybe we could explore the feasibility of using some knowledge as a "guidance" for web multimedia object classification by transfer learning methods, which could achieve better classification performance comparing with existing approaches.

B. Content of the Paper

Inspired by the above observation, in this paper, **we propose a cross-domain transfer learning method for utilizing web multimedia objects without true labels in performing supervised classification tasks.** Such objects without true labels is relatively easier to obtain from web, and could be used to extract knowledge through the unsupervised learning. Different from the previously proposed multi-view learning methods [2], [3], [4], [5] which process each information source separately and then combine them together at either feature level, semantic level, or kernel level, we focus on cross-domain knowledge transfer where we first discover the correlation knowledge between different domains (*i.e.*, text domain and image domain) and then apply it to classification. In particular, in order to discover the correlation knowledge, we consider the semantic correlation among the attributes of the data concepts and aim at finding out the "correlation" between the images and their corresponding textual descriptions at the semantic level. Therefore, we firstly need to adopt more meaningful and effective models to represent images and texts respectively. In this case, for text, the textual words with specific meanings (*e.g.*, house, water, face, and etc.) are good choices to represent various textual descriptions of web images. For image, however, it is originally modeled as a bag of visual words, which contain very limited semantic information because each region/patch represented by a visual word might come from different parts of the object. For example, as shown in Fig. 2, visual word *A* can not distinguish the American flag from the striped shirt and the adidas sport shoes, as they share visually similar red and white stripes. However, the combination of visual word A and B, *i.e.*, the visual phrase *AB*, can effectively distinguish the American flag from the other two. (Here, the definition of "phrase" is similar with the definition of that in [6], [7], [8], [9].) Therefore, we incidentally propose a novel image representation model named bag-of-visual-phrases (BoP) to represent images in a more meaningful and effective way. Specifically, we firstly obtain visual words via the hierarchical clustering, and generate visual phrases by discovering the spatial co-location patterns of visual words, then add these generated phrases into the vocabulary. In this way, the spatial and semantic distinguishing power of image features can be enhanced.

Given the enriched visual vocabulary as well as the textual words, we calculate the frequently co-occurring pairs among them to construct a cross-domain correlated graph in which
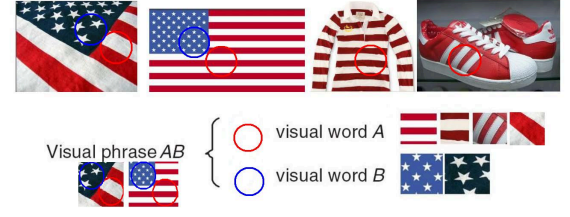
Fig. 2. An example of visual phrase.

the correlation knowledge, *i.e.*, a set of strongly correlated groups (*e.g.*, the collapsed buildings in image patches as well as the textual words "house" and "damaged" in Fig. 11), is mined by a backtracking algorithm [10]. After that, in order to evaluate our proposed cross-domain learning method, we apply the above knowledge to perform the task of web multimedia object classification. Here, we investigate two different strategies (*Enlarging Strategy* and *Enriching Strategy*) to apply the correlation knowledge to enrich the feature space for classification.

Our method can be viewed as a form of transfer learning [11], [12], because the discovered cross-domain correlation knowledge provides "guidance" to learning tasks which can be facilitated especially when the size of available labeled training samples is not large enough. However, different from heterogeneous transfer learning [13], translated learning [14] and SocialTransfer [15] which are designed for learning across different feature spaces by using labeled data from one feature space to enhance the learning tasks of other feature space, our method aims at discovering the semantic correlation between different domains via unsupervised learning on unlabeled data from multiple feature spaces and then applying such knowledge to facilitate the classification. Empirical experiments on two different datasets of web multimedia objects are conducted to demonstrate the efficacy and effectiveness of our proposed cross-domain transfer learning method.

C. Paper Contribution and Organization

In summary, the contribution of this paper is four-fold:

- **An unsupervised learning method for discovering cross-domain correlation knowledge**: We propose a novel cross-domain method to first discover the correlation knowledge between different domains via unsupervised learning on unlabeled data from multiple feature spaces, and then apply it to perform supervised classification tasks. To demonstrate the efficacy of our proposed method, compared with the classification results provided by three existing multi-view learning methods for classification mentioned in the first part of Section II, our proposed cross-domain transfer learning method outperforms the others on two different datasets of web multimedia objects.
- **A novel two-level image representation model**: Unlike the traditional bag-of-visual-words (BoW) model, in order to represent images in a more meaningful and effective way, we incidentally propose a novel image representation model (*i.e.*, bag-of-visual-phrases (BoP))

to represent images at both the word-level and the phrase-level. Here, these phrases are higher level visual components which could incorporate the spatial and semantic information into our image representation model to improve the performance of the basic BoW model. According to the experimental validation in Section V, once we represent images with both visual words and visual phrases, the classification accuracy can be improved, which shows that by combining visual words with phrases, the distinguishing power of image features is enhanced.

- **Two different strategies for correlation knowledge utilization**: In the stage of knowledge utilization, we investigate two different strategies (*Enlarging Strategy* and *Enriching Strategy*) to utilize the correlation knowledge to enrich the feature space for classification. By transferring such correlation knowledge, both the strategies can handle the situation when one information source is missing, especially the most common situation that the textual descriptions of a small portion of web images are missing. Empirical experiments demonstrate, comparing with the base line classification methods without using such knowledge, the classification results of our proposed two correlation knowledge based classification methods are significantly improved; and the *Enriching Strategy* performs better than the *Enlarging Strategy*.

- **A wide variety of new applications**: By effectively transferring the cross-domain correlation knowledge to new learning tasks, our proposed method can ont only be applied in some specific domain (*e.g.*, disaster emergency management), but also be used in general domain (*e.g.*, social-media images organization, etc.).

The rest of this paper is organized as follows. Section II reviews some related work. Section III presents the algorithmic details of the BoP model and the cross-domain transfer learning method. Section IV describes two novel applications based on the cross-domain transfer learning. Section V provides a detailed experimental evaluation and analysis. Finally we conclude the paper in Section VI.

II. RELATED WORK

The previous work related to this paper involves the following three aspects:

A. *Learning from Multiple Information Sources*

In many real-world applications, the data is naturally multimodal, in the sense that they are represented by multiple sets of features. Therefore, many research publications [2], [3], [4], [16], [5], [17] aim at designing multi-view learning algorithms to learn classifiers from multiple information sources via integrating different types of features together to perform classification. Besides classification, multi-view learning algorithms are also exploited and applied in the field of multimedia retrieval and automatic image annotation [5], [18], [19], [20], [21].

In general, from the information fusion perspective, most of the above multi-view learning methods can be categorized into three different groups based on the way that information from different sources is used.

(1) Feature Integration:
This approach enlarges the feature representation to incorporate attributes from various sources and produces a unified feature space. In particular, continuous data types will be converted into discrete levels and categorical data type will be mapped into similar discrete levels. The data are then transformed into the same features space and standard computational methods, such as prediction and clustering, can be performed. The advantage of feature integration is that the unified feature representation is usually more informative and allows many different data mining methods to be applied and systematically compared. One disadvantage is the learning complexity and difficulty are increased as the data dimension becomes large [2].

(2) Semantic Integration:
This approach keeps data intact in their original form. Computational methods are applied to each dataset separately. Results on different datasets are then combined by either voting [22], Bayesian averaging [23] , or the hierarchical expert system approach [24]. One advantage of the semantic integration is its ability of implicitly learning the correlation structure among different sets of features [4].

(3) Kernel Integration:
This approach can be viewed as a compromise between the feature integration and the semantic integration. Data sources are kept in their original form and they are integrated at the similarity computation or the kernel level [25]. Take the integration between two information sources for example, for data objects p_i and p_j, their total pairwise similarity or affinity is $S_{ij} = A_{ij} + B_{ij}$, where A_{ij} is computed from one information source and B_{ij} is obtained from another information source. Different weights can be applied to different data sources. In general, when a dataset has a varying local data distributions, method (3) is preferred, while method (2) is more stable especially when the size of the training dataset is small [26].

Different from the above three groups of methods which firstly process each information source separately and then combine them together at different levels, our method firstly discover the semantic correlation between different sources as knowledge and then apply it to facilitate new learning tasks.

B. *Learning with Knowledge Transfer*

In recent years, some works have been reported on incorporating the prior knowledge and experience for better learning new tasks [11], [12], [27], or learning across different feature spaces by using data from one feature space to enhance the learning tasks of other feature spaces [13], [14], [15]. In these methods, the prior knowledge and experience obtained from the source data can provide "guidance" to perform new learning task on target data such that the learning can be facilitated especially when the size of available labeled training samples is not large enough.

Different from the transfer learning [11], the knowledge extraction of our method is done through unsupervised learning. From this point of view, our method is closer to the self-taught learning [12] and the knowledge transfer [27]. However,

the knowledge extraction and utilization of our method is far different from those in [12]. And compared with [27], our method is a cross-domain learning approach involving both the image and the text, while [27] deals with the textual information only. Moreover, in order to extract more meaningful correlation knowledge, we identify the correlated pairs at both the word-level and phrase-level rather than the single word-level as in [27]. By doing this, the semantic distinguishing power of the correlated groups can be enhanced greatly. In addition, our method differs from heterogeneous transfer learning [13], translated learning [14] and SocialTransfer [15] in that the semantic correlation between different domains is discovered by using unlabeled data through unsupervised learning from multiple feature spaces and then applied such correlation knowledge to facilitate the classification, but not from one feature space to another feature space.

C. Image Representation Model

Existing methods for image representation can be roughly categorized into two groups: *global* representation and *local* representation. Typical global representation models (*e.g.*, color, texture, edge, shape, and etc.) are well accepted for simplicity. But in the visual object categorization system, global representation models may not work well as they are sensitive to viewpoint, object occlusion, and background clutter with respect to visual object categorization. On the contrary, by using local representation models, most of the local features are invariant to the scale, rotation, translation, affine transformation, and illumination changes [28]. Since the total number of local features varies in different images, these models may not be applicable to the follow-up processes. One classic solution is to represent local features of images with BoW model [29], [30] and *tf-idf* schema [31], and then perform classification. The basic idea of the BoW model for the image representation is to first group keypoints of images into a large number of clusters based on their similarities. Each cluster is then treated as a visual word that represents a particular local pattern or a patch shared by keypoints in the cluster. Consequently, a visual word vocabulary can be used to describe all local patterns or patches in the feature space. Based on the vocabulary, each keypoint is mapped into a visual word. Therefore, each image in the database is represented as a bag of visual words.

The BoW model was firstly explored for texture and material representations, based on the distribution of prototypical texture elements [32], [33], and then used for describing objects and scenes for the sake of image indexing or scene categorization [34]. Drawbacks of the BoW model mainly lie in its ignoring the spatial formation and lacking of semantic relationships (*e.g.*, polysemy and synonymy). To overcome the first drawback, some research efforts have been reported on either using the regular grid [29], spatial pyramid matching [35], or verifying geometric relationships after matching [36]; To deal with the second drawback, some research publications in recent years aim at designing higher-level visual representations to explore the semantic relationship between visual words by frequent adjacent patches pairs [6], frequently

co-occurring pattern [7], [8], minimal support region [9], or defined meaningful phrases [37]. However, relatively few research efforts are made to simultaneously and effectively address both of the above two drawbacks as well as the *background clutter* problem. In this paper, all of the above three problems can be well addressed via our proposed BoP model.

Notice that, different from the previously reported methods of mining association rules using visual bigrams [38], [39] which regards doublet visual words as those words with high probability in each topic after an initial run of pLSA or using frequent itemset mining algorithms to find out repeated groups of words from videos, in our work, we focus on calculating the frequent visual word pairs within a sliding sub-window which defines the spatial co-location patterns. Moreover, in the procedure of generating original visual words vocabulary, we adopt the agglomerative hierarchical clustering rather than the exclusive K-means clustering, which renders the visual vocabulary more semantic relationships to some extend.
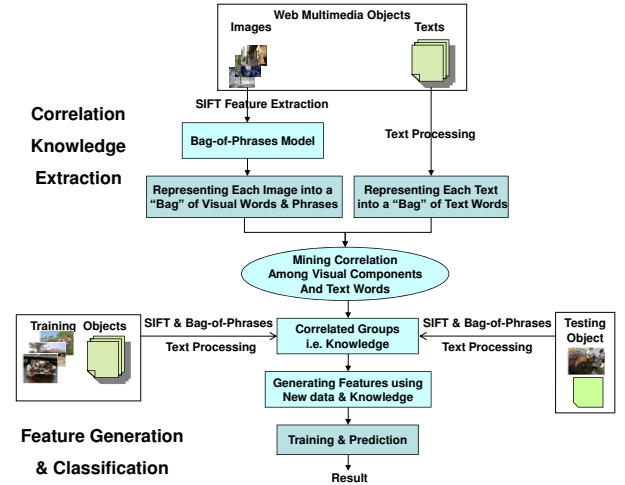
Fig. 3.   Framework of our proposed cross-domain transfer learning method.

III. Cross-Domain Transfer Learning

A. Framework Overview

Fig. 3 presents a brief framework of cross-domain transfer learning method, which is composed of a two-stage procedure: the knowledge extraction and utilization. Here, the first stage serves to extract correlation knowledge, and the second stage aims at applying such knowledge to generate new features which are finally applied to perform web multimedia object classification. The three major components in the framework, *i.e.*, BoP model, correlation knowledge extraction and its utilization in classification, are described in the following three subsections.

B. The Bag-of-Visual-Phrases (BoP) Model

1) Overview: The image representation is a prerequisite step due to its direct impact on the speed and accuracy of almost all kinds of image-based applications. In this section, we focus on building a novel model (*i.e.*, BoP) to represent

images at both the word-level and the phrase-level to improve the performance of the basic BoW model. Fig. 4 shows the flowchart of our proposed BoP model.
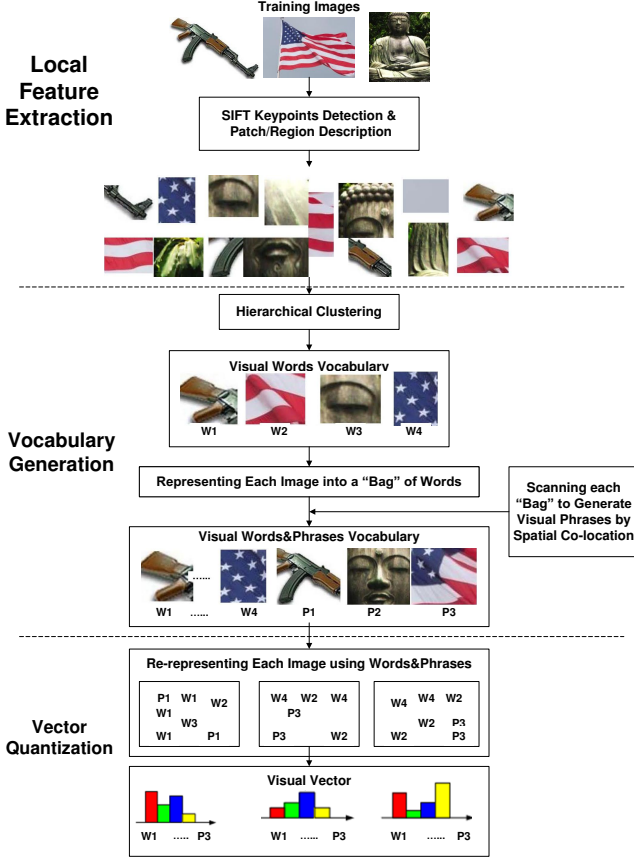
Fig. 4.   The flowchart of the proposed BoP model.

2) Hierarchical Clustering: In the basic BoW model, after local feature extraction and patch/region description, the K-means clustering is applied to generate the visual words vocabulary. Here, all the keypoints extracted from training images grouped in a mutually exclusive fashion so that if a certain keypoint belongs to a specific cluster then it could not be assigned to another cluster. However, in the original image, each keypoint represents a certain patch/region in the original image, and these patches/regions represented by each keypoint are often **overlapped** with each other or **being contained** in other patches/regions. Consequently, it is natural to use the hierarchical clustering to explore the semantic relationships among visual words. In this paper, we group local feature keypoints with the unsupervised agglomerative hierarchical clustering algorithm [40] to generate the vocabulary tree in which each visual word acts as a node.

For example, in Fig. 5, the intersection of visual words "eye", "nose" and "mouth" in K-means algorithm is zero; however, all three words together could represent a more general semantic concept ("face") in the hierarchical vocabulary tree.

Notice that, once no visual word in the testing image is classified as the visual word represented by a node at level i, there is no need to calculate the similarity between this testing visual word and all the other visual words in the same branch
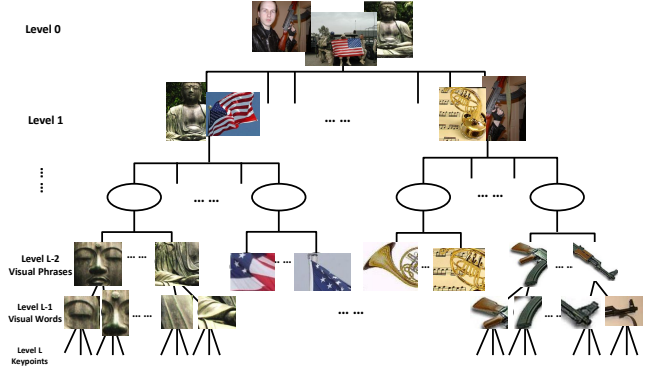
Fig. 5.   An example of generating the vocabulary tree using the agglomerative hierarchical clustering algorithm.

of the node. Therefore, representing images using this kind of hierarchical structure is more effective and efficient when performing the task of object categorization.

3) Spatial Co-location: The basic BoW model regards an image as a set of inter-independent visual words, *i.e.*, it assumes that the occurrences of visual words in the image are independent. However, the hypothesis is not reasonable because some visual words often appear together. Following this observation, we discover the co-location patterns to identify the subsets of features frequently located together.

Some works have been previously reported on using bi-grams for classification [38], and itemset mining to find out repeated groups of words [39]. However, the problem of discovering spatial co-locations is different from the above mining association rules because in spatial data sets there is no unified definition of transactions as well as items in traditional association rules [41]. In this paper, as shown in Fig. 6, we first scan all training images with a $M * M$ sub-window to obtain the frequent visual word pairs (*i.e.*, the frequency of their co-occurrences in a sub-window exceeds a certain threshold). Here, the sliding sub-window defines the spatial neighborhoods and each sub-window is equivalent to a transaction, while the visual words are equivalent to the items in association rules. If the window is too small, lots of co-location information might be lost; if the window is too large, too much irrelevant information (*e.g.*, background) would be involved and the accuracy of the co-location patterns cannot be guaranteed. In this paper, we choose the suitable size of the sliding sub-window M^2 according to the empirical experiments.

Notice that unlike words in text, there is no syntax information for these visual words. Therefore, given a window with n visual words, it can be transformed into $\binom{n}{2}$ word pairs. For example, if there are three visual words (a, b, c) in the window, they can be converted into three word pairs, (a, b), (a, c) and (b, c). Each visual word pair can be considered as a 2-itemset, but only the word pair whose frequency is higher than the given threshold can be defined as a 2-word visual phrase.

Finally, as shown in Fig. 4, the vocabulary is enriched by these generated visual phrases which are acting as the general visual words in the basic BoW model. In other words, these generated visual phrases are added into the vocabulary and
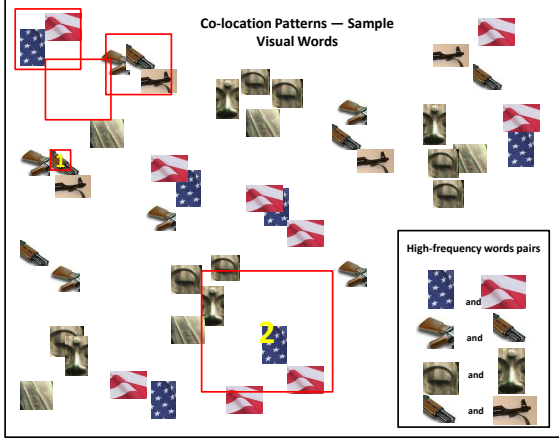
Fig. 6. An illustration of discovering spatial co-location patterns of visual words to generate visual phrases.

Fig. 7. An illustration of finding k-Cliques in the cross-domain correlated graph. The highlighted parts represent two k-Cliques.

regarded as its components as visual words in the basic BoW model. Therefore, the feature vector of an image is formed by first concatenating the histograms of visual words and phrases, and then performing *Principal Component Analysis(PCA)* on it.

Obviously, the spatial and semantic distinguishing power of image features can be enhanced via our proposed BoP model. In addition, parts of background information can be filtered. Generally speaking, objects in the same category are usually similar to each other while their backgrounds often differ from each other. Hence, the counts of the word pairs representing objects will be summed up and thus will be large, while those representing backgrounds will be much smaller. By specifying a proper sliding window and a suitable frequency threshold for the word pairs, the frequent object-related phrases will be maintained while parts of the infrequent background-related word pairs will be excluded successfully. Therefore, our method is capable of handling the *background clutter* problem effectively.

C. Correlation Knowledge Extraction

Given a collection of web multimedia objects, *i.e.*, images and their corresponding textual descriptions, in order to discover the correlation knowledge, both image and text are represented by corresponding data sources respectively. Specifically, for text, we extract features from textual descriptions assigned to the corresponding images via MAL-LET [42] and the *tf-idf* schema; For image, we adopt the classic SIFT(Scale Invariant Feature Transform) feature [28] and our BoP model together to represent each image as a bag of visual words/phrases (*i.e.*, visual components), each of which describes a patch/region of the original image. After that each multimedia object is represented by both visual components and textual words. Finally, we use correlation mining techniques to obtain the correlation knowledge, *i.e.*, a collection of strongly correlated groups of visual components and textual words.

A pair of visual component and textual word is said to be correlated if they frequently co-occur in the image and
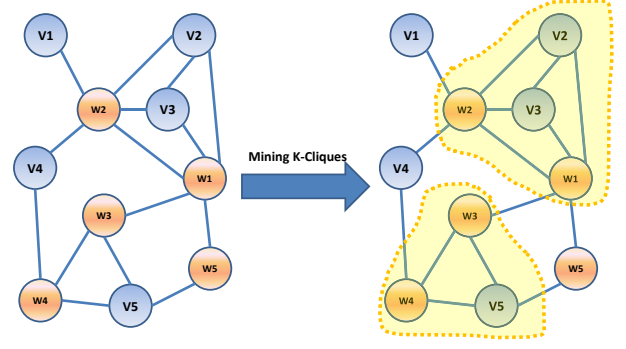
the corresponding text. The procedure of the "visual∼textual" correlation knowledge extraction is described in Algorithm 1.

Algorithm 1: Correlation Knowledge Extraction

Input: $O(v)$: the set of objects that contain the visual component v; $O(w)$: the set of objects that contain the textual word w; τ: a pre-defined threshold; $|\cdot|$: the size of the set

1   $I(v,w) = O(v) \cap O(w)$;
2   $U(v,w) = O(v) \cup O(w)$;

3   **if** $\frac{|I(v,w)|}{|U(v,w)|} > \tau$ **then**
4   the visual component v and the textual word w are correlated;
5   **return** correlation $v \sim w$;
6   **end**

Notice that the similar procedure could be applied to obtain the "visual∼visual" and "textual∼textual" correlations. Then a set of visual components and textual words

$$S = \{v_1, v_2, ..., v_m, w_1, w_2, ..., w_n\}$$

is said to be strongly correlated *iff* every pair (v_i, w_j), or (v_i, v_j), or (w_i, w_j) in S is correlated.

In our method, to generate correlated groups, we construct an undirected correlated graph where each node represents a visual component or a textual word after identifying all the correlated pairs, and each edge between any two nodes represents these two nodes are correlated according to Algorithm 1. Then we use the backtracking algorithm [10] to find k-Cliques in the graph (in this paper, k ≥ 3), as shown in Fig. 7. Here, a clique is a maximal complete subgraph of three or more nodes which are also adjacent to all the other members of the clique. Each correlated group of visual components and textual words, marked as C_k, is regarded as a concept with specific semantic meaning.

D. Feature Generation and Classification

The discovered correlation knowledge can be used to enrich the feature space for classification. With such enrichment,

the feature space for web multimedia objects includes both the actual components/words contained in objects and the collection of the correlated groups.

Given a new multimedia object for classification, it can be represented as an *(m+n)-dimension* vector:

$$\mathbf{x}^{(i)} = < v_1^{(i)}, ..., v_m^{(i)}, w_1^{(i)}, ..., w_n^{(i)} >,$$

where m and n represent the number of visual components and textual words respectively, $v_j^{(i)}$ is the number of occurrence of the *j*-th visual component in *i*-th object (O_i), and $w_j^{(i)}$ is the number of occurrence of the *j*-th textual word in O_i. For each correlated group C_k, we define a feature mapping value:

$$\Phi_{c_k} = \frac{|O_i \cap C_k|}{|C_k|},$$

i.e., the ratio of the number of components/words shared by C_k and O_i to the number of components/words in C_k. In other words, the value of Φ_{c_k} could reflect the proportion that the *i*-th object O_i contains how many (not necessarily all) words in the correlated group C_k. Then we investigate the following two strategies to generate features for unseen data.

(1)Enlarging strategy:

This strategy is to insert a *l-dimension* vector whose elements are the feature mapping value Φ_{c_k} of each C_k into the original feature vector [27], and then form an *(m+n+l)-dimension* vector:

$$\mathbf{x}^{(i)} = < v_1^{(i)}, ..., v_m^{(i)}, w_1^{(i)}, ..., w_n^{(i)}, \Phi_{c_k 1}^{(i)}, ..., \Phi_{c_k l}^{(i)} >,$$

where l is the total number of generated correlated groups.

(2)Enriching strategy:

This strategy is to enrich the original feature space via replenishing components/words that actually do not appear in the object into the original feature vector. Specifically, for a certain correlated group C_k, once the value of Φ_{c_k} is greater than a pre-defined threshold ϵ, we firstly calculate the mean value $\mu^{(i)}$ of the number of occurrence of the components/words shared by C_k and O_i(*i.e.*, components/words in $O_i \cap C_k$), and then assign it to the remaining components/words (*i.e.*, components/words in $C_k - (O_i \cap C_k)$) in the original feature vector. After processing all the correlated groups, the resulting vector is still of *(m+n)-dimension*:

$$\mathbf{x}^{(i)} = < v_1^{(i)}, ..., \mu^{(i)}, ..., v_m^{(i)}, w_1^{(i)}, ..., \mu^{(i)}, ..., w_n^{(i)} >,$$

where the original values of $\mu^{(i)}$ in the above feature vector are all zeros since they actually do not appear in the object.

In both of the strategies, an object may involve a concept with specific semantic meaning if it contains some (not necessary all) of the components in the corresponding correlated group. Therefore, both the strategies can handle the situation when one information source is missing. However, the first strategy enlarges the feature dimension by simply concatenating two vectors together, while the second one explores the semantic information hidden in the correlation knowledge and consequently renders the features more informative.

IV. Applications

Based on the aforementioned algorithms, the proposed cross-domain transfer learning method can be utilized for many real world applications, including some general-purpose applications, such as the social-media images organization, and some domain-specific applications, for example, the images classification in disaster emergency management.

In this section, we present two typical applications, both of which are difficult to be realized purely by traditional multimedia techniques, *i.e.*, without incorporating cross-domain correlation knowledge from unlabeled data through unsupervised learning.

A. Disaster Emergency Management

It is well known that disasters such as hurricanes and earthquakes cause immense damages in terms of physical destruction, loss of life and property around the world. For instance, in 2005, hurricane Katrina reported a total property damage of $81 billion. Thousands of people died in the actual hurricane and in subsequent floods. And in 2010, the Haiti earthquake affected billions of people, and an estimated 550,000 buildings collapsed or were severely damaged. In order to reduce such loss, emergency managers are required to not only be well prepared but also provide rapid response activities [43].

In recent years, with the proliferation of smart devices, disaster responders and community residents are capturing footage, pictures and video of the disaster area with mobile phones and wireless tablets. Therefore, a system that can effectively integrate multiple information sources such as textual reports and multimedia data would greatly assist emergency managers in making a better assessment of a disaster situation and performing efficient and timely responses correspondingly. Our proposed cross-domain transfer learning method can be potentially applied to disaster emergency management by effectively transferring the cross-domain correlation knowledge to disaster-related multimedia objects classification.

In this paper, for the application of disaster emergency management, in order to show the capability of our method in incorporating multiple information sources to assist emergency management, we crawled a disaster-related multimedia objects dataset from CNN News [1], including colored web images and their corresponding textual descriptions about "the aftermath of disasters". As is shown in Fig. 1, the dataset CNN News has several characteristics: 1) images in the same category may be differently visually but quite similar in terms of semantic concepts; 2) some images focus on the whole event whereas the others reflect only one single aspect of the whole event. For example, as shown in Fig. 1(a), the first image shows the macro overview of the whole event that Mississippi suffered extensive damage from hurricane Katrina, whereas the latter two images reflect two different aspects of the event respectively, one is an American flag flying in front of a damaged home, and the other is a line of rescue workers and working dogs; and 3) the corresponding textual descriptions

---

[1]http://www.cnn.com/

are usually composed of several paragraphs, instead of the textual tags, which is more challenging.

B. Social-Media Images Organization

Social media platforms, such as Facebook [2], Twitter [3], LinkedIn [4], Google+ [5], etc., play an increasingly important role nowadays because they provide users new ways to read the news, share the latest updates, interesting images, videos, and so on. As we all know, Flickr [6] is an social image hosting and video hosting website, and is home to over eight billion of the world's photos. In addition to being a popular website for users to share and embed photos, the service of Flickr is widely used in blogs and social media. Undoubtedly, users can benefit from those huge amounts of images and videos. However, as more data is coming, it has to be effectively retrieved by users. Thus, how to efficiently organize these images and videos to satisfy the needs of quick search of web users in such huge multimedia databases is becoming an urgent problem.

In this paper, we focus on the organization of social-media images and use a Flickr image dataset as an example. Since classification is a crucial step of image retrieval and indexing, meanwhile, social media users assign text descriptions to the images, a possible solution to the multimedia data organization is to take both textual information and image information into consideration to design an effective multimedia object classification method. However, sometimes users might be interested in images related to the weakly-defined textual topics [44]. As shown in Fig. 8, the textual descriptions (i.e., in the form of textual tags posted by individuals), which are provided to describe the general content of Flickr images, are sometimes ambiguous: the tag "pottedplant" might indicate that the image contains a flower of a pottedplant, a plant of a pottedplant, a pot of a pottedplant, a fertilizer for a pottedplant, a pot rack for displaying pottedplants, rather than a pottedplant itself. Considering the above requirements posted by social media users, and the capability of our proposed cross-domain transfer learning method in efficiently combining the textual and image information for classification, it can be potentially applied to social-media images organization.

V. EXPERIMENTAL EVALUATION

In this section, our experiments are divided into the following three parts. Firstly, we introduce two different datasets of web multimedia objects (consisting of images and their corresponding textual descriptions) used in our experiments. Secondly, we simply introduce the experimental tools and evaluation measures. Finally, in order to demonstrate the efficacy and effectiveness of our proposed image representation model and the cross-domain transfer learning method, we report a detailed experimental comparison among our proposed model/method and the different existing methods for web image categorization mentioned in Section II on both of the above different datasets.

Fig. 8. Example images and their partial corresponding textual descriptions of the Flickr.

A. Real World DataSets

In order to evaluate the efficacy and effectiveness of our proposed image representation model and the cross-domain transfer learning method, two different datasets of web multimedia objects are used in our experiments.

CNN News (355 *objects,* 4 *topics,* 76.9 *MB SIFT descriptors*): For this dataset, we manually collected 355 colored web images and their corresponding textual descriptions about "the aftermath of disasters" from CNN News, which include 4 different topics: "*building collapsing*", "*flood*", "*oil spilling*", and "*animal death*" , as shown in Fig. 1. Each topic includes 101, 101, 53 and 100 objects respectively, and each object either shows the macro overview of the whole event or just reflects one aspect of the event. The entire dataset is split into two parts: 176 objects (50, 50, 26, 50 from four topics respectively) are randomly selected for knowledge extraction and the rest 179 objects are taken as classification data, where about 50% objects (91 in total) are randomly selected for training and the rest 50% objects (88 in total) are taken as the test data.

Flickr (67,634 *objects,* 19 *categories,* 14.4 *GB SIFT descriptors*): This dataset consists of 67,634 web images along with their corresponding textual descriptions from a public image dataset obtained from Flickr [44] (see Fig. 8 for some examples). The entire dataset includes 19 different categories (see Fig. 13). We split objects in each category into two parts, thus 33,818 objects are randomly selected for the knowledge extraction and the rest 33,816 objects are taken as the classification data, in which about 50% objects (16,909 in total) are randomly selected for training and the rest 50% objects (16,907 in total) are taken as the test data.

B. Experimental Tools and Evaluation Measures

Vocabulary Building Tools:

As we mentioned before, we adopt the hierarchical clustering to build the vocabulary. For our first dataset – CNN News, many state of art toolkits can be used for vocabulary building purpose; while for our second dataset – Flickr, which contains 14.4 GB SIFT descriptors, none of the toolkits could handle the data with such scale. Therefore, we seek help from the

well-accepted open source Map-Reduce [45] framework by Apache – Hadoop [7], which is considered as the "expert" for processing the large scale data. We implement the hierarchical clustering over Hadoop, and then build the vocabulary through it.

Classification Tools:

In our experiment, LIBSVM [46] is utilized as the base classification tool, and the RBF kernel is adopted as the kernel function of svm classifier. The parameter tuning is done via *k*-fold cross validation. In the *k*-fold cross-validation, all the data in data set is divided into *k* subsets of equal size. The SVM classifier is trained for *k* times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the accuracy of the SVM classifier. This process is repeated until all the subsets have been used for both training and testing, and the computed average classification accuracy was used as the performance measure for the designed SVM classifier.

Evaluation Measures:

Due to the randomness of selecting objects for learning and classification, we run all the experiments 10 trials, and take the average classification accuracy as the performance measure in our experiments. Generally, the higher the accuracy value, the better the classification result.

C. Experimental Results

In this section, our experiments are divided into the following three parts. Firstly, we evaluate the validity of our proposed BoP model by comparing with the the other 5 models. Secondly, we compare our proposed two correlation knowledge based classification methods with the base line classification method without using such knowledge. Finally, in order to evaluate the efficacy of our proposed two cross-domain learning methods, we report an experimental comparison of the three different multi-view learning methods for classification mentioned in the first part of Section II, and then compare their performance with those of our proposed two cross-domain learning methods.

1) BoP Model: as we all know, one of the key parameters of the BoW model is the size of vocabulary, which directly impacts its performance. Therefore, we use 6 different vocabulary sizes and compare the classification accuracy of 6 different methods, some of which are mentioned in [30]. Specifically, these methods in our experiments include: (1) *K-means + BoW*, *i.e.*, the basic BoW model; (2) *K-means + Phrase Only*; (3) *K-means + BoP*, *i.e.*, representing images by both visual words and phrases; (4) *K-means + BoP + PCA*, *i.e.*, performing *PCA* to reduce the dimensions of the integrated features in method (3); (5) *Hierarchical + BoP*, *i.e.*, using the hierarchical clustering instead of K-means in method (3); (6) our proposed BoP model, *i.e.*, performing *PCA* on the features in method (5). Here, we cut the vocabulary tree generated from the hierarchical clustering at the layer which has the same vocabulary size with K-means for comparison.

As shown in Fig. 9 (a), the basic BoW model achieves the best performance when the vocabulary size is 300 (CNN
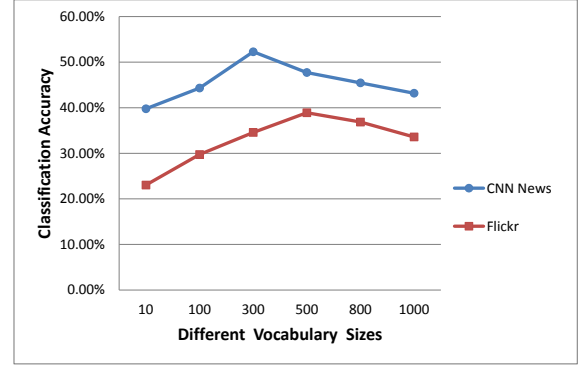
(a) Different Vocabulary Sizes
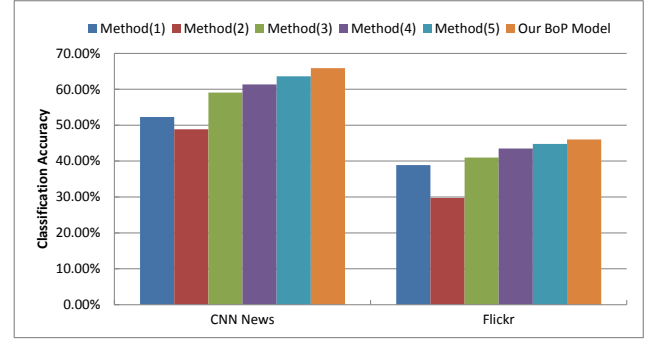
(b) Different Methods

Fig. 9. Comparisons on the categorization accuracy of different vocabulary sizes, and different methods for image representation.

News) and 500 (Flickr) respectively, which illustrates that sizes 300 and 500 are the optimal clustering size of our training datasets respectively. And as shown in Fig. 9(b), we evaluate the efficacy of using visual phrases, hierarchical clustering and *PCA* in our proposed image representation model. From the comparison results, we have the following observations:

1) The best results of object classification using single visual information (method (1) and (2)) are relatively low. However, once we represent images with both visual words and visual phrases, the accuracy can be improved. It shows that *by combining visual words with phrases, the distinguishing power of image features is enhanced.*

2) Method (5) outperforms method (3) and our proposed BoP model outperforms method (4). The BoP model achieves the best performance with the accuracy 65.9091% on CNN News and 46.0224% on Flickr. The reason behind the performance improvement is straightforward: *replacing K-means with the hierarchical clustering provides more semantic relationships* between visual words into method (5) and (6). In other words, it helps to generate the classifier from the semantic image patches/regions related to visual words.

3) Compared with method (5), the accuracy of BoP model increases about $2\% \sim 3\%$. Our proposed BoP model can outperform method (5) because *PCA could get rid of some noise information* which helps improve the performance. Similarly, method (4) outperforms method (3).
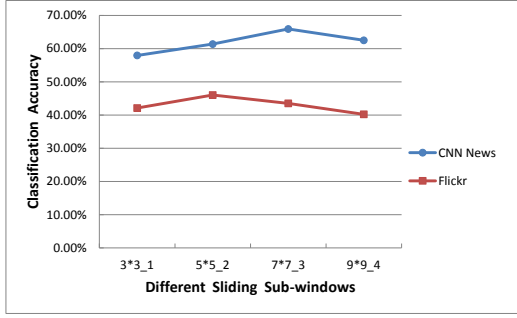
Fig. 10. Comparison on classification accuracy of 4 different sizes and moving steps of the sliding sub-window. Here, $M * M\_s$ represents the window size is set to M and the window moves s pixels each time.

Fig. 11. Examples of the correlated groups on CNN News.

Another two parameters in our proposed BoP model are the frequency threshold of words pairs and the size of the sliding sub-window. In this paper, we mainly focus on demonstrating the usefulness of phrases and the best way of identifying the first parameter will be our future work. Based on the experimental evaluation, the minimum support threshold of words pairs is 0.2 on CNN News and 0.1 on Flickr respectively. The second parameter defines the spatial neighborhoods to be involved while mining visual phrases. Taking the average size of images into account, we adopt the following 5 sizes and moving steps of the sliding sub-window. Here, in Fig. 10, using the above minimum support thresholds, we present the classification accuracy of different sliding sub-windows ($M * M\_s$ indicates that the window size is set to M^2 and the window moves s pixels each time). From the diagram, we observe that with the size $7 * 7\_3$ and $5 * 5\_2$ of sliding sub-window respectively, we obtain the best classification accuracy on the two datasets. The reason for the overall trend of each diagram rising first and then declining is straightforward: *using too small window size will omit some spatial co-location information, while using too large window size might involve inaccurate co-location information* which would lead to relatively low accuracy.

2) Correlation Knowledge based Classification: In our proposed transfer learning method, the results can be divided into the following two parts:

Correlation Knowledge Extraction:

There is a pre-defined threshold τ when discovering the correlation knowledge. According to the observation in our experiments, the higher the value τ is, the fewer the correlated pairs we can obtain. For the dataset CNN News, based on experimental evaluation, the correlated groups are generated with a threshold $\tau = 0.25$. Fig. 11 shows some example results of the correlated groups obtained in our mining process [8], where each row represents a correlated group. From the results, we observe that *each correlated group of visual components and textual words can be regarded as a concept with specific semantic meaning.*

Feature Generation and Classification:

After obtaining the above correlation knowledge, we gen-

[8]Due to the space limit, we only present part of the experimental results obtained on CNN News here. Similar results can be obtained on Flickr.
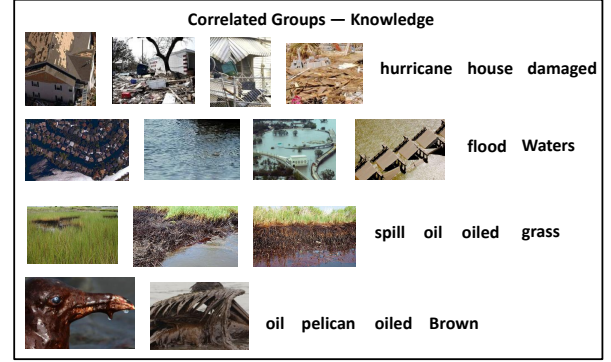
erate new features to perform classification via investigating two different strategies to utilize the correlation knowledge to enrich the feature space. An important parameter of the *Enriching Strategy* is the pre-defined threshold ϵ. If ϵ is always the same for all object categories, it may not be robust enough for classification due to the various of object categories. Therefore, We further compare the classification performance on each category under different thresholds ϵ. The confusion tables of accuracy under different ϵ for datasets CNN News and Flickr are shown in Fig. 12 and Fig. 13. Here, each black part represents the best classification performance. From the results, we observe that: (1) Our proposed correlation knowledge based classification method achieves its best performance on two datasets when ϵ is 0.8 and 0.9 respectively; (2) In Fig. 12 (b), the topic 1 and 3 achieve best results when ϵ is 0.8, whereas in Fig. 12 (c) and Fig. 12 (a), topic 2 and 4 obtains its best performance when ϵ is 0.9 and 0.7 respectively. And the similar observations could be obtained in Fig. 13.

(a) $\epsilon = 0.7$ (b) $\epsilon = 0.8$ (c) $\epsilon = 0.9$
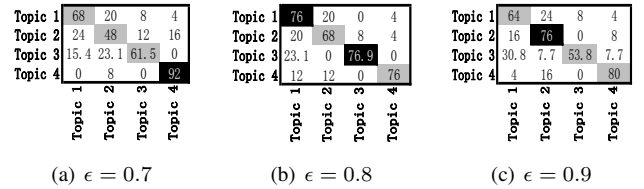
Fig. 12. The confusion tables of accuracy of 4 topics under different ϵ for CNN News.

Finally, we compare our proposed two correlation knowledge based classification methods with the base line classification methods without using such knowledge. The experimental results using different information sources are shown in Table I. From the results, we have the following observations: (1) the classification results are significantly improved using the correlation knowledge; and (2) the *Enriching Strategy* beats the *Enlarging Strategy* for the reason that the *Enriching Strategy could benefit from the semantic information hidden in the knowledge and consequently renders the features more informative whereas the first strategy only makes use of raw knowledge information.*

3) Cross-Domain Transfer Learning: For further comparison, we firstly implement three existing multi-view learning methods for classification mentioned in the first part of Sec-

(a) $\epsilon = 0.7$

	aero	bic	bird	boat	bot	bus	car	cat	chair	cow	dog	horse	per	sheep	sofa	train	moto	pot	tv
aeroplane	62.8	0	0	3.6	0	0	3.8	0	14.7	0	0	14	0	0	0.2	0.9	0	0	0
bicycle	0	72.1	0	0	0	0	0	0	0	0	0	27.9	0	0	0	0	0	0	0
bird	0	0	57.7	0.2	0	0	0.5	0	0	0	0	41.4	0	0	0	0.2	0	0	0
boat	0	0	0.1	74.5	0	0	0.4	0	0.8	0	0.1	23.3	0.1	0	0.1	0.6	0	0	0
bottle	0.1	0.1	0	0.5	55.5	0.3	1.5	0.1	1.3	0	0	39.8	0.1	0	0	0.7	0	0	0
bus	0	0	0	0	0	85.1	0.1	0	0.1	0	0	14.7	0	0	0	0	0	0	0
car	0	0.1	0.7	0.1	0	0.1	78.5	0	0	0	0.7	19.5	0	0	0.3	0	0	0	0
cat	0	0	0	0.3	0	0	0.8	57.5	1.2	0	0.6	39.6	0	0	0	0	0	0	0
chair	0.5	0.2	0	0.3	0	2	1.8	1.3	60.6	0	1.4	29.7	0	0	0.7	1.5	0	0	0
cow	0	0	0	0	0	0	0	0	0	54.4	0.1	44.8	0.7	0	0	0	0	0	0
dog	0	0	0	0	0	0	0.2	0	0.1	0	56.4	43.2	0	0	0.1	0	0	0	0
horse	0	0	0	0	0	0	0	0	0	0	0.1	99.8	0	0	0.1	0	0	0	0
person	0.2	0.2	0	1.7	0	0	4.3	0.1	5.4	0.1	0.4	35.4	51.4	0	0.2	0.6	0	0	0
sheep	0	0	0	0	0	0	0	0	0	0	0.3	31.1	0	68.6	0	0	0	0	0
sofa	0	0	0	1	0	0	3.9	2.5	7.9	0	2.4	31.9	0	0	49.5	0.9	0	0	0
train	0	0	0	0	0	0	0.1	0	0	0	0	7.3	0	0	0	92.6	0	0	0
motorbike	0	0.1	0	0.7	0	0.1	1.5	0	4.6	0	0	58.5	0.4	0	0	0.3	33.8	0	0
pottedplant	1.6	0	0	2.7	0	0	5.4	0	10.8	0	0.5	14.5	1.1	0	2.2	2.2	0	59.1	0
tvmonitor	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	50

(b) $\epsilon = 0.8$

	aero	bic	bird	boat	bot	bus	car	cat	chair	cow	dog	horse	per	sheep	sofa	train	moto	pot	tv
aeroplane	96.2	0	3.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bicycle	0	87.4	12.5	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0
bird	0	0	99.6	0.3	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0
boat	0	0	20.3	79.6	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
bottle	0	0.2	4.3	0	94.8	0.1	0.1	0.3	0.1	0	0	0	0	0	0	0	0	0	0.1
bus	0	0	1.6	0	0	97.7	0.5	0	0.1	0	0	0	0	0	0	0.1	0	0	0
car	0	0.1	1.8	0.2	0	0.1	97.2	0	0	0.5	0	0	0	0	0.1	0	0	0	0
cat	0	0	5.2	0	0	0	0	93.7	0	0	1	0.1	0	0	0	0	0	0	0
chair	0	0	3.2	0	1.3	0.1	0.1	0	94.4	0	0	0	0	0	0.9	0	0	0	0
cow	0	0	11.1	0	0	0	0	0	0	88.6	0	0.3	0	0	0	0	0	0	0
dog	0	0	2.9	0	0	0.2	0	0.1	0	0	96.4	0	0.2	0.1	0.1	0	0	0	0
horse	0	0	2.5	0	0	0	0	0	0	0	7.9	89.6	0	0	0	0	0	0	0
person	0	0.1	4.5	0	0	0	0.1	0	0	0	0	0	95.3	0	0	0	0	0	0
sheep	0	0	4.5	0	0	0.1	0.4	0	1.5	0.1	0	0	0	93.4	0	0	0	0	0
sofa	0	0	31.9	0.1	0	0	0.1	0.7	0	0.2	0	0	0	0	67	0	0	0	0
train	0.1	0	14.5	0.1	0	0	0	0	0	0	0	0	0	0	0	85.2	0	0	0
motorbike	0	0	6.3	0	0	0.1	0	0	0	0.1	0	0.1	0	0	0	0	93.3	0	0
pottedplant	0	0	15.1	0	0	0	1.6	0	0	0	1.1	0	0	0	0	0	0	82.3	0
tvmonitor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

(c) $\epsilon = 0.9$

|	aero	bic	bird	boat	bot	bus	car	cat	chair	cow	dog	horse	per	sheep	sofa	train	moto	pot	tv
aeroplane	99.9	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0
bicycle	0	73.1	0	8	0	0.5	10.9	0	0	0	0.1	4.3	0	0	0	3.1	0	0	0
bird	0	0	99.8	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
boat	0	1.1	0.1	46.7	0	24.8	1.8	0	0	0	0.3	2.5	18.4	0	0	4.3	0	0	0
bottle	0	0	0	0	99.5	0.1	0	0.1	0.2	0	0	0	0.1	0	0	0	0	0	0
bus	0	0.6	0	4.6	0	81.4	4.4	0	0.2	0	0.5	0.1	0	0	0	8.2	0	0	0
car	0	1.5	1	6.8	0	7.9	76.5	0.2	0	0.7	1.3	0.3	0	0	3.8	0	0	0	0
cat	0	0	0	0	0	0	0	97.8	0	0	1.1	0.1	1	0	0	0	0	0	0
chair	0	0	0	0	0	0	0	0.2	98.1	0	0	0	0	0	1.6	0	0	0	0
cow	0	0	0	0	0	0	0	0	0	89.3	0	10.7	0	0	0	0	0	0	0
dog	0	1.8	0.1	0	0	0.1	0	0.1	0	0	97.3	0	0.2	0.3	0.1	0	0	0	0
horse	0	0.3	0	1.8	0	0.3	5.3	0.1	0	0	11.4	74.4	0	0	0	6.4	0	0	0
person	0	0	0	0	0	0	0.1	0	0	0	0	0	99.9	0	0	0	0	0	0
sheep	0	0	0	0	0	0	0	0	0	1.8	0	0	0	97.7	0	0	0	0	0
sofa	0	0	0	0	0.1	0	0	0	0.1	0	0	29.3	0	0	70.5	0	0	0	0
train	0	5.4	0	13.5	0	29	6.1	0	0	0.1	3.2	10.9	0	0	0	31.8	0	0	0
motorbike	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0	99.6	0	0
pottedplant	0	0	0	0.5	0	0.5	0	0	0	0	0.5	0.5	0	0	0	0	0	97.9	0
tvmonitor	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	75

Fig. 13. The confusion tables of accuracy of 19 categories under different ϵ for Flickr.

Dataset	Information Source	Without Knowledge	With Knowledge	
			Enlarging Strategy	Enriching Strategy
CNN News	Image Only	65.9091%	67.0455%	68.1818%
	Text Only	62.5%	67.0455%	70.4545%
	Our Method	63.6364%	**69.3182%***	**73.8636%***
Flickr	Image Only	46.0224%	51.4402%	56.4914%
	Text Only	66.7534%	67.546%	80.4282%
	Our Method	69.2731%	**78.1925%***	**84.2018%***

TABLE I

COMPARISON OF CLASSIFICATION ACCURACY. THE RESULTS WITH $*$ ARE SIGNIFICANTLY BETTER THAN THE OTHERS (UNDER $p < 0.005$), AND THE RESULTS OF *Enriching Strategy* SHOWN HERE ARE THE BEST ACCURACY ON TWO DATASETS WHEN ϵ IS 0.8 AND 0.9 RESPECTIVELY.

tion II, and then compare their classification performance with those of our proposed two methods using *Enlarging Strategy* and *Enriching Strategy*. The reason for we choosing these three existing methods is that each one of them is well-accepted and typical representative of that type of methods. The concrete methods we adopted in our experiments will be described as follows:
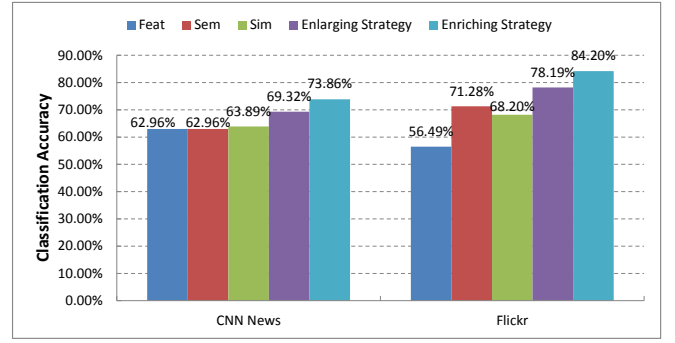
Fig. 14. Comparison among different multi-view methods. The results of *Enriching Strategy* shown here are the best accuracy on two datasets when ϵ is 0.8 and 0.9 respectively.

- Feature Integration (*Feat* for short): In the experiments, we adopted one popular method proposed in [2]. Here, we treat unique terms as text features and extract image features using BoP model, and combine the features of text and image together by simply concatenating these two types of features to form a larger vector as the input of SVM classifier;
- Semantic Integration (*Sem* for short): We train two classifiers based on textual features and image features (BoP) respectively, and then ensemble these two classifiers to be an integrated version via voting schema, similar to the method proposed in [22];
- Kernel Integration (*Sim* for short): We compute the pairwise similarities using textual features and image features (BoP) respectively, and then use the weighted summation [25] of these two types of similarities as the similarity measurement between objects. Note that different weights can be assigned to the features of different data sources. We tune the weight factor to find the optimal one via the empirical comparison;
- Our proposed cross-domain transfer learning method using *Enlarging Strategy*: utilize the extracted correlation knowledge to enrich feature space for classification by inserting a new vector into the original feature vector;
- Our proposed cross-domain transfer learning method using *Enriching Strategy*: utilize the extracted correlation knowledge to enrich feature space for classification via replenishing visual components and textual words that actually do not appear in the object into the original feature vector.

As shown in Fig. 14, compared with the classification results provided by three existing methods, our proposed two cross-domain transfer learning methods outperform the others. The reason behind the performance improvement is straightforward: *our proposed two methods utilize the extracted correlation knowledge to enrich feature space for classification*. In other words, the discovered cross-domain correlation knowledge can provide "guidance" to the new learning tasks, (*i.e.*, classification) such that the classification procedure can be facilitated. In addition, the improvement of overall performance on dataset Flickr is larger than that on

dataset CNN News because the textual descriptions assigned to the images in Flickr are textual tags which can be seen as more accurate textual information after preprocessing.

VI. CONCLUSION

In this paper, we empirically studied the problem of transferring and applying the cross-domain correlation knowledge among visual components from images and textual words from corresponding texts. In order to discover more meaningful correlation knowledge, we propose a novel image representation model (*i.e.*, bag-of-visual-phrases (BoP)) which incorporates the spatial and semantic information to represent images at both the word-level and the phrase-level. While transferring and applying such correlation knowledge, we investigate two different strategies to utilize such knowledge to enrich the feature space where the classification is performed. The proposed method can not only handle large scale web multimedia objects, but also address the problem that the textual descriptions of a small portion of web images are missing. Moreover, by representing images using the BoP model, it also can successfully classify the images in which the objects and the background information are mixed together. In addition, our proposed method can apply in both specific domain (*e.g.*, disaster emergency management) and general domain (*e.g.*, social-media images organization). Empirical experiments on two different web multimedia objects datasets demonstrate the efficacy and effectiveness of our proposed image representation model and the cross-domain transfer learning method.

For future work, we are interested in exploring an adaptive selection mechanism to automatically choose the optimal sliding sub-window and the pre-defined threshold for each object. Moreover, the process of knowledge exaction is off-line and the running time is slightly more, which is still a problem to be solved. In this paper, we mainly focus on demonstrating the validity and feasibility of using cross-domain correlation knowledge to help multimedia object classification, and solving this problem will be our future work. Therefore, we are studying the online method for real-time applications right now. Also, to fulfill the requirements of web users, it is better to explore the feasibility of combining three or more data sources(text, image, video, audio, etc.) for web multimedia object classification and retrieval.

REFERENCES

[1] Z. Yin, R. Li, Q. Mei, and J. Han, "Exploring social tagging graph for web object classification," in *Proceedings of ACM SIGKDD*, 2009, pp. 957–966. 1

[2] L. Wu, S. Oviatt, and P. Cohen, "Multimodal integration-a statistical view," *IEEE Transactions on Multimedia*. 2, 3, 11

[3] Y. Wu, E. Chang, K. Chang, and J. Smith, "Optimal multimodal fusion for multimedia data analysis," in *ACM Multimedia*, 2004, pp. 572–579. 2, 3

[4] T. Li and M. Ogihara, "Semisupervised learning from different information sources," *Knowledge and Information Systems*. 2, 3

[5] Y. Wang and S. Gong, "Refining image annotation using contextual relations between words," in *Image and video retrieval*, 2007, pp. 425–432. 2, 3

[6] Q. Zheng, W. Wang, and W. Gao, "Effective and efficient object-based image retrieval using visual phrases," in *ACM Multimedia*, 2006, pp. 77–80. 2, 4

[7] J. Yuan, Y. Wu, and M. Yang, "Discovery of collocation patterns: from visual words to visual phrases," in *CVPR*. IEEE, 2007, pp. 1–8. 2, 4

[8] S. Zhang, Q. Tian, G. Hua, Q. Huang, and S. Li, "Descriptive visual words and visual phrases for image applications," in *ACM Multimedia*. ACM, 2009, pp. 75–84. 2, 4

[9] Y. Zheng, M. Zhao, S. Neo, T. Chua, and Q. Tian, "Visual synset: towards a higher-level visual representation," in *CVPR*. IEEE, 2008, pp. 1–8. 2, 4

[10] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *ACM*. 2, 6

[11] R. Raina, A. Ng, and D. Koller, "Constructing informative priors using transfer learning," in *Proceedings of ICML*, 2006, pp. 713–720. 2, 3

[12] R. Raina, A. Battle, H. Lee, B. Packer, and A. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proceedings of ICML*. 2, 3, 4

[13] Y. Zhu, Y. Chen, Z. Lu, S. Pan, G. Xue, Y. Yu, and Q. Yang, "Heterogeneous transfer learning for image classification," in *AAAI*, 2011. 2, 3, 4

[14] W. Dai, Y. Chen, G. Xue, Q. Yang, and Y. Yu, "Translated learning: Transfer learning across different feature spaces," *NIPS 2008*, pp. 353–360, 2008. 2, 3, 4

[15] S. Roy, T. Mei, W. Zeng, and S. Li, "SocialTransfer: cross-domain transfer learning from social streams for media applications," in *ACM Multimedia*, 2012, pp. 649–658. 2, 3, 4

[16] T. Berg and D. Forsyth, "Animals on the web," in *CVPR*, vol. 2. IEEE, 2006, pp. 1463–1470. 3

[17] F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting image databases from the web," *PAMI, IEEE Transactions on*, no. 99, pp. 1–1, 2011. 3

[18] H. Li, J. Tang, G. Li, and T. Chua, "Word2image: towards visual interpreting of words," in *ACM Multimedia*. ACM, 2008, pp. 813–816. 3

[19] X. Liu, B. Cheng, S. Yan, J. Tang, T. Chua, and H. Jin, "Label to region by bi-layer sparsity priors," in *ACM Multimedia*. ACM, 2009, pp. 115–124. 3

[20] J. Hare and P. Lewis, "Automatically annotating the mir flickr dataset," in *Multimedia information Retrieval*, 2010. 3

[21] ——, "Semantic retrieval and automatic annotation: Linear transformations, correlation and semantic spaces." SPIE, 2010. 3

[22] R. Carter, I. Dubchak, and S. Holbrook, "A computational approach to identify genes for functional RNAs in genomic sequences," *Nucleic Acids Research*. 3, 11

[23] C. Bishop, *Pattern recognition and machine learning*, 2006. 3

[24] M. Jordan and R. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*. 3

[25] B. Schölkopf and A. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, 2002. 3, 11

[26] W. Lee, S. Verzakov, and R. Duin, "Kernel combination versus classifier combination," *Multiple Classifier Systems*, pp. 22–31, 2007. 3

[27] J. Zhang and S. Shakya, "Knowledge Transfer for Feature Generation in Document Classification," in *ICMLA*, 2009, pp. 255–260. 3, 4, 7

[28] D. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*. 4, 6

[29] L. Fei-Fei, "Visual Recognition: Computational Models and Human Psychophysics." 4

[30] T. Tuytelaars, C. Lampert, M. Blaschko, and W. Buntine, "Unsupervised object discovery: A comparison," *IJCV*, vol. 88, no. 2, pp. 284–302, 2010. 4, 9

[31] G. Salton and M. McGill, "Introduction to modern information retrieval," 1986. 4

[32] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *IJCV*. 4

[33] O. Cula and K. Dana, "3D texture recognition using bidirectional feature histograms," *IJCV*. 4

[34] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *CVPR*. 4

[35] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*. 4

[36] K. Palander and S. Brandt, "Epipolar geometry and log-polar transform in wide baseline stereo matching," in *ICPR*. IEEE, 2009, pp. 1–4. 4

[37] M. Sadeghi and A. Farhadi, "Recognition using visual phrases," in *CVPR*. IEEE, 2011, pp. 1745–1752. 4

[38] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman, "Discovering objects and their location in images," in *ICCV*, vol. 1. IEEE, 2005, pp. 370–377. 4, 5

[39] T. Quack, V. Ferrari, and L. Van Gool, "Video mining with frequent itemset configurations," *Image and Video Retrieval*, pp. 360–369, 2006. 4, 5

[40] J. Han and M. Kamber, *Data mining: concepts and techniques*. Morgan Kaufmann, 2006. 5

[41] Y. Huang, S. Shekhar, and H. Xiong, "Discovering colocation patterns from spatial data sets: a general approach," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1472–1485, 2004. 5

[42] A. McCallum, "MALLET: A Machine Learning for Language Toolkit," 2002, http://mallet.cs.umass.edu. 6

[43] L. Zheng, C. Shen, L. Tang, T. Li, S. Luis, S.-C. Chen, and V. Hristidis, "Using data mining techniques to address critical information exchange needs in disaster affected public-private networks," in *ACM SIGKDD*, 2010, pp. 125–134. 7

[44] M. Allan and J. Verbeek, "Ranking user-annotated images for multiple query terms," in *British Machine Vision Conference*. 8

[45] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008. 9

[46] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," 2001. 9

Georgia

Institute

of Technology

Characterizing the Intelligence Analysis Process:
Informing Visual Analytics Design through a Longitudinal Field Study

Youn-ah Kang

Georgia Institute of Technology

John Stasko

Georgia Institute of Technology

ABSTRACT

While intelligence analysis has been a primary target domain for visual analytics system development, relatively little user and task analysis has been conducted within this area. Our research community's understanding of the work processes and practices of intelligence analysts is not deep enough to adequately address their needs. Without a better understanding of the analysts and their problems, we cannot build visual analytics systems that integrate well with their work processes and truly provide benefit to them. In order to close this knowledge gap, we conducted a longitudinal, observational field study of intelligence analysts in training within the intelligence program at Mercyhurst College. We observed three teams of analysts, each working on an intelligence problem for a ten-week period. Based upon study findings, we describe and characterize processes and methods of intelligence analysis that we observed, make clarifications regarding the processes and practices, and suggest design implications for visual analytics systems for intelligence analysis.

KEYWORDS: Intelligence analysis, qualitatvie user study.

1 INTRODUCTION

Visual analytics applies to many domains and problem areas, but one area of particular study since the beginnings of the field has been intelligence analysis. Intelligence analysis is a cognitively demanding process, one that seems ideal for the application of visual analytics tools. Accordingly, a growing number of systems have been built for it [2, 11, 25, 30].

Research in human-computer interaction also teaches us to deeply analyze and understand end-users and their problems in order to design appropriate computational solutions. We question whether visual analytics systems, including some of our own, have been based upon a deep enough understanding of the discipline. Relatively few studies of intelligence analysts, their tasks, and their work processes exist. Notable exceptions [4, 12, 19, 23] provide initial insights into the field, but we have frequently interacted with analysts who feel that their practices are misunderstood and that visual analytic systems often fail to address their most important problems.

To address these concerns and to learn more about the analysis process, we conducted a longitudinal, observational field study of intelligence analysis on real world problems. Unfortunately, getting access to working, professional analysts is challenging. Even if they are available, it is difficult or impossible to study them for an extended period of time while they work on real tasks without having some type of special access that simply was not available to us. As an alternative, we studied analysts-in-training who are soon to become working professionals. More specifically,

ykang3@gatech.edu
stasko@cc.gatech.edu

we studied groups of students from the Department of Intelligence Studies at Mercyhurst College as they conducted a term-long intelligence project.

We were given deep access to the students, the materials they examined, the tools they used, and their final intelligence products. We interviewed the teams multiple times and observed their group meetings. Additionally, we interviewed their instructor to learn his impressions of the process. Our goal was simply to better understand what these young analysts do, the challenges they face, and how we might be able to help them. Thus, the contributions of our research include a characterization of the processes and methods of intelligence analysis that we observed, clarification and reflection of several beliefs about intelligence analysis processes and practices, and resultant design implications for visual analytics systems for intelligence analysis.

Munzner has argued of the importance and the need for more domain characterization research like this [17]. She notes that such research is both difficult and time consuming to do properly, but the visualization community could benefit greatly from it.

2 BACKGROUND

One of the most widely used models in the visual analytics community is Pirolli and Card's sensemaking model [19] for intelligence analysis. While the model broadly characterizes processes used in analysis activities and has guided the design of visual analytics tools, the model does not provide rich details of how intelligence analysts work in the real world. More empirical, descriptive explanations of the intelligence analysis process are required to provide appropriate visual analytics system solutions.

Several studies have captured and characterized the work practices and analytical processes of individual or collaborative analysis through a qualitative approach. Chin et al. [4] conducted an observational case study with professional intelligence analysts in which participants worked on real-world scenarios. The researchers revealed various characteristics of the analytical processes of intelligence analysts. Gotz et al. [7] also recognized the lack of studies examining analyst behavior and conducted a user study to explore the ways in which analysts gather and process information. Another study by Robinson examined how analysts synthesize visual analytic results by studying domain experts conducting a simulated synthesis task using analytical artifacts [20]. Based on analysis of video coding results, he identified several characteristics in the process of collaborative synthesis. While these studies did not evaluate specific visual analytic tools or features per se, they provide valuable implications to inform design directions for future support tools.

Johnston [12], an anthropologist, conducted an ethnographic study of the CIA for a year and identified variables that affect intelligence analysis and requirements for techniques and procedures to reduce analytic error. While he made useful recommendations to improve analytic performance, his approach was primarily intended to understand organizational culture and describe current community practices, rather than identifying leverage points for designing support systems.

In our study, we aim to deeply understand the analysis process with an eye toward designers. We seek to provide meaningful implications for developing technological support for analysts.

3 METHODS

In order to investigate the intelligence analysis process in-depth, we conducted an observational study of teams of analysts conducting an in-class intelligence project. In the term-long (ten-week) project, each team addressed a real intelligence problem proposed by a client. We observed three teams, monitoring their status and process throughout the project. At the end of the project, each team had to produce final deliverables and present their findings and analysis to decision makers.

3.1 Participants

We recruited three groups of students, one team of four undergraduate students (Team A) and two teams of five graduate students (Teams B and C), from the Department of Intelligence Studies at Mercyhurst College [5]. Mercyhurst's Intelligence Program, started in 1992, provides education for students who want to pursue a career as an intelligence analyst. It is recognized as one of the top programs for intelligence studies in the United States, offering a broad range of classes and degrees for students seeking a career as an analyst in national security, law enforcement, or the private sector.

We recruited students who were taking the courses named "Strategic Intelligence" (undergraduate) and "Managing Strategic Intelligence" (graduate), in which teams are required to conduct an analysis project over a ten week term. The two courses are very similar with respect to the projects. The students all were close to graduation, with past internship experience, and most of whom had already received job offers.

While these student teams clearly are not practicing professional analysts, there was not a significant difference between the way the students worked and the way real analysts work, according to the instructor. The analysis process used in the class was modeled directly after the process employed by the US National Intelligence Council to produce its strategic reports, the National Intelligence Estimates [18]. The instructor also intentionally stayed relatively detached from the students, acting as a mentor and limiting his supervision so that the teams could autonomously work on the project. The teams were diverse in expertise on the subject matter, which is common for teams in the intelligence community. One key difference from real world practice was the relative absence of administrative and bureaucratic overhead affecting the student teams, as well as issues relating to security clearances. They operated in a much more "sanitary" environment than the real world.

3.2 Task

Different types of intelligence questions exist - we focused on one of the most common types, strategic intelligence. Strategic intelligence is "intelligence that is required for the formulation of strategy, policy, and military plans and operations at national and theater levels [8]." Strategic intelligence is exploratory and long-term in nature. The requirement for tasks within the class was that "the questions should be relevant and relatively important to the client's success or failure but outside their control." We served as a client/decisionmaker for team A in order to observe the process even closer, whereas Teams B and C worked with external organizations. The specific issues each team addressed were:

Team A
The strategic assessment of potentially influential factors to the evolution of computer-mediated undergraduate and graduate distance education: What aspects of computer-mediated distance education will likely influence R1 institutions during the next 5, 10, 20 years with specific, but not exclusive, emphasis on undergraduate education and computer science?

Team B
Who are the key people, technologies and organizations that likely currently have or will develop the potential to disrupt or replace traditional US national security Intelligence Community (IC) analytic work flows and products with commercially available products available over the next 24 months?: Criteria that will be used to identify these key players are:
- *Those that are not beholden to the IC or US Government as primary sources of funding.*
- *Those that are looking at future based events or actions that are outside the control of the forecaster/predictor.*

Team C
What are the most consistent and identifiable characteristics displayed by potential insider threats to (a defense department)?
- *An insider threat will be defined as an individual or collection of individuals employed directly or indirectly by the department who violate security or access control policies with the intent of causing significant damage to the department's personnel, operations, or information.*
- *Within the broad range of insider threats, special priority will be given to violent threats and improper diversion of information or physical assets.*

The teams updated the status and the process of the project on a wiki site. At the end of the semester, they needed to produce a final report that synthesizes analytical results, and strategies of the entire analysis process.

3.3 Study Protocol and Procedures

The analyst teams conducted the project for ten weeks - from the week of September 1 through the week of November 10, 2010 for Team A, and from the week of December 1 to the week of February 14, 2011 for teams B and C. Normally, strategic intelligence projects range from a couple months to years; ten weeks is short but within normal limits for strategic intelligence.

Before the project began, the external clients formulated a draft of their initial intelligence problem. In the first week of the project, the clients conducted a conference call with the analyst team to discuss the scope and requirements of the problem. During the next two weeks, the analysts refined the problem and wrote a formal statement of the intelligence question, which they call "Terms of Reference (TOR)". Upon approval from the decisionmakers, the teams began working on the problem, which took another seven weeks.

The wiki platform was used as a workspace for analysts to document their process and findings, and we were able to monitor the wiki's status throughout the project period. The final report of the projects also was placed on the wikis.

During the project period, we conducted two face-to-face meetings with each team – one in week 7 and the other in week 10. In the meetings, we interviewed each team as a group and the class instructor in order to learn more details about the project's status, process, difficulties, and future steps. Each interview took approximately an hour. While the interview was semi-structured, we followed an interview guide containing several key topics [3, 15], including:

- How do the analysts perceive their analysis process?
- What barriers and difficulties do they encounter?
- Tools and aids being used - where and why?

- Collaborative aspects in the analysis process
- Where in the process can technology help?

We also observed two team meetings firsthand, which took about 3 hours in total.

3.4   Data Collection and Analysis

Most of the process descriptions and produced artifacts were stored digitally. The teams reported methodologies, tools used, sources, as well as the findings on their own website (wiki). To further understand the process, we analyzed interview notes and audio recordings from the focus group interviews. We used the artifacts produced by the analysts, such as drawings, wiki pages, tables, and slides as further data. Additionally, we had access to history logs of wiki page changes.

We transcribed each interview's audio recording and then coded the transcripts based on the grounded theory approach [26]. We began by identifying major themes and categories from the text. One emergent theme focused on the analysis process, including methodology and challenges encountered.   Another theme was collaboration, focusing on how and to what degree the analysts collaborated and what types of collaboration existed. Throughout the coding process we iteratively refined the categories. We then elaborated on supporting evidence from the data for each category through a deductive approach.

4   RESULTS & FINDINGS

4.1   Overall Analysis Process

Through the project, we found that four component processes were essential to the overall analysis: constructing a conceptual model, collection, analysis, and production. While the four processes did not occur linearly, this section describes the importance of each and how the analyst teams worked on each.

Phase 1: Constructing a Conceptual Model
Once the teams and clients/decisionmakers finalized the requirements of the intelligence question, the teams started to build a conceptual model, which is a map of issues and concepts that the team will be investigating to address the problem. The conceptual model illustrates the areas the analysts need to research by helping them to visualize the question at hand. The question is placed in the center, and then several high-level components of the question surround the question (Figure 1). Each component branches out and creates a bigger map, from which the team gains an idea of the areas with less/more information that they need to research. This allows the team to focus on collecting a set of data with an appropriate scope.

While the significance of the conceptual model differs depending on the question and the team, it plays a key role for the team to understand the domain area and determine the direction of research. We were told that analysts often construct this conceptual model implicitly, rather than externalizing it, which we found quite interesting. The instructor commented:

> *In most cases, it's implicit. People don't write it down. But it's the way they are actually doing it. There's a model in people's heads, and that's far more important than the data. There's research that says analysts' judgments are far more driven by the way they think about the problem than the data itself. So making the way you think about the problem explicit would allow analysts to identify whether they disagree about how to think about this model, and to merge their best thinking about this model. So the process happens, it's just the degree of to which it's made explicit, that is unusual.*
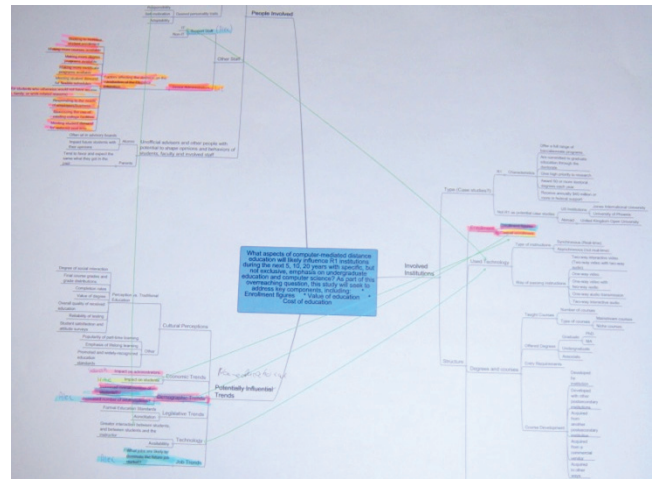

Figure 1. Conceptual Model. Printed from Mindmeister

Phase 2: Collection
While working on the conceptual model, the teams also assigned areas/concepts to each member. Next, they collected information from various sources including online and offline sources (e.g., interviews with experts), which they call "all-source intelligence". While each analyst was responsible for collecting data about their assigned topic(s), the team shared their sources using Zotero, a web browser plug-in for gathering and organizing source material. This allowed teammates to view the data like a common library – other team members might already have found information that they need. More specifically, the data collection process typically involved these steps:
- Once an analyst identifies needed information, they search the Internet using search engines and various keywords.
- The analyst also establishes RSS news feeds on websites of interest using Google Reader.
- Whenever they find useful sources, whether for their own topic or someone else's, the analysts place the link into a Zotero group library.

Phase 3: Analysis
The analysis phase exhibited various characteristics depending on the requirements and analytical methods used. In this phase, analysts processed data that they collected from many different sources in order to convert "information to knowledge." While team A directly began writing short format analytical reports on each topic, team B and C used a more structured format (e.g., spreadsheets) to quantify information and rank the significance of each topic or entity. No matter which method they used, the initial analysis of each topic/entity was undertaken and written by one person in accordance with the assigned topic. However, everyone on the team could review and comment on the others' work via the wiki pages. In all cases, the analysis phase was incorporated with the collection and the production phase.

Phase 4: Production
Once individual collection and analysis was almost finished, the teams met and tried to synthesize findings from each part, which led to the "key findings" – the major product of the analysis. Production was an intensive reading/writing process in which the team collaborated tightly with each other. This stage was more to prepare a presentation for the decisionmakers. Team members repeatedly checked their sources and findings to make sure that they were consistent and logical.

Reiterating, while we separated these four components for the sake of clarification, the process was not simple and it was not clear which phase the team was in throughout the project. Instead,

the characteristics of the question and the analytic method chosen most influenced the process. In our study, we observed two different styles of intelligence analysis process. The difference in approaches resulted from the type of the intelligence question.

4.1.1 Intuitive Analysis – Team A

Team A addressed potentially influential factors to online distance education in the near future. Because the requirements were rather broad and intuitive, the team decided to take a top-down approach, investigating meta-information sources such as research that forecasts future education trends.

Instead of using a specific analytic method, this team depended considerably on the conceptual model and used it as a guide throughout the entire project. They put significant time and effort into constructing it, revising the conceptual model until the seventh week of the project. Because of time constraints, they were not able to cover all the topics in the model. Through discussions, they chose a number of concepts they felt most worth exploring and divided up the concepts for each member.

After collecting and reading information for their designated topic, each analyst wrote a short format analytical report that synthesized the information. Most of the analysis simply involved reading. For a few topics that required careful weighing of alternative explanations, the team employed analysis of competing hypothesis (ACH) [9]. While documenting results, everyone was able to review and edit the others' drafts on the wiki page, and team members frequently discussed others' analysis (short write-ups) both online and face-to-face. Therefore, everyone was responsible for the reporting of each topic.

After working on the individual topics, the team met to write key findings together. This team invested considerable efforts in synthesizing their findings because their narrative was extremely important for their intuitive type of analysis.

4.1.2 Structured Analysis – Teams B and C

Teams B and C used structured analysis with quantified information because their research questions tended to be more specific and required rank-ordering of entities (e.g., top x indicators, key people/companies). Both teams built their conceptual model in the beginning as a base model. For these teams, however, the model was more of a collection plan rather than an actual conceptual model. Although they used the model to collect information and divide up the work, they did not refer to it for the remainder of the project. Instead, they started building a matrix in a spreadsheet to collect and analyze data from diverse sources. The matrix was rather a re-interpretation of the conceptual model, and each cell in the matrix indicated a collection requirement.

The purpose of the matrix was to evaluate each entity based on criteria chosen and identify the most influential ones, those of most interest to the decisionmaker. Team B, that was asked to identify key people, technologies, and companies that might affect intelligence community products, created a matrix and chose criteria while collecting information. They identified 180 entities and graded each based on the criteria, noting the ones with highest scores. Team C, that was asked to identify indicators displayed by potential insider threats to a defense department, analyzed data from the 117 case studies about crimes using a matrix (Figure 2). They used it to compare the relationship between crimes and motivations, as well as crimes and indicators.

In both teams, the matrix captured the conceptual model and how each team was thinking about the question. Filling in the cells was a time-consuming part as analysts needed to read and analyze each case/source to fill in one cell, addressing "the devil in the details."

However, this type of analysis required additional efforts in the production phase. Initially, the teams converted qualitative information from sources into quantitative information for rank-ordering. Once they had completed the matrix, the teams needed to transform its data into a story so that it could be made useful to decisionmakers.

Upon the completion of the projects, the instructor evaluated the teams' performances as being in the "top 10% of the projects over 8 years." He commented that all three teams performed the analysis well and in one case, the decisionmaker briefed the head of his organization with the team's results.

Figure 2. Case Study Matrix of Crimes

4.2 Tools and Methods Used

The teams used various software tools and analytical methods to develop hypotheses, arrive at analytic estimates, and create written reports and multimedia products.

Wikispaces/Google Sites: The teams used a wiki platform (Team A&B – Wikispaces, Team C- Google Sites) to exchange gathered information, aid administration, and share organizational details. The wiki sites became part of the final product, displaying the key findings, terms of reference, and all analytic reports.

Mindmeister (conceptual model): Mindmeister is an online mindmapping tool the teams used to build a conceptual model [16]. A conceptual model provides a revisable platform to view the requirements and their components. As research and facts begin to support or refute initial ideas, main ideas become more solidified and focused.

Zotero: The teams used Zotero as a source collection database [31]. Downloaded as an Add-on to Mozilla Firefox, Zotero allows the analyst to search websites and save the sites in a database that is accessible through the Zotero website. The teams used the Group Library feature to place their sources in a single database.

Website Evaluation Worksheet: To evaluate the credibility of the online sources, all the teams used the Dax Norman Trust Scale [19]. This matrix allows scores to be applied based upon criteria such as clear bias, corroboration of information, and the analyst's overall perception of the source. Based on the sum of scores, the source can score a High, Moderate, Low, or Not Credible rating.

Analytic Confidence: Each report includes an analytic confidence section that conveys to the decisionmaker the overall doubt connected with the estimative statement(s). While assessing the level of analytic confidence, the teams used Peterson's method [21]. Peterson identified seven factors that influence analytic confidence: the use of structured analytic methods, overall source reliability, source corroboration, level of expertise on subject, amount of collaboration, task complexity, and time pressure. In the analytic confidence section, the teams addressed these six factors as applicable to the particular estimate.

Social Network Analysis: Team C employed social network analysis using i2's Analyst's Notebook [11] to see relationships within industry. The team analyzed the social network analysis based on betweenness and eigenvector scores.

5   UNDERSTANDING THE INTELLIGENCE ANALYSIS PROCESS

Observing analyst teams helped us to better understand their goals and processes. In particular, the study highlighted a number of misconceptions we harbored about the intelligence process. Other visual analytics researchers may or may not share these preconceived beliefs, but we think that they have the potential for misunderstanding and are thus worth exploring.

Intelligence analysis is about finding an answer to a problem via a sequential process.
Some existing models of the intelligence analysis view it as an answer-finding process with a sequential flow, as noted in several models of the intelligence analysis process [12, 14, 27]. This perception presumes that the process is linear, sequential, and discrete by step. Pirolli and Card's sensemaking model [22] includes the notion of iterations and revisions between steps, but the fundamental assumption is that separate stages exist throughout the process and that analysts transition between stages.

However, this model was not the intelligence process we observed. Instead, the process appeared to be more parallel and organic, as one analyst described:

> *Intelligence analysis is not about getting from point A to point B along the route, but it is better associated with basic research where you don't necessarily know where you are going to go. You're cutting a path through the jungle that's never been explored. That's what you're doing in most intelligence analysis projects. It's not a mechanical process in a sense that an assembly line is. It's a very exploratory activity by nature.*

The key part of the intelligence process is the analysis of a specific set of data.
Visual analytics systems often manipulate pre-processed data for analysis. A primary misconception about intelligence analysis is that the data analysis process, in which investigators analyze a set of collected data, is the most difficult part and takes the most time. This belief assumes that analysis occurs after investigators collect all data required for the analysis.

This view, however, needs to be changed. Although analysis is important, we observed that the process of "constructing a frame," as described in the Data-Frame theory [13] is more important. In other words, intelligence is about determining how to answer a question, what to research, what to collect, and what criteria to use. This process becomes part of the analysis - analysis implicitly occurs during the process of the construction. Analysts also explore different sets of analytic techniques to address a problem. Deciding which method to use is important, but it often changes during analysis as the way that analysts think about the problem evolves, depending on information that constantly flows in.

Understanding that collection and analysis are integrated together in the process of building a frame is extremely important. Systems are not likely to be successful in supporting intelligence without acknowledging that fact. One analyst commented:

> *Intell analysis is not like that you have a set of data in hand and run a program. It's like a conundrum from the very beginning. You have to learn how to learn, how to frame the question, and how to answer it through collecting and evaluating sources.*

Analysts do not often collaborate.
One common perception of intelligence views analysts as isolated individuals who prefer to work alone, struggling with pieces of information, rather than as collaborative teams [4]. However, a faculty member at Mercyhurst countered this perception:

> *Collaboration is almost all intelligence analysts have done in the context of the team. In the CIA or DIA, working as a team is pretty normal. While working on a particular topic within an agency is typical, also typical is working on an interagency team that consists of analysts from different agencies such as state department teams, DIA teams, and NSA teams.*

> *Analysts are normally organized by function or geographical region. These typically operate as loose teams. Strategic projects almost always involve a team as do crisis projects (for example I am sure there are multiple Libya teams that did not exist a month ago). In short, teamwork is the norm although the teams differ in the degree of formality and to the degree that there is a designated leader.*

During our study, we also observed many collaborative elements of intelligence analysis. Collaboration is commonplace in intelligence analysis, and understanding how that occurs is important because it influences one's whole notion of the process. The intelligence community itself has recognized the importance of improved collaboration since 9-11 [6]. Although collaborative tools have been built and they are pushing users into tighter collaboration, it is still important to understand where tighter collaboration will be beneficial and where it may not help much.

We found that multiple layers of collaboration exist in intelligence analysis and that the degree of collaboration differs depending on the type of task and the group dynamics. We observed that analysts usually do not collaborate tightly on data and content – the actual collection and analysis. Although the teams had meetings frequently – twice or three times per week – the main purpose was to discuss their status, issues, and the next steps, as two analysts said:

> *We come up with an agenda before meeting, a list of what we're supposed to talk about - what we did, what we want to do, what the questions we need to solve as a group. We didn't really plan that way, but it just happened. It's the way it is.*

> *There was no detailed, content-based work during the meeting although sometimes we had discussions on controversial issues. We basically do work in our own time, get preliminary ideas of it, and get together and discuss what issues we had.*

The teams often worked together in the same lab, but it was rare that they worked on the same document or content. They worked on their own part, but they often talked to each other about issues and questions. Essentially, they took advantage of being in the same place at the same time. To better support analysts' work, we need to understand the collaborative aspects of the intelligence process and where technology can leverage collaboration. Some tasks are inherently done better by an individual.

We can help intelligence analysts by developing sophisticated analytic tools that assist their thinking process.
Visual analytics researchers often seek to help intelligence analysts by developing technologically advanced analytical tools, thereby assisting their cognitive processes. The tools support specific types of analysis, specific analytical methods, and specific stages of the process. Such tools certainly can be helpful, especially to assist analysts to handle a flood of information.

However, our study revealed that analysts want something more than that. Currently, more than 50 analytic methods exist in the intelligence community [10], and analysts try many different kinds of techniques depending on the problem. Consequently, their dependency on a specific analytical technique is relatively low. Instead, the ability to manage the intelligence process

effectively and employ various analytical methods and tools quickly is more important, as the instructor and an analyst said:

Everything is fragmented. I've got Mindmeister here, Mindmeister doesn't interface with my search technology and Google reader. I've got to manually go out and figure out what all those bullets are. No help from a computer…But there isn't any set that ties all these, the pieces are there, Mindmeister, Zotero, RSS Google reader, MS Word, the wiki, are here, but nothing links all that in one seamless thing so I can go from the requirement to a product in a single package, in a single way.

6 REFINING OUR THEORIES

6.1 Rethinking the Intelligence Analysis Process

6.1.1 Linear vs. Parallel

One might believe that the way intelligence analysts work is quite simple and straightforward. First they specify requirements, build a conceptual model of what to research, then collect information, analyze data using various techniques, and finally write a report. This belief is a common misconception about intelligence as mentioned in the previous section. The reality is quite different. Rather than working linearly, analysts work on *everything* during almost the entire project. That is, analysts do not hold writing until enough information is collected; they keep revising analysis and writing as new information flows in. Analysts do not decide what to research and move on to collecting information; they start searching for information even when they are not sure what to research. Analysts do not produce final products after they are done with analysis; they already have an idea or a structure of final products in the very beginning, although it may be rough.

This "parallelism" is portrayed well in Wheaton's model of the intelligence process (Figure 3). In each phase, one of the core processes is emphasized most but all other functions operate in parallel. Wheaton argues that "All four functions begin almost immediately, but through the course of the project, the amount of time spent focused on each function will change, with each function dominating the overall process at some point [29]."

Although several distinct elements exist in the analysis process, all are very closely coupled and the connection is very organic. One can easily observe an analyst working on collecting new information while analyzing and checking the credibility of previously collected sources at the same time. In our study, we observed that a team's conceptual model changed drastically in the middle of the process, that a new information source was added ten days before the deadline, and that a previous analysis report was discarded and new analysis began in a late stage. The matrices also kept changing as new information arrived. While the teams were working on the matrix, they were collecting information at the same time to make sure that they were familiar with the area. Several quotes better explain this:

But it isn't as rigidly isolated as it's on that (traditional) cycle because you can't build a good conceptual model without knowing what's out there. So there's little bit of collection as you're building the model and we refined it.

Our conceptual model is changing. It doesn't get set in phase 1 and we drive it, that's the difference between this process and an outline. An outline drives your production. But we are using it differently. As it changes, we're changing our analytic focus, we're making decisions about production, who's going to write something, who's going to do the analysis, based on how it's changing and that's being informed by new information that comes in.
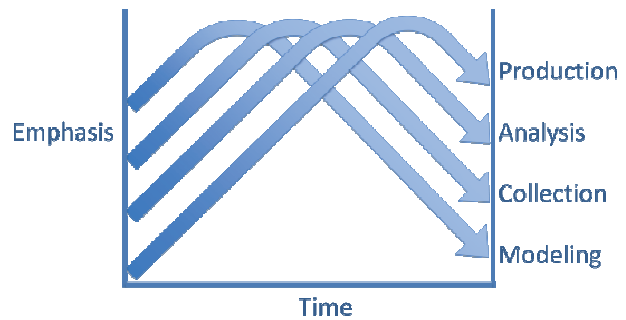

Figure 3. Wheaton's Multi-phasic Model of the Intelligence Process [29]

6.1.2 Pirolli and Card's sensemaking model

How does this new way of thinking about the intelligence process relate to Pirolli and Card's sensemaking model [19]? Because it is the most widely used model in the Visual Analytics domain, we were curious how well their model explains real-world intelligence analysis processes.

Pirolli and Card's model provides new insights about the intelligence process, suggests leverage points for analysis tools, and has guided the design of many visual analytics systems. However, we argue that the model still implies sequential, discrete stages of the intelligence process although it acknowledges that analysts can move either top-down or bottom-up or jump to different stages. For example, the model does not explain why analysts so frequently jump from one state to another state that is not adjacent. Many visual analytics tools thus support specific states only (e.g., shoebox and evidence file, evidence marshalling, foraging), and often they do not blend into the entire process of intelligence analysis.

More importantly, the model describes how information transforms and how data flows, rather than how analysts work and how they transition. It gives a very insightful illustration of how the form of information evolves from raw data to reportable results. However, it does not quite fit analysts' mental model of their work process because they do not work as information is transformed. Rather, information is transformed by how analysts proceed. Similarly, all different states of the model can exist at any point during the process. Analysts may have polished reports on certain sub-topics, drafts of analysis, structured matrices, and a collection of documents at a time.

The Pirolli-Card model identified various leverage points for visual analytics tools, but the linearity of the model could give researchers an inaccurate impression of the process. While models are inherently abstract and stage-based, it is important to understand the context and the purpose of the model. We would characterize their model as more of an information-processing process rather than intelligence analysis process. Pirolli-Card explicitly state that he model was suggested as a starting point to investigate the domain. While it has contributed to visual analytics researchers understanding of the domain, now we need to change our assumptions to build systems that better help intelligence analysts with their work.

6.2 Where and How Collaboration Occurred

6.2.1 Collaboration throughout the process

Throughout the project, the teams worked tightly together although the degree to which they collaborated differed depending on the phase of analysis. Once the project started, the team set up weekly meetings. The first thing they had to decide was to specify requirements of the problem, and then they collaboratively worked on building the conceptual model. Whether the team kept using this model or changed to a matrix, it played a role as a representation of their "group thinking," as an analyst described:

6

You want to say that this is the way I'm thinking about this problem. These are some of things I need to think about. And what we've done by building the conceptual model is to have that sort of group interaction, which is not necessarily harmonious action. There can be disagreements about how we should be thinking about this. And if there's shifting, moving it around, that represents an evolution of the way of our thinking.

Once the team had an idea of the areas to explore, they divided up the work and assigned concepts to each analyst. While each one worked on different concepts, they collaborated in collecting information by using a group library. Although this seems to be loose collaboration, the benefit the team gained was invaluable because it could significantly save time and effort in collection. An analyst explained how they worked in collection using Zotero:

Zotero is a good example of one way we collaborate. Each person creates a group library on the Zotero server. If I find a website that I think is useful, whether for my topic or someone else's topic, I add it to our group collection, and then other members can see it before they go searching the Internet for something. And if she doesn't find that in Zotero, then she might go out Google. So..try Zotero first, you might already have it.

While working on and analyzing their own topics, team members often met with each other to check status and discuss issues. When not all team members were available, they used typewith.me [28], a web-based collaboration tool for writing. When most of the areas they had planned to explore were covered and analyzed, they collectively wrote the key findings – the crux of the analysis project. Very tight collaboration occurred in this work. They met together and spent significant time to synthesize findings from all the topics and write the key findings.

6.2.2 Sharing vs. Content vs. Function

We found that three different types of collaboration exist when analysts discuss the topic: sharing, content, and function.

Sharing is a way to collaborate by sharing information. In our study, analysts shared sources to better assist their search process and understanding of the topics. At a higher level, however, this can be the sharing of analytical products as well as information sources. When people mention the importance of collaboration in the intelligence community, they primarily refer to the sharing of information sources and analytical results [1]. This type of collaboration can be significantly supported by technology.

Collaboration also occurs at the **content** level. This type of collaboration, in which analysts work together to create analytic products, can be seen more often in a small-size team. Examples in this project include constructing a conceptual model together, dividing concepts and assigning to each analyst, commenting on each other's analysis, working on ACH together, and writing the key findings together. However, in our study, once work was divided, then each part was done individually. The degree of tightness in this type of collaboration may directly affect the quality of analysis. The more closely the team works together, the more that output is coherent and logical. However, in reality, it is difficult to collaborate on content because of efficiency. This type of collaboration is also difficult to facilitate via technology because so many subtle issues – such as social dynamics, politics, teamwork, and motivations – are involved.

Functional collaboration is needed to execute practical tasks for completing the project, such as editing, creating a matrix structure, specialized analysis on a specific topic, and polishing deliverables. Whereas analysts work on the same thing and divide up the analytic product in the content level, functional collaboration naturally emerges at the later stage of the process as the team begins to think about allocating multiple functions. In this type of collaboration, analysts reinforce their strength. For example, if one is a good editor and has a detailed eye, then that person would do the editing, as one analyst explained:

There was a lot of collaboration. A spent a lot of time working on wiki stuff. B spent time doing the ACH stuff. C did a lot of technical stuff. So each of us spent extra amount of time doing something specific. Whichever parts of this you want to take on, those are the parts that get divided up. As different as I do this (analysis), I do this concept, I do this concept.

Olson et al. [20] similarly characterized collaborative activities of groups by analyzing design mettings from four software teams. Focusing on the time spent in meeting activities, they found similar patterns across design teams. In meetings, teams spent 40% of the time in direct discussions of design. 30% of meeting time was spent taking stock of their progress, and coordination activities consumed approximately 20% of the time. Clarification of ideas across these activities took one-third of the time, indicating that participants spent a large amount of time sharing and explaining expertise.

7   HOW VA CAN HELP: DESIGN IMPLICATIONS

How can visual analytics help intelligence analysis? Based on our study findings and reflections, we suggest several design implications for systems supporting intelligence analysis.

Externalize the thinking process - Help analysts continuously build a conceptual model
"Good analytic practices encourage continuous improvement upon the conceptual model throughout research, which continues through the end of the project."

The analysts in our study told us that the process of making sense of a problem and building a conceptual structure is one of the most important parts of intelligence analysis as it decides the direction of analysis. In most cases, analysts encounter a situation in which they need to learn about new subject matter, but it takes time and effort until they become familiar with the domain. Because they cannot build a good mental model of the problem without knowing what information is available, they struggle to know more about the domain until the later stage of analysis.

Using the power of representation, visual analytics systems can help analysts build a conceptual model or a structure of the problem and domain. For example, the system can take the main question the analyst has and suggest a number of possibly related concepts and keywords based on online encyclopedias, table of contents of books, tagging services, etc. The system should allow the analyst to refine the concepts so that it can repeat the search and suggest other relevant concepts. By connecting, grouping, and organizing concepts, analysts can continuously build up their conceptual model or structure of the area throughout the process. One analyst cited experience:

Ok, I got to model something, I've got to do a report on Ghana, I don't know anything all about Ghana, where's the tool that if I hit the button, it gives me a picture of what the relationship is, the model how to think about Ghana? It gives me 60-70% of the solution. But it gives me the ability to input and tweak and change those. Because I want to have a role in that, I can't allow the computers to do all my thinking, you know.

Support for this externalization should occur throughout the analysis process because as analysts learn more about the domain, they alter their way of thinking and refine their visual model.

Externalizing the thinking process also can assist analysts when they review their analysis after the project terminates. Supporting this activity would be especially useful because it will inform how the analysts could have done better and the areas that need to be examined if they did a similar project, as the instructor said:

The other thing this model helps you do is at the end of the project you can look back and go, "What did we not have time to do? And how does that impact our company, our estimates?" Because whatever reason we didn't get to it, this was important, we thought this define the space…We can sit back and go, ok, how confident are we on our estimates, knowing that our analysis is always at some level incomplete? And it's always incomplete, but how does it impact our confidence in our product? That's another way to use this representation.

Support source management - enable managing both pushed and pulled information and organizing sources meaningfully
One prominent characteristic of how analysts think about sources is that they have to be always vigilant of new sources. They often search for the same keywords again to see if any new materials have been added regarding the topic (pulled sources). They also receive news articles through RSS feeds everyday and check if they have received interesting information (pushed sources).

This process of searching sources takes more time than one may think, and systems should allow analysts to manage both pushed and pulled information associated with concepts they have identified. For example, a system could populate several concepts chosen by the analyst and store all the pulled sources in a database such as Zotero. Based on sources already found, the system also could recommend push resources such as blogs and news articles. For each source collected, the analyst could express if it is a useful source or not. Analysts commented on this functionality:

Sources are what we have to get, but where is the tool where I can integrate them? My RSS feeds dump into me every morning. But then I do searches as well. Where's the tool that allows me to integrate all data, the information that is useful for me?

If that kind of system exists, I have the ability to go back and find all my sources. Automatically, this (keywords, phrases) gets populated. And every point, I have the ability to say no or yes, no or yes to a source. But the actual extraction or the pulling, and the organization of that is automatic from that.

Then the list of sources can be organized in a meaningful way – for example, by keyword queries, by tags the analyst annotated, or by date the source was added. The system also could provide several ways of representing source results such as summary and tag clouds. Further support for analysis or visualization of collected sources as a group would be extremely beneficial.

All these technical capabilities currently exist in visual analytics systems. Now it is important that they be integrated together appropriately.

Support analysis with constantly changing information - integrate collection and analysis in a single system and help analysts use structured methods during collection
As described in the previous section, collection and analysis are not separate, but highly integrated processes. Analysts do not wait until all the data are gathered; rather, they start analysis even when they have only a few pieces of information. Through the repeated process of collection and analysis, they revise a frame and use the collected data as supporting evidence for the frame.

Currently, many systems provide analytical support assuming that processed data is available. If a system does not support a

seamless transition between collection and analysis, it is likely to be less successful in assisting the analysis. Analysts collect during analysis and they analyze during collection. This differs from statistical analysis, in which a structure or a frame about how to analyze the data is clearly defined and analysis is done with clean dataset. An analyst mentioned:

If they had more reliable, structured data, I'd use statistical analysis. But intelligence data is unstructured and dirty. You don't know what the best way to analyze it is until the middle of the process, or even the end of the process.

Multiple visual analytics systems provide analytical capabilities. By supporting more flexible data manipulation so that analysts can easily import and remove data from the analysis pool, these systems will be more usable, with better integration into the analysis process.

If the processes of collection and analysis are integrated in a single system, this helps analysts apply structured analytic methods such as ACH, social network analysis, geospatial mapping, and decision matrix. In our interviews, two teams mentioned that if they had more time, they would have tried other analytic techniques. Analysts always want to push their findings and triage, aggressively reshuffling their analysis. One of the most effective ways to do this is to employ multiple analytic methods and compare and contrast findings from each. The ability to try various techniques with the data can help analysts find effective ways for addressing questions and strengthening their analysis.

We had this time crunch. We pretty much got rid of the process of re-evaluating our hypothesis, finding what's the most important to make it perfect, and hitting on that, and going back to the stuff that we didn't deem as important. If we had time, we would fill that in.

Help analysts create convincing production – support insight provenance and sanity checks of analytical products
Production is what differentiates intelligence analysis from general sensemaking which does not necessarily entail external representation. Even when analysts finish their analysis, they need to convert the results into a concise format so that decisionmakers can understand their findings. This can be tedious and time-consuming part of the intelligence process.

When asked about the most difficult part of their project, two teams mentioned production. Interestingly, this difficulty comes from sanity checking and insight provenance, not simply from formatting and writing issues. The sanity check, or qualitative double-check, takes time because data and findings are derived from many sources and analysts have meshed them through the process of collection and analysis. Analysts need to return to original sources and provide a rationale by which their statements are made. They also have to add references to their statements, for which they have to revisit original sources. The following quote from an analyst illustrates those difficulties:

Most difficult part…basically going back through all the sources we used to grade these technologies, people, and companies, then taking basic pieces from those and making a narrative out of it. So explaining why we thought they are the keys and then relating it to the rest of the other findings.

A system that promotes simple insight provenance during analysis could help analysts save their time in production.

Support asynchronous collaboration rather than synchronous collaboration for exploratory analysis

We discussed three different layers of collaboration in the intelligence process and that the degree to which technology can contribute varies. In particular, visual analytics systems seem to have the potential to help collaboration in "sharing" and "content."

From our study, we found that these types of collaboration tend to occur asynchronously, rather than synchronously. When meeting face-to-face, analysts did not work on actual tasks but spent time checking their status, coordinating next steps, and discussing issues. Even when they worked in the same lab for several hours, team members took their own computer and worked individually. Although they often talked to each other, it was for simple coordination issues or specific questions about the content. One analyst stated about his perception on collaboration:

We discussed how each of us interprets the data. We're very group-oriented when it comes to discussing to a consensus. Other than that, we prefer to work individually especially for the actual analysis. Of course we collaborate even when we work on our own parts, but there's no one who really knows about those concepts or entities like you do.

In a nutshell, analysts collaborate cognitively. Rather than trying to build a system that allows analysts to work at the same time in the same workspace, providing a system that promotes individual workspaces but also provides asynchronous collaborative features - such as the ability to share sources and data, view and comment on others' work, and merge individual work together - would appear to be more beneficial.

Note that our findings are based on strategic intelligence. In other types of intelligence such as tactical and operational intelligence, which form the basis for immediate action, real-time collaboration is also important because such intelligence must be shared and used quickly.

Unifying the pieces

Because their typical processes of requirements gathering, collection, analysis, and production are so intertwined, and it takes considerable time to coordinate between different software systems, it appeared to us that analysts want an all-in-one system that can streamline the analysis process and save their time. When asked about their 'dream' system, a few analysts answered:

If I had to go back to the beginning and start all the way over, I should be able to jump back and forth seamlessly between all of these processes. We need a tool that compensates for that.

It should be one program. We spend more time to make it work together. Nothing's compatible with others. We want a program that syncs all the documents. Help us do our visualization with the documents. A program that is compatible with Excel spreadsheet. Don't want to open 20 different programs.

Thus, a hypothetical tool that simplifies the intelligence analysis process would function as follows:
- The analyst enters requirements into the system.
- The system suggests various concepts associated with key terms, phrases, and ideas in the requirements.
- The system automatically draws connections between concepts, but it also allows the analyst to draw connections, group, and organize them.
- The system takes the concepts and starts populating them, collecting information sources using the concepts as keywords (pull sources).
- The system uses sources the analyst identified and suggests new articles relevant to the sources (push sources).

- All of these pulled and pushed sources are integrated into a source repository.
- For documents in the database, the analyst can highlight important facts and annotate his/her thoughts. On demand, the system extracts entities requested by the analyst.
- For intuitive analysis, the analyst can write reports in a preferred format, walking through each document.
- For structured analysis, the system helps the analyst try a variety of structured methods. It takes all the information identified by the analyst and integrates it directly into the methods.
- At the end of the process, when the analyst produces final output, the system automatically links each statement to relevant sources and the process by which the statement was derived.

Thus, analysts could flexibly move between conceptual model, collection, analysis, and production. The system accompanies the analyst from requirements to product in a single platform, speeding up the process, as expressed in one analyst's comment:

If I had something like that, I'd be blazingly fast. I mean I would be able to do this 10-week project in three weeks.

Interestingly, our suggestions reiterate the findings of other researchers who identified the importance of unifying disparate tools in a different domain. In an observational study of the scientific data analysis process, Springmeyer et al [24], concluded that "an effective data analysis environment should provide an integrated set of tools which supports not only visualization, but some of the additional functionality" such as capturing the context of analysis and linking materials from different stages of analysis.

8 CONCLUSION

In this paper, we described an empirical study to understand intelligence analysts and their processes. We observed three teams of student analysts working on typical intelligence problems. Our contributions include documentation of the processes and methods they followed, clarification of issues regarding the intelligence process, and design implications for visual analytics systems for intelligence analysis.

The study has several limitations. We followed only three teams (14 analysts). Also, the analysts were not working professional analysts, but were student analysts-in-training. The analytic questions studied were from strategic intelligence, one type of analysis. Possible future work includes the study of more cases, particularly with professional analysts working on similar or other types of intelligence problems. Of course, the design implications can serve as motivation for new visual analytics systems, ideally created through participatory design with analysts.

REFERENCES

[1] D. C. Andrus. The wiki and the blog: toward a complex adaptive intelligence community. In *Studies in Intelligence*, 49(3), September 2005.

[2] E. Bier, S. Card, and J. Bodnar, Entity-based collaboration tools for intelligence analysis. Proceedings of *IEEE VAST'08 (Columbus, Ohio, October 19-24, 2008)*, pages 99–106. IEEE Press, October 2008.

[3] K. Charmaz. Qualitative interviewing and grounded theory analysis. In *Handbook of Interview Research: Context and Method*, (J. F. Gubrium and J. A. Holstein, eds.), CA: Sage Publications, 2002.

[4] G. Chin, O. A. Kuchar, and K. E. Wolf. Exploring the analytical processes of intelligence analysts. Proceedings of *ACM CHI'09* (Boston, Massachusetts, April 4-9, 2009), pages 11–20. ACM Press, April 2009.

[5] Department of Intelligence Studies at Mercyhurst College. http://intel.mercyhurst.edu/ (accessed March 15, 2011).

[6] A. B. Eli and J. Hutchins. Intelligence after Intellipedia: improving the push pull balance with a social networking utility. Research Report in *Information Science and Technology Directoratet*. Defense Technical Information Center, February 2010.

[7] D. Gotz, M. X. Zhou, and Z. Wen. A study of information gathering and result processing in intelligence analysis. In *IUI 2006 Workshop on IUI for Intelligence Analysis*, 2006.

[8] J. G. Heidenrich. The State of strategic intelligence. In *Studies in Intelligence*, 51(2), 2008.

[9] R. Heuer. Psychology of intelligence analysis. Center for the Study of Intelligence, Central Intelligence Agency: Washington, DC, 1999.

[10] R. J. Heuer and R. H. Pherson. Structured Analytic Techniques for Intelligence Analysis, CQ Press College, 2010.

[11] i2 – Analyst's Notebook. http://www.i2group.com/us (accessed March 15, 2011).

[12] R. Johnston. Analytic culture in the United States intelligence community: an ethnographic study. Central Intelligence Agency: Washington, DC, 2005

[13] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: a macrocognitive model. In *IEEE Intelligent Systems*, 21(5), September/October 2006.

[14] L. Krizan. Intelligence essentials for everyone. Joint Military Intelligence College Occasional Paper Number 6, June 1999.

[15] D. R. Millen. Rapid ethnography: time deepening strategies for HCI field research. Proceedings of *ACM DIS'00 (New York, New York, August 17-19, 2000)*, pages 280-286. ACM Press, August 2000.

[16] Mindmeister. http://www.mindmeister.com/ (accessed March 15, 2011).

[17] T. Munzner. A nested model for visualization design and validation. In *IEEE Transactions on Visualization and Computer Graphics*, 15(6), pages 921-928. IEEE Press, 2009.

[18] National intelligence estimates. http://www.cfr.org/iraq/national-intelligence-estimates/p7758#p3 (accessed June 07, 2011)

[19] D. R. Norman. Websites you can trust. In *American Libraries*, 37(7), page 36. American Library Association, Aug 2006.

[20] G. M. Olson, J. S. Olson, M. Carter, and M. Storrøsten. Small group design meetings: An analysis of collaboration. In *Human-Computer Interaction*, 7, pges 347-374. 1992.

[21] J. J. Peterson. Appropriate factors to consider when assessing analytic confidence in intelligence analysis (Master's Thesis). Retrieved from http://scip.cms-plus.com/files/Resources/Peterson-Appropriate-Factors-to-Consider.pdf on March 15, 2011.

[22] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proceedings of *International Conference on Intelligence Analysis'05 (MacLean, Virginia, 2005),* May 2005.

[23] A. Robinson. Collaborative synthesis of visual analytic results. Proceedings of *IEEE VAST'08* (Columbus, Ohio, October 21-23, 2008), pages 67–74. ACM Press, October 2008.

[24] R. R. Springmeyer, M. M. Blattner, and N. L. Max. A characterization of the scientific data analysis process. Proceedings of IEEE VIS'92, pages 235-242. IEEE Press, 1992.

[25] J. Stasko, C. Gorg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. In *Information Visualization*, 7(2), pages 118–132, 2008.

[26] A. Strauss and J. Corbin. Basics of qualitative research: Grounded theory procedures and techniques. Sage Publications: Newbury Park, Calif, 1990.

[27] G. F. Treverton. Reshaping national intelligence in an age of information. Cambridge: Cambridge University Press, 2001.

[28] Typewith.me http://typewith.me (accessed March 15, 2011).

[29] K. Wheaton. Wikis in intelligence. Unpublished manuscript, 2011.

[30] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis: concepts and methods. Proceedings of *ACM CHI '06* (Montreal, Quebec, April 22-27, 2006), pages 801–810. ACM Press, April 2006.

[31] Zotero. http://zotero.org (accessed March 15, 2011).

Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw

Carsten Görg, *Member, IEEE* Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Haesun Park,
and John Stasko, *Senior Member, IEEE*

Abstract—Investigators across many disciplines and organizations must sift through large collections of text documents to understand and piece together information. Whether they are fighting crime, curing diseases, deciding what car to buy, or researching a new field, inevitably investigators will encounter text documents. Taking a visual analytics approach, we integrate multiple text analysis algorithms with a suite of interactive visualizations in order to provide a flexible and powerful environment that allows analysts to explore collections of documents while sensemaking. Our particular focus is on the process of integrating automated analyses with interactive visualizations in a smooth and fluid manner. We illustrate this integration through two example scenarios: an academic researcher examining InfoVis and VAST conference papers and a consumer exploring car reviews while pondering a purchase decision. Finally, we provide lessons learned toward the design and implementation of visual analytics systems for document exploration and understanding.

Index Terms—Visual analytics, information visualization, sensemaking, exploratory search, information seeking, document analysis.

✦

1 INTRODUCTION

EVERYDAY, analysts and investigators confront large collections of data as they make decisions, solve problems, or simply seek to understand a situation better. Frequently, the data collections include text documents or documents with key text components. While numerical or structured data is more amenable to statistical and computational analysis, text data is conversely often messy and noisy, requiring a very sequential, slow processing (reading documents one-at-a-time, in order).

Investigators working with such document collections gather bits of information as they explore the data, hoping to form new insights about the issues at hand. Large, unstructured document collections make this task more difficult; the investigator may not know where to begin, what is important, or how concepts/events are related. The following situations are examples of these kinds of tasks:

- An academic researcher moves into a new area and seeks to understand the key ideas, topics, and trends of the area, as well as the set of top researchers, their interests, and collaborations.
- A consumer wants to buy a new car but encounters a large variety of possible models to choose from, each of

- *C. Görg is with the Computational Bioscience Program, University of Colorado, Aurora, CO, 80045; Z. Liu is with the Department of Computer Science, Stanford University, Stanford, CA, 94305; J. Kihm is with Cornell University, Ithaca, NY, 14850; J. Choo, H. Park, and J. Stasko are with the School of Interactive Computing & School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, 30332. E-mail: Carsten.Goerg@ucdenver.edu, zcliu@cs.stanford.edu, jk2443@cornell.edu, {joyfull,hpark,stasko}@cc.gatech.edu*

which has ten to twenty "professional" reviews and a web forum with hundreds of postings.
- A family learns that their child may have a rare disease and scours the web for documents and information about the condition, easily encountering many articles.
- A police investigator has a collection of hundreds of case reports, evidence reports, and interview transcripts and seeks to "put the pieces together" to identify the culprits behind a crime.

Such processes, sometimes called Sensemaking [39, 50, 54], Information Seeking Support [44], or Exploratory Search [43, 66], go beyond the initial retrieval of data or the simple return of the "right" document. Instead, they involve analysts browsing, exploring, investigating, discovering, and learning about the topics, themes, concepts, and entities within the documents, as well as understanding connections and relationships among the entities.

One approach to this problem is the computational analysis of document text, including text mining [3, 22]. However, as many researchers have noted [37, 58], simply performing computational analysis of the documents may not be sufficient for adequate understanding of a document collection—the investigator inevitably will think of some question or perspective about the documents that is either not addressed by the computational analysis or not represented accurately enough to draw a conclusion.

Another approach leverages information visualization to show information about document contents [40, 47, 59]. However, interactive visualization itself may not be sufficient for sensemaking either—as the size of the document collection grows, interactively exploring the individual characteristics of each document may simply take too much time.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. X, NO. Y

2

Our approach to the problem combines these two analytics methods: (1) automated computational analysis of the text documents and (2) interactive visualization of the documents and of the analysis results. Such a combination is described as a *visual analytics* approach [36, 58], and it leverages the strengths of both the human and the computer. Humans excel at the interactive dialog and discourse of exploration and discovery. They develop new questions and hypotheses as more and more information is uncovered. They reason about the importance of new facts that are discovered. The computer excels at complex analyses to calculate metrics, correlations, connections, and statistics about the document collection. It can rapidly analyze large collections of documents in ways that would be prohibitively time-consuming for people to do.

Relatively few systems to date have deeply and smoothly incorporated both automated computational analysis and interactive visualization while providing a tight coupling between the two. Systems (as discussed in the related work) usually focus on one of the two approaches and provide a few elements from the other. For instance, computational analysis tools sometimes provide rudimentary visualizations to depict analysis results. Alternatively, interactive visualization systems may provide a few simple analysis techniques such as filtering or statistical analysis of the data.

Elaborating on this notion, Keim et al. [36] state:

> Visual analytics is more than just visualization. It can rather be seen as an integral approach to decision-making, combining visualization, human factors and data analysis. The challenge is to identify the best automated algorithm for the analysis task at hand, identify its limits which can not be further automated, and then develop a tightly integrated solution which adequately integrates the best automated analysis algorithms with appropriate visualization and interaction techniques.

In this article we explore this coupling through Jigsaw [55], a system for helping analysts explore document collections. Jigsaw is a relatively mature system, and has garnered trial use in the field by analysts in law enforcement, investigative reporting, fraud detection, and academic research, among other areas. An initial user study of the system showed its potential in helping investigators work with documents and in supporting different analysis strategies [34].

Earlier versions of Jigsaw emphasized multiple, coordinated visualizations but provided relatively little computational analysis of documents' text. The system primarily visualized connections between entities across documents to help investigators follow trails of information. More recently, we have added a variety of automated text analyses to the system including analysis of document similarity, document sentiment, document clusters by content, and document summarization through a few words or sentences. These new analyses aid investigators in determining the documents to examine first, the documents to focus on or discard, and the documents that may be related to different investigative angles.

Our focus is not on developing novel innovative algorithms for computational text analysis. Instead, we explore ways to smoothly integrate existing computational analyses into an interactive visual interface in a seamless manner that will provide a natural and fluid user experience. Furthermore, new computational analysis algorithms frequently are developed for well-defined tasks or problems with carefully constructed inputs and data. Real-world visual analytics systems, conversely, encounter messy, noisy data and must support open-ended analytical reasoning and sensemaking. Thus, our research also examines how computational analysis techniques can be used throughout visual exploration on challenging real-world data.

The contributions of this research include (1) methods for fluidly integrating computational text analysis and visualization, (2) illustration of the utility of such an approach through two example usage scenarios, and (3) lessons learned toward the design and construction of visual analytics systems for document exploration and understanding. Additionally, we provide implementation advice and experience on the integration of text analysis algorithms as a broader benefit for other researchers.

2 RELATED WORK

Computationally-aided analysis and visualization of text and documents to assist human investigators with sensemaking has been a topic of intense research interest recently. Furthermore, different subdisciplines of computer science each bring their own focus to the problem. Thus, a comprehensive examination of related work likely would take a complete paper itself. Here, we highlight some of the existing research most strongly related to our work to provide the reader with greater context and familiarity of the varied approaches others have taken.

Systems in this area typically focus on some aspect of a document or document collection to present. Broadly, they visualize (1) metadata about the documents; (2) the document source text (words); (3) computed features and attributes of the documents including entities; and/or (4) general concepts, themes, and models across the documents. Visualization techniques have been developed for single documents or large collections of documents, though the techniques for individual documents often can be generalized to collections.

Systems with a specific focus on helping people understand various attributes of an academic paper collection are a good example of presenting **metadata about a set of documents**. PaperLens [40] employs a variety of bar chart, list, graph, and text-based visualizations to show author, topic, and citation data of past CHI papers. The system uses a clustering analysis to help group papers by topic as well. Selecting an author, paper, or concept in one visual representation loads related items into the other visualizations. A follow-on system, NetLens [33], focuses on visualizing content-actor data within document collections such as scientific publications and authors. NetLens uses bar charts, histograms, and lists to represent the data and help analysts understand statistics and trends from conference papers and their citations.

A number of innovative visualization techniques have been developed to represent the **words and source text of documents**. The SeeSoft system [21] represents a line of a text document by a row of pixels on the screen, with the length of the text line (number of characters) mapped to the length of the row of pixels. The goal of the technique is to visually depict

documents that are larger than what can normally be shown on one screen. Other well-known source text visualization techniques such as TextArc [47], Word Clouds [61], Word Trees [64], and Phrase Nets [59] actually still show text, unlike SeeSoft. They also show frequency and relationships of particular words or terms within documents.

Many systems, in fact, inhabit a conceptual space that transitions from visualizing document source to visualizing **computed metrics or features of a document or documents**. For example, Themail [60] analyzes collections of email messages using a variant of the term-frequency inverse document-frequency (TF-IDF) algorithm that focuses on each sender. The system's visualization is temporally-based and shows lists of keywords from the emails to characterize the main topics of messages during each month and over entire years.

Other techniques such as Arc Diagrams [63], DocuBurst [12], and Parallel Tag Clouds [13] compute metrics about a set of documents and visualize the computed metrics in unique ways. The PaperVis system [10] combines a relevance-determination algorithm with visualization to show relationships among academic articles. PaperVis performs citation and keyword analysis and presents the results through bullseye and radial space filling visualizations. The size of a node (document) and its distance to other documents denote its importance and relevance, respectively.

Keim and Oelke [38] perform numerous text analysis measures not seen in other document analysis systems including measures such as average word length, sentence length, number of syllables per word, and other measures such as Simpson's index and Hapax Legomena and Dislegomena (number of words occurring once and twice). The visualization of the analysis results for each of these measures uses a heatmap style display. Together with colleagues they subsequently added sentiment analysis to their measures [45] and added node-link network visualization to communicate relationships among the documents' sentiments [46].

One particular computed attribute sometimes visualized by systems is an entity within a document or documents. Identifying entities may be as simple as looking for particular strings or expressions within a document's text or it may involve complex computations to determine unique entities and their types. Different systems then choose to visualize the results of the computation in unique ways.

FeatureLens [19] uses text mining to identify words or expressions that recur across a set of documents. The system presents lists of the frequently occurring words and expressions, small overview rectangles representing each document with term positions identified by small marks, graphs of appearance count across documents, and textual views with terms highlighted. Primary users of the system may be literary scholars or journalists reviewing books or speeches. A follow-on system, POSVis [62], performs word-based part-of-speech analysis on documents and then displays the results using pixel-based overviews, word clouds, and network diagrams.

Entity Workspace [4] focuses on entity-based analysis and provides a "snap-together" visualization of entities and their attributes. Its analysis capabilities include spreading activation techniques to calculate degree-of-interest for the entities.

The IVEA system [57] uses entities of interest from a document collection to support faceted navigation and browsing across the collection. The system employs a matrix-style visualization with semantic zooming to represent the facets within documents.

Another set of systems move beyond the calculation of specific features, entities, or linguistic metrics of documents. These systems employ sophisticated text mining techniques to compute document **models and abstractions, often including concepts or themes across the documents**. Models and abstractions become especially useful as the size of the document collection grows.

The ThemeRiver technique [28] uses a river metaphor to represent temporal themes across a document collection. The river visualization extends from left-to-right to show the chronological progress of documents, and individual currents (colored bands) within the river represent different concepts. The vertical width of a current portrays its strength at a certain point in time.

Document topic modeling through latent Dirichlet allocation (LDA) [6] has become a popular technique for driving visualizations of document collections. TIARA [41] performs LDA analysis to identify themes throughout documents, and it portrays the results using a ThemeRiver-style visualization that has been augmented with word clouds. The system thus shows how topics grow and decline in focus over time. The system also supports user interaction to drill down and provide more detail on concept regions and to see the actual documents (emails) generating the concepts. TIARA can be used in many domains such as consumer reviews, email, and news. TextFlow [14] extends TIARA, showing how topics emerge and merge over time, how keywords relate to topics, and critical events within topics.

Parallel Topics [20] also employs LDA to model topics across a document collection and uses a ThemeRiver style visualization to present the results, coupled with a Topic Cloud to show important terms within topics, and a parallel coordinates visualization to show how individual documents contribute to the different topics. Other systems use LDA but provide different visualizations of the identified topics including word and topic lists [9], word clouds and sentences [23], force-directed networks [27], or custom-designed 2-D projections [11].

The FacetAtlas system [8] helps an analyst understand relationships between entities and facets within collections of documents sharing traits similar to academic articles. FacetAtlas uses a density map-style visualization with bundled edge connections between facets and entities along with rich interactive operations to present complex relationships between concepts in a document collection. Users can either search for specific concepts or interactively explore through the visualization interface.

The IN-SPIRE [26, 30] system takes a different approach to visualizing document themes. It utilizes powerful automated analysis, clustering, and projection capabilities, primarily operating at the document level. IN-SPIRE computes high-dimensional similarities of documents and then visualizes these relationships through galaxy or themescape style pro-

jected representations that show the documents grouped into multiple clusters.

Finally, some visual analytics systems focus not on unique visualizations of text and documents but on creating environments where an analyst can analyze and reason about the documents. Often these systems use visual representations to help analysts explore the documents and develop hypotheses, and their target domain is frequently intelligence analysis. The systems' main goal typically is to give an investigator a faster, better understanding of a large collection of documents to help understand plots, themes, or stories across the collection.

nSPACE/TRIST/Sandbox [32, 67] provide sophisticated document analysis including entity identification and relations, trend analysis, clustering, and other automated operations. The systems present the documents through views of the documents' text or via groups of documents as small icons, but they augment this representation with sophisticated user interface flexibility for analysts to reason and develop stories about the data.

Commercial tools such as i2's Analyst Notebook [31] help intelligence, law enforcement, and fraud investigators work with document collections, among other types of data. Analyst's Notebook primary visualization is a node-link graph that shows connections between key entities in an investigation. Typically, however, the human investigator establishes these connections and constructs linkages.

As we will show in the following sections, *our contribution beyond this vast body of related work centers around the breadth of computational analysis techniques paired with a suite of rich interactive visualizations and integrating the two in a fluid, consistent manner*. Jigsaw provides multiple, varied perspectives to portray analysis results that allow the investigator to rapidly explore documents in a flexible manner. The particular emphasis on communicating entity connections across documents within concept-, temporal-, and sentiment-based perspectives also distinguishes it from existing systems.

3 COMPUTATIONAL TEXT ANALYSES

An earlier version of Jigsaw, described in [55], focused on interactive visualization support rather than on computational modeling and analysis of documents' text. In an evaluation study [34] we found that the system was overall useful and supported a variety of strategies participants used to conduct their investigations on a document collection. However, we also found a number of situations in which the participants might have benefitted from additional information provided by computational text analysis, especially to get started with their investigation.

Some participants first read many of the documents to gain familiarity with the collection. Automated text summarization could have helped them to speed up the initial reading by reducing the amount of text to examine; document metrics, such as documents' date or length, could have provided order and structure to make the initial familiarization more efficient. Other participants focused early in their investigation on certain entities and tried to learn everything about them.

Document similarity measures or features for recommending related documents could have supported this task by highlighting related information in other documents; showing documents clustered by content also could have helped them to step back and see the topics already examined or overlooked. Another group of study participants first randomly selected a few documents for acquiring evidence on which to start their investigation. Clustering documents by content could have been beneficial to help them to choose documents from different clusters for broader initial evidence.

We made similar observations on the potential benefit of computational analyses from our own use of Jigsaw, especially through our participation in the VAST Contest and Challenges [24, 42], as well as from other researchers' use of the system [49, 53]. In addition, we noticed that sentiment analysis would be another useful computational technique since product reviews are a natural document set for an investigation.

Computational text analyses are not without their own set of issues and concerns, however. As Chuang et al. note, text mining algorithms generate *models* of a document collection to be visualized, as opposed to source data about the documents [11]. When models are presented to the analyst, *interpretation* and *trust* arise as important concerns. In Jigsaw, we use an extensive suite of interactive visualizations to provide multiple perspectives on analysis results, thus enabling the analyst to review and explore the derived models and determine which are most helpful.

We now describe the suite of computational analyses added to the system and, most importantly, we focus on how the analyses integrate with different visualizations. First, we explain each analysis measure and how Jigsaw presents its results. Subsequently, we provide two example usage scenarios that illustrate how an analyst explores a document collection with the system (Section 4) and we present the implementation details of the analysis algorithms (Section 5). Our main focus has been on developing techniques for smoothly combining the computational analyses with interactive visualizations. We have emphasized an integrated user experience throughout, one that provides information where and when it is most helpful and that ideally feels natural and coherent to the analyst using it for an investigation.

Document Summarization

Jigsaw provides three different techniques to summarize a document or a set of documents: one sentence summaries, word clouds, and keyword summaries. A one sentence summary— a determination of the most significant sentence—of a single document helps analysts first to decide whether to read the full text of the document and subsequently to recall the content of a document read earlier. Jigsaw presents a one sentence summary above the full text of each document in its Document View (Figure 4). Additionally, the one sentence summary appears via tooltip wherever a document is presented through icon or name. Word clouds, the second type of document summary, help analysts to quickly understand themes and concepts within sets of documents by presenting the most frequent words across the selected documents. Jigsaw presents

word clouds of selected documents in its Document View and flexibly allows a fewer or greater number of words to be shown. The final type of summary, keyword summaries of document sets, labels sets of grouped documents in the Document Cluster View (Figure 5) and Document Grid View (Figure 6) in order to help an analyst know what the group is about. Keyword summaries are based on different metrics: word frequency in each set, word uniqueness across sets, or a combination of both. Summaries based on word frequency help to understand the content of each set, word summaries based on uniqueness help to analyze differences among sets. Jigsaw allows the analyst to interactively change the metric chosen. Overall, document summarization helps an analyst to quickly decide whether a document (or set of documents) is relevant for a specific task or question at hand and whether it should be investigated further.

Document Similarity

The similarity of two documents is measured in two different ways in Jigsaw: relative to the text within the documents or to the entities connected to the documents. The latter similarity measure is of particular interest for semi-structured document collections such as publications in which metadata-related entities (e.g., authors, years, and conferences) are not mentioned in the actual document text. Document similarity measures help an analyst to determine if a document is unique (an outlier in the collection) or if there exist related documents that should be examined as well. We implemented a new view in Jigsaw (the Document Grid View) to present, analyze, and compare document similarity measures. The view organizes the documents in a grid and provides an overview of all the documents' similarity to a selected document via the order and color of the documents in the grid representation. In all other views showing documents, an analyst can retrieve and display the five most similar documents to any document through a simple menu command.

Document Clustering

Clustering of similar and related documents also is based on either document text or on the entities connected to a document. Clusterings can be either computed fully automatically (using default values for the parameters of the clustering algorithm), or the analyst can specify the number of clusters and themes within clusters by selecting seed documents. Additionally, the analyst can interactively change clusters and define new clusters based on identified entities or keyword searches across the document collection. Document clustering partitions the documents into related groups to help an analyst explore the collection more systematically. Jigsaw presents clusterings in its Document Cluster View. The Document Grid View also provides an option to organize the documents by cluster when showing document metrics.

Document Sentiment Analysis and Other Metrics

Jigsaw computes a document's sentiment, subjectivity, and polarity, as well as other attributes such as a document's length and its number of connected entities. These metrics help an analyst seeking documents that are particularly high or low in key attributes. Jigsaw integrates and presents these metrics in its new Document Grid View. One metric can be used to

determine the order of the documents within the grid, and a second metric (or the first metric again) can be mapped to the documents' color. The combined representation of any two of these metrics (by the documents' order and color) provides a flexible and powerful analytical view.

Identifying Entities in the Documents

The initial version of Jigsaw used a statistical entity identification approach from the GATE [15] package. We have added additional packages for automated entity identification and Jigsaw now provides three different approaches for automatically identifying entities of interest in text documents: (1) statistical entity identification, (2) rule-based entity identification, and (3) dictionary-based entity identification. It uses statistical approaches from GATE, Lingpipe,[1] the OpenCalais webservice,[2] and the Illinois Named Entity Tagger [52] to identify a variety of entity types, including person, place, organization, date, and money. For the rule-based approach we define regular expressions that match dates, phone numbers, zip codes, as well as email, web, and IP addresses. The dictionary-based approach allows analysts to provide dictionaries for domain-specific entity types that are identified in the documents using basic string matching.

The automatic identification of entities is still error-prone, especially in noisy, real-world data. Therefore, Jigsaw also provides functionality to correct errors in the set of identified entities. Within different visualizations, an analyst is able to add entities that were missed (false negatives), remove entities that were wrongly identified (false positives), change the type of entities, and define two or more entities as aliases.

Recommending Related Entities

To find embedded connections among entities (that might be connected via a long chain of other entities and documents) Jigsaw recommends related entities for further examination. The recommended entities are computed by searching for connecting paths between two or more entities in the document-entity network. The chain(s) of connected entities and documents are presented in the Graph View.

4 INVESTIGATIVE SCENARIOS

To better understand how these computational analysis techniques operate within Jigsaw and aid an investigation, we present two example use scenarios: a researcher exploring academic publications to learn about a research area and a consumer exploring product reviews to help make a purchase. The two scenarios involve relatively small document collections (in the hundreds) in order to make the presentation here more concise. We have used Jigsaw on larger collections numbering in the thousands of documents, however, and have found the new computational analysis capabilities to be even more useful at this larger scale. Because the static descriptions in this article cannot adequately convey the dynamic nature of the investigator's interaction with the system, we refer the reader to the accompanying videos for further illustration and elaboration of similar scenarios.

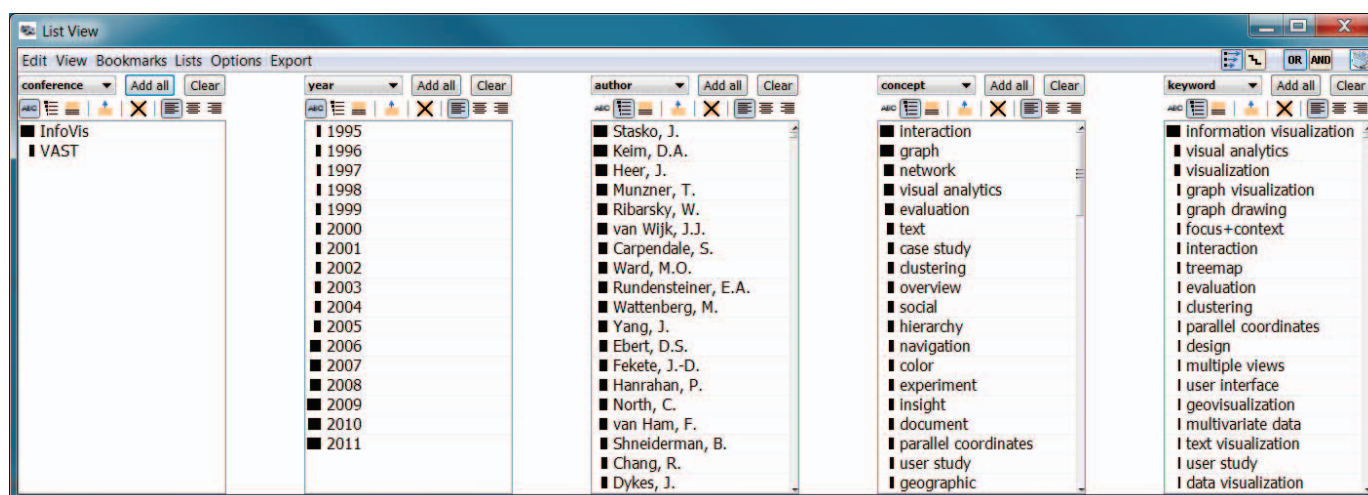1. http://alias-i.com/lingpipe
2. http://www.opencalais.com

Fig. 1. List View showing conference, year, author, concept, and keyword, with the last three sorted by frequency.

4.1 Investigative Scenario: InfoVis and VAST Papers

In this scenario we illustrate an investigation of a dataset involving all of the IEEE InfoVis and VAST conference papers from 1995 to 2011; the InfoVis conference has run from 1995 to 2011 and VAST from 2006 to 2011. The dataset contains 578 documents, one for each paper presented at either of the conferences; each document includes the title and abstract of the article it represents; its entities are the paper's authors, index terms, keywords, conference, journal, and year. Additionally, we added an entity type "concept" including 80 domain-relevant terms such as *interaction*, *brushing*, *network*, and *evaluation* to be found within the articles' titles and abstracts.

To generate this dataset, we gathered information about the papers from the IEEE Digital Library. Throughout the data gathering process we performed a few cleaning steps and we resolved aliases for authors. We unified each unique author to one specific name because it was not uncommon to find initialized names or inconsistent inclusion of middle names. For keywords, we unified terms effectively meaning the same thing to one common string identifier. For example, the terms "Treemap", "tree-map", "treemaps", all were changed to the string "treemap". Jigsaw's List View (Figure 1) was very useful in this data cleaning phase as we could enumerate all the instances of any entity type in alphabetical order and easily check for similar strings. Additionally, we identified a set of documents to serve as seeds for clustering the documents.

Clearly, our domain knowledge helped in this initial data cleaning and entity resolution. Such transformations are typically necessary in any analysis of semi-structured text document information [2]. Jigsaw allows the results of such a process to be saved as an XML data file for sharing with others. In fact, we have made this conference paper dataset available on the web.[3]

For the purpose of this scenario, we introduce a hypothetical academic researcher, Bill, who works in the database area. Bill has developed a new technique for representing database schemata as graphs or networks and he has worked with a student to build a visualization of it. Bill knows a little

about visualization research but not much detail about the IEEE InfoVis and VAST Conferences. He would like to learn whether one of these conferences would be a good fit for his paper, and if so, which one. Questions such as the following naturally arise in such an investigation:

- What are the key topics and themes of the two research areas?
- Have these topics changed over the history of the conferences?
- Who are the notable researchers in the different areas?
- Which researchers specialize in which topics?
- Are particular topics relating to his work present?
- Are there specific papers that are especially relevant?

Bill starts the investigation by examining statistics about the dataset to gain an overview of the conferences and areas. Jigsaw's Control Panel (not shown here) indicates that 1139 different researchers have contributed papers. These authors self-identified 1197 keywords and IEEE designated 1915 index terms for the papers. 78 of the 80 concepts (we generated) appeared in at least one title or abstract.

After gaining a general overview, Bill wants to learn more specifics about the key topics and authors so he opens Jigsaw's List View (Figure 1). He displays conference, year, author, concept, and keyword, then changes the list ordering from alphabetic to frequency-of-occurrence on the final three entity types to see the top-occurring entities. The small bar to the left of each entity denotes the number of documents in which it occurs. The general terms *information visualization* (101 occurrences), *visual analytics* (42), and *visualization* (40) are unsurprisingly the most frequent author-identified keywords. More interesting are the next most-common terms: *graph visualization* (18), *graph drawing* (17), *focus+context* (16), *interaction* (16), *treemap* (16), *evaluation* (14), *clustering* (13), and *parallel coordinates* (13). The term *interaction* (96) was the most frequent concept found in titles and abstracts, followed by *graph* (91), *network* (63), *visual analytics* (63), *evaluation* (55), and *text* (43). While these notions are likely familiar to someone within the field, they help a relative outsider such as Bill to understand some of the most important ideas in the research area.
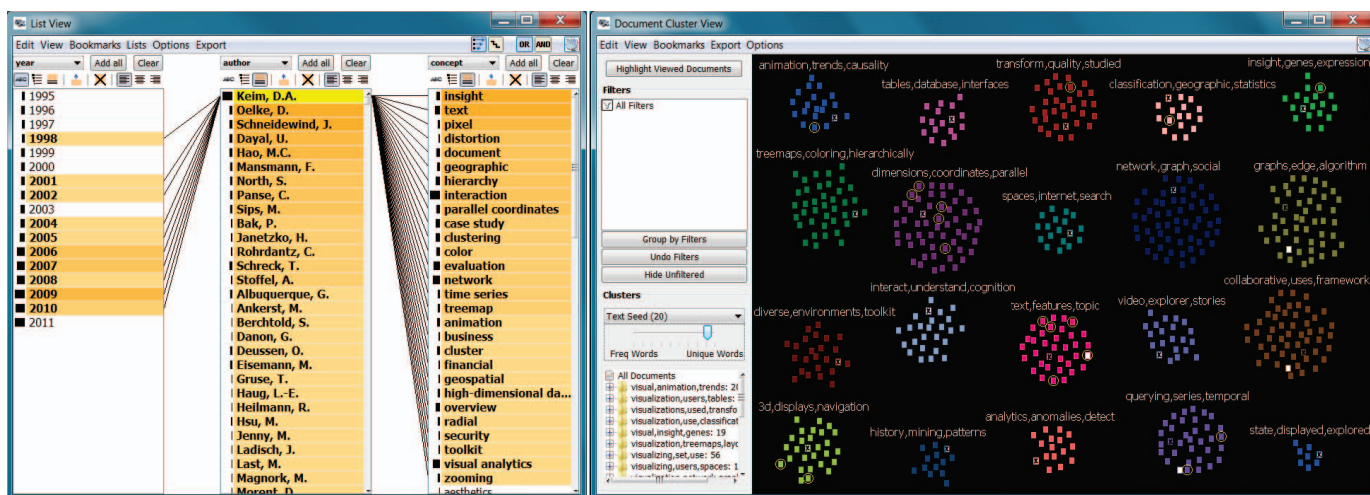
Fig. 2. List View (left) showing years, co-authors, and concepts connected to *Keim*. Document Cluster View (right) showing different clusters of related papers (small rectangles in different colors). Papers authored by Keim are selected (surrounded by a yellow circle).

Examining the author list, Bill notes that his old friend from database research, *Daniel Keim*, is one of the very top authors at the conferences. Bill is curious about Keim's papers at the conferences and decides to explore this further. He selects *Keim* in the List View and reorders the author and concept lists by strength of connection to that selection in order to see the entities most common with him (Figure 2, left). Connections in Jigsaw are defined by document co-occurrence, either of identified entities in the document text, such as concepts, or of meta entities of the document, such as authors. Connection strength is defined by the number of document co-occurrences: more co-occurrences signify stronger connection. (Further details of Jigsaw's connection model are described in [55].) The List View highlights entities connected to the selection via an orange background, with darker shades indicating stronger (more frequent) connections. Entities with white backgrounds are not directly connected. The terms *insight*, *text*, *pixel*, *distortion*, *document*, and *geographic* are the most connected concepts. Keim's most frequent co-authors are *Oelke*, *Schneidewind*, *Dayal*, *Hao* and *Mansmann*; he has published frequently from 1998 to 2010.

Bill now wants to explore ideas related to his own research. He notes that the concepts *graph* and *network* are the second and third most frequent, suggesting his work might be a good fit for these conferences. He selects the concept *graph* to learn which authors work on the topic. Jigsaw shows the most connected authors *van Ham*, *Abello*, *Hanrahan*, *Munzner*, and *Wong* and illustrates (dark shade of orange for recent years) that this has been a strong topic recently (Figure 3). Selecting *network* shows the most connected authors *Brandes*, *Ebert*, *Fekete*, *Hanrahan*, *Heer* and *Henry Riche* and that the topic also has been important recently. Surprisingly, the two author lists have many different names, which puzzles Bill since the two topics seem to be closely related.

To investigate further and gain a better understanding of the different topics within the conferences based on the articles' titles and abstracts, Bill switches to the Document Cluster View that displays each document in the collection as a small
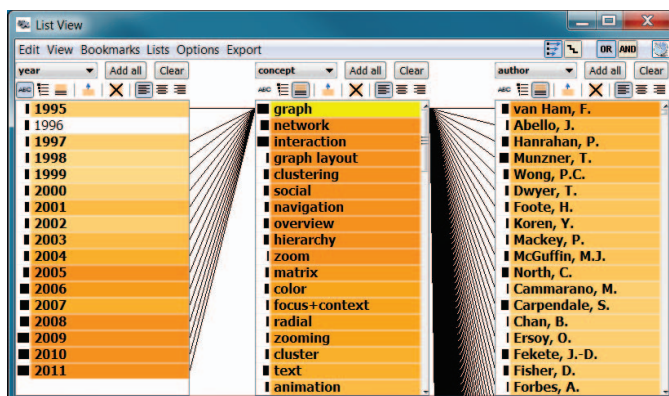

Fig. 3. List View with the concept *graph* selected, showing strongly connected years, concepts, and authors.

rectangle. Upon starting Jigsaw, Bill ran Jigsaw's automated computational analyses that calculated the similarities of all documents and a set of clusters based on these similarities.

The Document Cluster View (Figure 2, right) shows the 578 papers divided into 20 clusters resulting from the cluster analysis. The groups are each assigned a different color and are labeled with three descriptive keywords commonly occurring in the titles and abstracts in each cluster. If the summary terms are selected based solely on their frequency, common terms such as "data" and "visualization" represent many clusters which likely is not useful. The Cluster View provides a word frequency slider (left, lower center) for the investigator to interactively modify to show either more common or more unique terms affiliated with each cluster. Bill moves the frequency slider to the right, thus labeling clusters with terms more unique to that cluster. The resulting cluster labels represent important topics in these areas including toolkits, treemaps, text, animation, parallel coordinates, social networks, 3d, and databases (Figure 2, right).

Bill is curious which clusters his friend Daniel Keim's papers fall into. He applies cross-view selection and filtering [65], one key capability of Jigsaw. It can, for example, show the topics (clusters) in which an author publishes simply
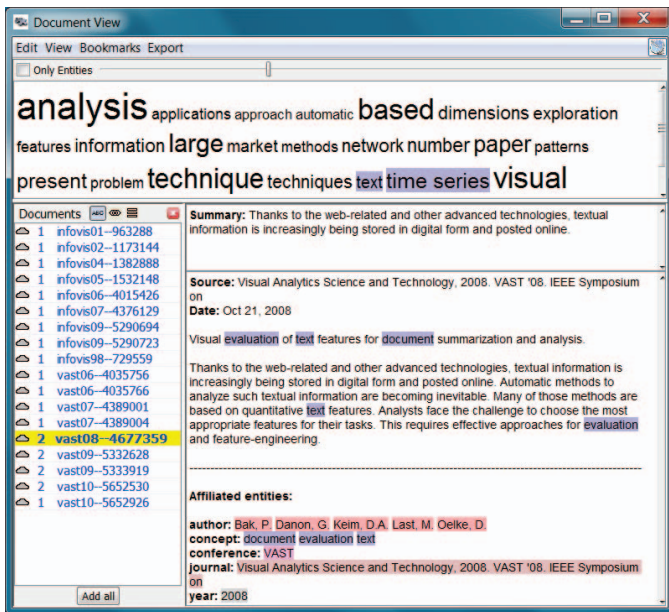
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. X, NO. Y

8

Fig. 4. Document View showing all the papers authored by *Keim*. Above the selected document's text (right) is a one sentence summary and below are the affiliated entities. The word cloud (top) summarizes all documents loaded in the view.

by selecting that author in any other view. Selecting *Keim* in the List View (Figure 2, left) immediately updates the Cluster View (Figure 2, right) and highlights (yellow circles around document rectangles) the papers *Keim* has authored. As shown in the figure, his work is relatively focused with five papers each in the "dimensions, coordinates, parallel" and "text, features, topic" clusters, and eight other papers scattered among six other clusters. Knowing Keim's research, Bill is quite surprised to see none of his papers in the cluster with "database" as a descriptive word. He decides to load all of Keim's papers into a Document View to examine them more closely.

The Document View (Figure 4) presents a list of documents (left) with the selected (yellow highlight) document's text and related information presented to the right. Below the text are the associated entities and above the text is the one sentence summary of the document computed by Jigsaw's summary analysis (described in Section 5.2). The word cloud at the top shows the most common words (with highlighted keywords and concepts) in the abstracts of these loaded papers. Bill reviewed all the papers quickly and noticed that indeed none were about database research. He grows a little concerned about whether these conferences would be a good fit for his paper.

Next, Bill wants to understand the evolution of topics in the conferences over time to learn which have waned and which have been growing in importance recently. To do so, he selects the first four years (1995 to 1998, all InfoVis) in the List View and notices strong connections to the "internet", "toolkit", and "3d" clusters in the Cluster View; additionally the List View shows strong connections to the concepts *interaction*, *case study*, *navigation*, and *animation*, with the concepts *network* and *graph* as the sixth and seventh most frequent. Selecting
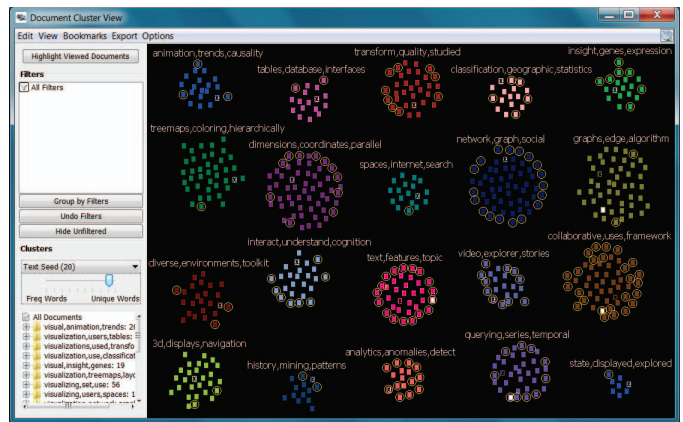

Fig. 5. Document Cluster View with the VAST Conference papers highlighted. Note the clusters where they provide a strong presence.

the most recent four years (2008 to 2011, both InfoVis and VAST) illuminates strong connections to multiple clusters but only connections to one document in the "3d" cluster and to two documents in the "internet" cluster. These topics clearly have waned over time. The terms *graph* and *network* are each in the top five connected concepts; thus, Bill sees how they have remained strong notions throughout the history of the conferences.

Bill next wants to better understand how the two conferences differ, so he explores the key concepts and ideas in each. He selects each conference, one at a time, in the List View and observes the connections. Among the ten most common concepts for each conference, five terms appear in both: *interaction*, *network*, *evaluation*, *graph*, and *case study*; the five other unique terms for InfoVis are *overview*, *hierarchy*, *color*, *navigation*, and *experiment*, and for VAST are *visual analytics*, *text*, *collaboration*, *clustering*, and *insight*. As shown in Figure 5, VAST papers (far fewer in number) occupied more than half of the "analytics, anomalies, detect", "video, explorer, stories", and "collaborative, uses, framework" clusters. These simple interactions help Bill begin to understand the subtle differences in the two conferences. His work still appears to fit well into either, however.

To learn more about the papers potentially related to his own work, Bill uses cross-view filtering in an opposite manner as he did earlier. He selects an entire cluster in the Document Cluster View and observes the resulting connections in the List View. For example, selecting the potentially related "network, graph, social" cluster shows that *Shneiderman*, *Fekete*, *Henry Riche*, *McGuffin*, *Perer*, and *van Wijk* are highly connected authors to its papers. Another potentially related cluster to Bill's work, "graphs, edge, algorithm" has top authors *Koren*, *Munzner*, *Abello*, *Ma*, and *van Ham*, all different than those in the previous cluster.

Bill decides to explore the papers in the "graphs, edge, algorithm" cluster. Since there are many, he moves his mouse pointer over the small rectangles in that cluster to quickly read a one sentence summary (tooltip) of each document. This document summary tooltip is available in other views such as the Document Grid View (Figure 6) and the Graph View where small iconic representations of documents are shown. None of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. X, NO. Y   9
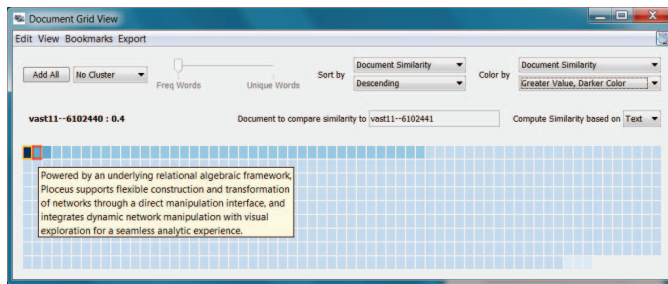
Fig. 6. Document Grid View with the document (small rectangle) order and shading set to correspond to the document's similarity to the selected Orion paper.

the papers in this cluster seem to be relevant to his research; they are not about the general representation of structure and relationships in networks but about specific details of layout techniques and their mathematical optimizations. Therefore, he moves on to the "network, graph, social" cluster. Here, he discovers a paper whose summary sparks his interest: "Despite numerous advances in algorithms and visualization techniques for understanding such social networks, the process of constructing network models and performing exploratory analysis remains difficult and time-consuming." Bill decides to load all the papers from this cluster into a Document View and selects this paper's icon in the Document Cluster View, thus also displaying it in the Document View. He reads the abstract of the VAST '11 paper by Heer and Perer about their Orion system and notices that it is definitely related to his work.

Bill now wants to know if papers similar to the Orion one have been published at the conferences. To find out, he uses Jigsaw's Document Grid View. The Document Grid View displays all the documents and is able to sort them by various text metrics, one being similarity to a base document. The Document Grid View in Figure 6 shows the similarity of papers compared to the Orion paper.

Bill decides to examine the most similar papers more closely, so he selects the eight most similar ones and displays them in a Document View (Figure 7). He observes that four of the eight papers are from InfoVis and four are from VAST. However, the paper most similar to the Orion paper is also from VAST '11 and is titled "Network-based visual analysis of tabular data." Upon reading the abstract, Bill learns that his work is quite similar to that done in this paper. Thus, he has both found some very relevant related work to explore further and he has determined that his new paper likely would fit in either conference, but VAST may be a slightly better match.

Through this abbreviated scenario, we illustrated how Jigsaw's analysis and visualization capabilities help analysts to gain quick insight on places to start an investigation, to learn about the key entities and topics in certain areas, and to explore connections and relationships in more depth. We also showed how it helps identify leaders, rapidly summarize sets of documents, compare and contrast information, find similarities and differences, and determine what should be investigated in more depth at a later point.

As shown in this scenario, investigative analyses of textual documents are often open ended and explorative in nature:
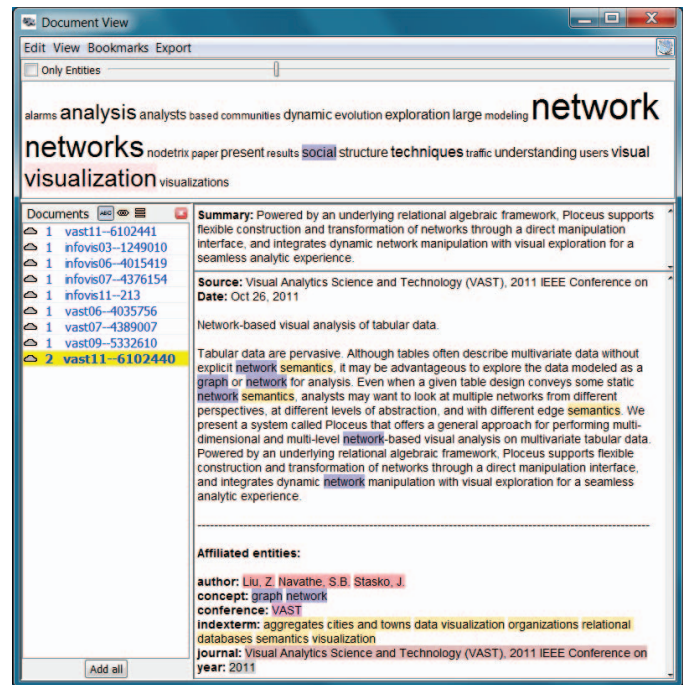

Fig. 7. Document View with the eight most similar papers to the Orion paper loaded. Selected here is the most similar document.

detailed questions or precise hypotheses may not be known at the beginning of an investigation but rather arise and evolve as the investigation unfolds. Analysts often switch back and forth between analyzing general trends, such as examining key topics, their relationships, and how they change over time, and more focused explorations about specific entities. Formulating new questions and finding supportive as well as contradictive evidence are fundamental tasks throughout these types of investigations.

4.2 Investigative Scenario: Car Reviews

The next scenario illustrates a different kind of investigation using documents—a consumer, Mary, who is shopping for a car. A colleague is selling his 2009 Hyundai Genesis, so to learn more about this particular model Mary examines a document collection consisting of 231 reviews of the car from the edmunds.com website. Mary wants to gain a general sense of consumers' views of the car and determine whether she should buy it. Specific concerns and goals that have arisen in her mind include:

- Identify and understand the important topics being discussed throughout the reviews,
- Learn the strong and weak points of the car,
- Determine whether perceptions of the car have improved or weakened over time,
- Identify the key competitive makes/models of cars,
- Judge whether particular attributes of the car such as its gas mileage, power, sound system, and reliability are good.

Mary could, of course, examine these 231 reviews one-by-one from the website just as anyone could do when exploring
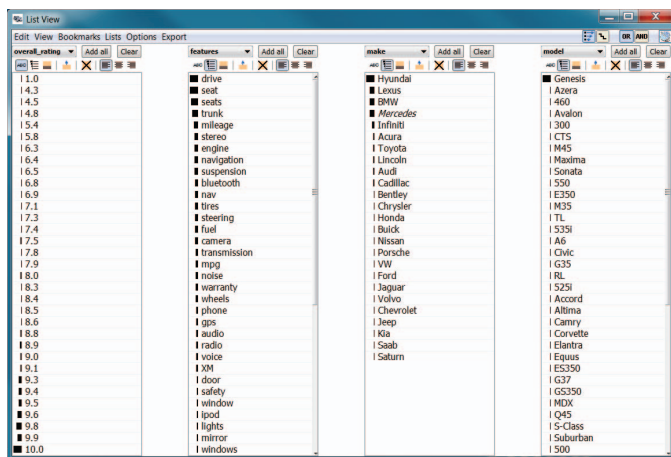
Fig. 8. List View showing the overall rating, feature, make, and model entity types and their values from the reviews. The last three are sorted by frequency.
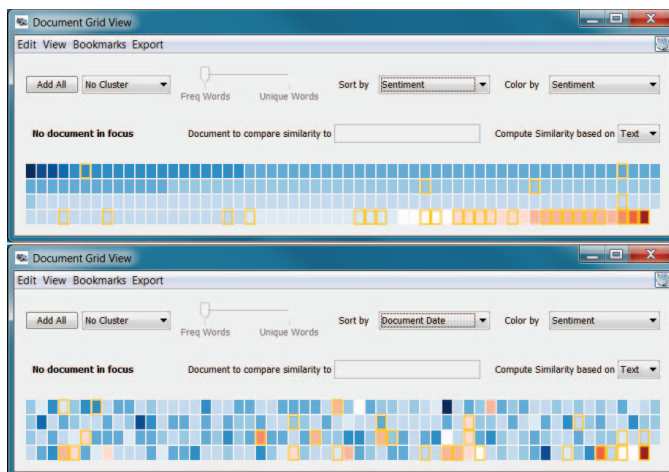

Fig. 9. Document Grid View showing all the reviews colored by sentiment: blue indicates positive, white neutral, and red negative. The top view displays the documents sorted by sentiment as well, while the bottom view shows them ordered by date ranging from the top-left (oldest) to bottom right (newest).

a collection of consumer reviews or webpages retrieved from a search engine. However, this process is tedious and may not illuminate well the key themes and connections across the reviews.

For illustrating Mary's use of Jigsaw in this scenario, we scraped reviews of the 2009 Genesis from the edmunds.com website and imported them into Jigsaw. Each review, including its title and main narrative written by a consumer, is represented as a document. The document's entities include various rating scores (e.g., exterior design, fuel economy, and reliability) that the review author explicitly designated. We also calculated an overall rating that is the average of all the individual ratings. We added three other entity types to be found within the document text (title and review narrative): car make (e.g., *Audi*, *Ford*, *Lexus*), car model (e.g., *525i*, *Avalon*, *ES350*), and car "feature", for which we defined 57 general terms about cars such as *seat*, *trunk*, *transmission*, and *engine*.

To get an overview of the reviews, Mary begins her investigation by invoking Jigsaw's List View (Figure 8). She displays the overall ratings from consumers, as well as the features, makes, and models discussed in the reviews, each sorted by frequency. Mary notices that the review ratings are generally high (indicated by longer frequency bars near the bottom of the first list); *drive*, *seat(s)*, *trunk*, and *mileage* are the most mentioned features; *Lexus*, *BMW*, *Mercedes*, and *Infiniti* are the most mentioned makes (excluding *Hyundai* itself); and *Azera*, *460*, *Avalon*, *300*, and *CTS* are the most mentioned models (excluding *Genesis* itself). This is useful information to know about the key competitive cars and most commented-upon features of the Genesis.

Although the ratings are generally good for the car, Mary wants to know more details about reviewers' thoughts. An analysis of the sentiment [26] of the reviews is useful here. To calculate sentiment, Jigsaw uses a dictionary-based approach, searching for positive or negative words throughout the document text. Here, Mary uses Jigsaw's capability to augment the dictionary by domain-specific words. For example, terms such as "quiet" and "sweet" are positive car sentiment words, while "lemon" and "clunk" indicate negative sentiment. Mary opens the Document Grid View and orders and colors the reviews

by sentiment (Figure 9, top). Positive reviews are colored blue and shown first, neutral reviews are colored white and appear next, and negative reviews are colored red and shown last. Darker shades of blue and red indicate stronger positive and negative sentiment, respectively. At first glance, the reviews for the Genesis appear to be positive overall, roughly mirroring the overall rating scores shown in the List View.

Mary once had a car that developed a number of problems after a year of driving it, so she is curious what the most recent reviews of the car express. Thus, she changes the order of the reviews in the Document Grid View to be sorted by date, as shown in Figure 9, bottom. The oldest review from 06/26/2008 is placed in the top-left position in the grid and the most recent review from 07/24/2011 is in the bottom-right position. The view indicates that the earlier reviews were generally positive (shaded blue) but the more recent reviews begin to show more negative (red) perceptions. The most recent review is, in fact, the most negative, which is a concern. This trend might indicate that some issues with the car were not apparent when it first appeared but were revealed over time as the car matured.

To learn more about the car's potential weaknesses, Mary sorts the feature entities in the List View by their strength of connection to these negative reviews with overall rating below 8 (Figure 10). The terms *seat*, *tires*, *transmission*, *steering*, and *suspension* appear as the features most connected to the negative reviews, and Mary wants to investigate perceptions of these particular car features further.

For this task, document clustering by concept in Jigsaw is useful. Mary switches back to the Document Grid View and sorts the reviews into ten clusters where document similarity is calculated by Jigsaw based on the set of entities connected to each review. The clusters are labeled with descriptive keywords and the documents within each cluster are ordered and colored by their sentiment (Figure 11, left). The majority of the negative reviews aggregate into clusters 1 and 8 described

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. X, NO. Y
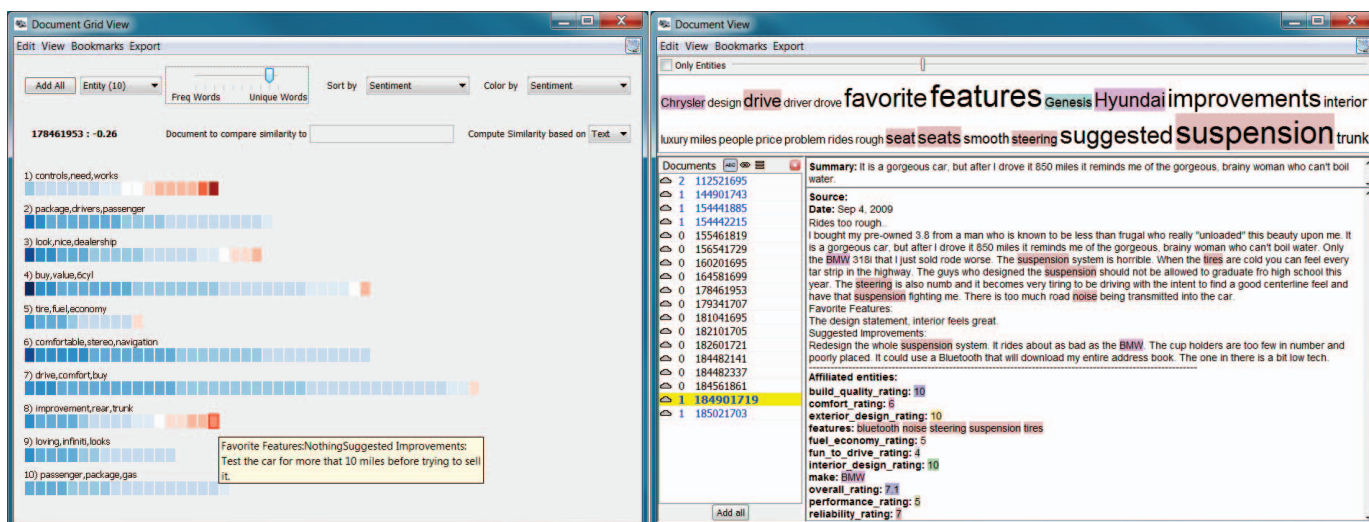
11

Fig. 11. Document Grid View (left) with the reviews grouped by similarity and ordered and colored by sentiment. Clusters 1 and 8 have the most negative sentiment. Document View (right) with the reviews from cluster 8 loaded. The word "suspension" is noteworthy within the word cloud at the top. The selected document illustrates an example of the views from reviews in this cluster.
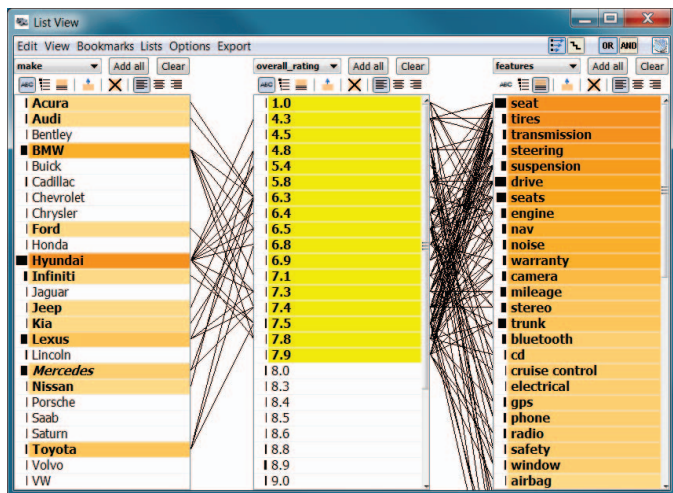

Fig. 10. List View showing the make, overall rating, and feature entity types. Low-rated reviews from 1.0 to 7.9 are selected and the feature list is sorted by connection strength to these selections.

by the terms "controls, needs, works" and "improvement, rear, trunk", respectively. It is not clear what each of these clusters is describing, so Mary loads the documents from each into a separate Document View to learn more.

The word clouds from each view highlight the most common words found in each review. The terms "suggested improvements" and "favorite features" are found in every review, so they are expectedly large. Similarly, the words "Hyundai" and "Genesis" also are common. However, the first cluster's word cloud also shows the word "transmission" in a large size, as does the second cloud for the word "suspension" (see Figure 11, right). This observation and the earlier similar finding from the List View suggests these may be key problems with the car. Mary decides to investigate further and reads all the reviews in cluster 8. She finds that the suspension is often described in a negative context, as shown in the review in

Figure 11, right. She concludes that the suspension may indeed be a weak point of the 2009 Hyundai Genesis. Even though Jigsaw only performs document level sentiment analysis, Mary was able to also determine a type of feature-level sentiment analysis by combining the results of multiple computational analyses and coordinating their results across different visual representations of the document collection.

Mary now recalls that far more reviews were positive than negative, so she decides to examine the good aspects of the car. She selects all of the reviews giving the car a perfect overall rating of 10.0 in the List View (48 reviews in total, shown in Figure 12). The features *drive*, *seat(s)*, *stereo*, *fuel*, and *navigation* show up as being most connected. The terms *drive* and *seat(s)* occur in many documents overall as indicated by the long bar in front of the terms in the List View, so they may not be as useful. Mary now loads the documents mentioning *stereo*, the next highest term, into a Document View and reads these reviews. She learns that the Genesis' sound system is a 17-speaker Lexicon system and the reviewers typically rave about it, a definite plus to her.

Mary also wants to learn what are the other top, competitive brands of cars to consider as alternatives. She is curious about reviews mentioning other makes of cars. Thus, she sorts the car make entity by frequency in the List View and selects the top four other mentioned makes (all luxury cars), *Lexus*, *BMW*, *Mercedes*, and *Infiniti*, one by one. She notices that, overall, the connected reviews for each receive high ratings, suggesting that the Genesis is being compared favorably with these other makes. The reviews mentioning *BMW* exhibit slightly lower overall ratings, however. Perhaps prior BMW owners are not quite as favorably impressed as owners of the three other car brands. She reads the reviews also mentioning BMW and confirms that this is true.

To learn more about the ride quality of the car, an important feature to her, Mary displays Jigsaw's Word Tree View for "ride" (Figure 13). A Word Tree [64] shows all occurrences
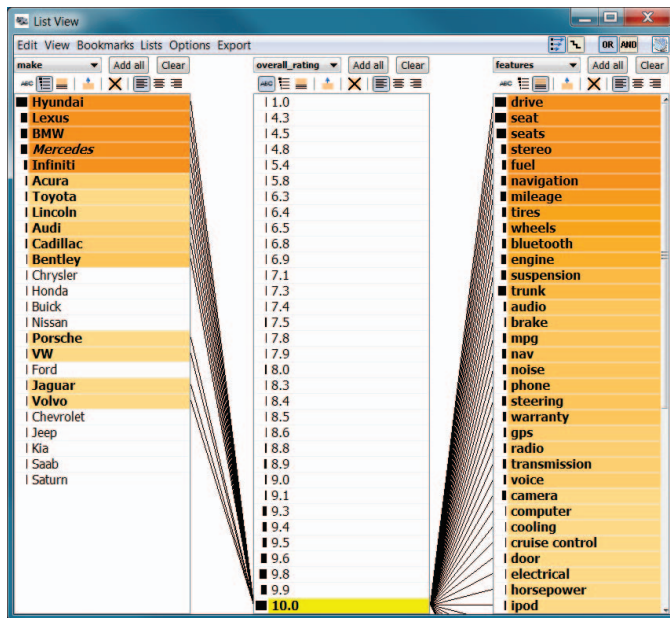
Fig. 12. List View showing the make, overall rating, and feature entity types. All the reviews with an overall rating of 10.0 are selected. The make list is sorted by overall frequency within the document collection and the feature list is sorted by connection strength to the 10.0 overall ratings.

Fig. 13. Word Tree View showing occurrences of the word "ride" and the most common phrases that follow the word in sentences within the review collection.

of a word or phrase from the reviews in the context of the words that follow it, each of which can be explored further by a click. The Word Tree View shows that reviewers have different opinions about the quality of the ride, ranging from "a little bumpy" and "rough and jittery" to "comfortable and quiet" and "excellent".

Mary's investigations of the Genesis' reviews have helped her understand overall perceptions of the car and what the most recent impressions are. The computational analyses, the sentiment analysis and document clustering in particular, facilitated the identification of the car features perceived most favorably and unfavorably by the reviewers, and Mary learned more about other competitive makes and models of cars. An important part of such an exploration is reading the individual reviews of note, which we have not emphasized here for obvious reasons of brevity. However, we must stress that this activity is a key aspect of any document corpus investigation like this. The newly integrated computational analyses in Jigsaw help to more rapidly identify the documents of note for any of a variety of attributes or dimensions.

5 COMPUTATIONAL ANALYSIS ALGORITHMS

In this section we provide a brief discussion of the text analysis algorithms we implement in Jigsaw, primarily for the reader interested in more detail. We integrate well-known algorithms for the different computational analyses, practical algorithms that can be readily implemented in Java (Jigsaw's implementation language) and that run in a "reasonable" time on computers that real clients would have. These descriptions and our experiences in designing and implementing the capabilities may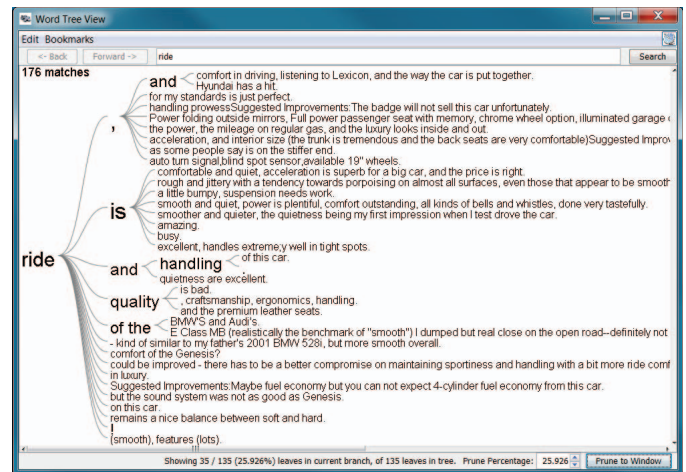 be beneficial for other researchers who wish to integrate enhanced automated computational analysis in their visual analytics systems.

5.1 Preprocessing

To apply computational analyses, text documents are typically converted to a certain form of numerical vector representation. We use the standard "bag of words" encoding scheme where each dimension corresponds to a unique term, and the value represents the term count in the document. In Jigsaw, the vocabulary that constitutes the entire set of dimensions can be based on either all the terms occurring in the document corpus or only the entities that are identified within the documents. Thus, we obtain either a term-document or an entity-document matrix.

Then, we follow standard preprocessing procedures for text data such as stemming and stop word removal. For stemming, we use the Porter Stemmer [51] implementation in the Lingpipe library. Additionally, we exclude the terms and entities that appear less than three times throughout the entire document set. (The terms and entities are only excluded from the computational analyses; they are not removed from the dataset.) Based on empirical experiments we determined that these terms do not affect the results of the computational modules significantly while the vocabulary size is reduced drastically, often up to 40%, which improves both the computation time and memory usage.

After building the term-document matrix, we apply TF-IDF weighting and normalization [1]. TF-IDF weighting penalizes the terms that broadly appear in many documents since they would not contribute to the differentiation of one document from another. Normalization transforms each document vector to a unit norm to overcome the dependency on the document length.

Based on this numerical encoding of textual documents, we integrate three text analytical modules into Jigsaw: document summarization, document similarity, and document clustering. Document sentiment analysis, our fourth module, operates directly on the original document text.

5.2 Document Summarization

This module summarizes documents by extracting significant sentences. It first computes the importance scores (described below) for all the terms and for all the sentences within a single document, and then ranks the sentences with respect to the scores. The sentence with the highest importance score is determined to be the most representative sentence in the document and chosen as a summary sentence. The scored and ranked terms are used to summarize multiple documents with keywords (described in Section 5.4). To determine the sentences and the terms in a document we use a sentence splitter and a tokenizer from the Lingpipe library.

To implement the summarization algorithm, we apply the mutual reinforcement learning method [68]. This method first decomposes each document into a set of all the terms $T = \{t_1, \ldots, t_m\}$ and a set of all the sentences $S = \{s_1, \ldots, s_n\}$. A weighted bipartite graph between T and S is built with a weight matrix $W = \{w_{ij}\} \in \mathbb{R}^{m \times n}$ where w_{ij} is the frequency of the term t_i in the sentence s_j. Then we randomly initialize two vectors, $u \in \mathbb{R}^{m \times 1}$ and $v \in \mathbb{R}^{n \times 1}$, of the importance scores of terms and sentences, and perform a power iteration, i.e., $u = Wv$ and then $v = W^T u$, normalizing after every step. This iteration continuously passes the importance scores between terms and sentences until they converge.

5.3 Document Similarity

This module computes all the pairwise similarity scores for the documents in the corpus. The computation of similarity between two documents can be based on various measures. Although the most widely used measure is the Euclidean distance, semantically, cosine similarity can be a better choice for textual data [56] and therefore we use it in our implementation.

To obtain semantically better results, we do not compute the similarity based on the original document vector. Instead, we first reduce its dimension by applying the latent semantic analysis (LSA) technique [17] and then compute the similarity in the resulting reduced dimensional space. By grouping semantically similar terms, LSA improves similarity scores against polysemy and synonymy problems. After experimenting with different values, we chose to set the number of reduced dimensions to 20% of the number of dimensions after the preprocessing step of removing terms that occur in less than three documents. LSA requires the computation of the singular value decomposition (SVD) of the term-document matrix. We use the JAMA library[4] for matrix computations such as SVD.

Using the term-document or the entity-document matrix we can compute document similarity based on either the entire document text or on the entities identified in the documents.

5.4 Document Clustering

This module groups the documents into a given number of clusters, where similar documents fall into the same cluster. The similarity can be based on the document text or on the

4. http://math.nist.gov/javanumerics/jama

entities identified in the documents. We adopt the spherical k-means clustering algorithm [18], which uses cosine similarity as a distance measure.

The clustering algorithm requires the number of clusters as an input parameter. Theoretically, it is crucial to choose the "right" number of clusters to get optimal clustering results. Although there exist methods to quantitatively evaluate clusters [29], semantically it is difficult to determine the right number of clusters and achieve satisfactory results on noisy real world data. Thus, by default, we choose 20 as the number of clusters. Our reasoning behind this choice is that, on the one hand, if the document set has fewer clusters, then our result would show a few similar clusters that can be merged into a single true cluster by humans' further analysis. On the other hand, if the document set has significantly more clusters, e.g., 50 clusters, analysts might have difficulties in understanding their structure due to an unmanageable number of clusters, even if they represent the correct clustering result. However, we also provide a user option to specify the number of clusters in case an analyst is familiar with a specific document set and has some knowledge about its structure.

In addition, the algorithm requires a list of initial seed documents for the clusters as an input parameter. Although the algorithm is not sensitive to this parameter if the document set has a clearly clustered structure, we observed that the results can vary significantly depending on the initial seeds for most real-world document sets that do not have well-defined clusters. Thus, we carefully choose initial seed documents using a heuristic in which seed documents are recursively selected such that each seed document is the least similar document to the previously selected seed document. Due to space limitation we do not discuss details, such as optimizations and exceptions of this heuristic. We also provide a user option to choose initial seed documents in case an analyst is interested in specific topics in a document set and wants to steer the cluster analysis by choosing the seed documents according to the topics of interest.

To enhance the usability of the clustering results, we summarize the content of each cluster as a list of the most representative terms within its documents. We use the algorithm described in Section 5.2 after aggregating all documents in a cluster into one single document. However, instead of the most representative sentence, we use three high-ranking terms as the summary of the cluster. We compute a number of alternative term summaries for each cluster. One summary is based only on the term frequency within a cluster, whereas another summary also takes term uniqueness across clusters into account and eliminates any terms that would occur in multiple summaries; additionally we compute summaries that are gradually more strict on the uniqueness of summary terms (i.e., eliminating any terms that occur in 10%, or 20%, ..., or 90% of the cluster summaries). As described in Section 3 the analyst can interactively switch between the different cluster summaries to gain different perspectives on the clustering result, either examining the content of individual clusters or understanding differences among the clusters.

5.5 Document Sentiment Analysis

This module provides two different implementations to characterize the text in a document on a positive-to-negative scale. It does not apply the preprocessing steps discussed in Section 5.1 but operates directly on the original document text.

One implementation is based on the classifier provided in the Lingpipe library. It applies the hierarchical classification technique described in Pang and Lee [48] and requires two classifiers: one for subjective/objective sentence classification and one for polarity classification. The technique involves running the subjectivity classifier on the document text to extract the subjective sentences first and then running the polarity classifier on the result to classify the document as positive or negative. We trained the subjectivity classifier with the data provided in [48]. To train the polarity classifier we used 2,000 product reviews (1,000 positive and 1,000 negative) extracted from amazon.com. We considered all reviews with a rating of 4 or 5 as positive and those with a rating of 1 or 2 as negative. We did not use reviews with a rating of 3.

An alternative implementation computes a quantitative sentiment score for each document on a scale from +1 (positive) to -1 (negative) via a dictionary-based approach, identifying "positive" and "negative" words in documents. We developed the list of words by creating two initial sets of negative and positive words and then iterating and checking results against known positive and negative documents. The results have been surprisingly good, particularly when characterizing documents strong in expected sentiment such as product reviews. We also allow the user to provide domain-specific dictionaries of positive and negative words to classify the documents. This feature was very useful for a collaborative analysis in which we examined wine reviews with a wine expert; we developed dictionaries of words that describe "good" and "bad" wines to classify the reviews.

5.6 Computation Time

The runtime of the computational analyses depends on the characteristics of the document collection (number of documents, average document length, and number of entities per document) and the available computational power (processor speed and memory size). The InfoVis and VAST papers dataset in our case study has 578 documents, the average length of a paper's title and abstract is 1,104 characters (min: 159; max: 2,650), and the average number of entities per paper is 17 (min: 4; max: 58). On a desktop computer with 8 GB of memory and two 2.4 GHz Quad-Core processors, computing the summary sentences and the sentiment analysis each took 2 seconds, the text-based similarity computation took 47 seconds, the entity-based similarity computation took 10 seconds, the text-based cluster computation took 20 seconds, and the entity-based cluster computation took 15 seconds, resulting in a total computation time of less than 2 minutes. The car reviews dataset is much smaller (231 documents) and all analyses finished in 12 seconds.

We also ran the analyses on other datasets with different characteristics. From a practical point of view, the computation time can be divided into three categories. For small datasets (about 500 documents) the computation time is a few minutes (coffee break), for medium-sized datasets (about 1,500 documents) the computation time is less than one hour (lunch break), and for larger datasets (about 5,000 documents) the computation time is several hours (over-night).

6 DISCUSSION

Investigations on document collections proceed with the analyst gathering nuggets of information while forming new insights and deeper understanding of the document contents. Especially when the documents are unfamiliar, an investigator may not know where to start, what is related, or how to dive more deeply into analysis. We believe that fluid integration of computational analyses with interactive visualization provides a flexible environment that scaffolds the investigator's exploratory process.

In exploration and sensemaking, investigators likely want to ask a broad set of questions and also develop new questions throughout the investigation process. Interactive visualization supports this dynamic conversation or dialog between the investigator and the data, and it makes the results of powerful computational analyses more easily accessible and contextually relevant. Our efforts to integrate enhanced computational analysis support into Jigsaw have taught us a number of lessons about this process (resulting both from an implementation perspective and from working with users of the system [7, 35]), but five in particular stand out:

1. Make different computational analysis results available throughout the system in a variety of different contexts and views, not in just one canonical representation. Chuang et al. [11] identify *interpretation* and *trust* as two key issues to the success of visual analytics systems. With respect to the results of computational text mining, trust seems to be a primary concern. We have found that portraying the results of mining algorithms under different perspectives better allows the analyst to inspect and interpret the algorithm's results. In particular, multiple analyses within Jigsaw appear in several different views and can be examined under different perspectives. For example, the single sentence document summaries are shown above the corresponding full document text in the Document View as one might expect, but they also are available as tooltips anywhere a document is represented iconically or by name. Clusterings are shown (naturally) in the Document Cluster View but also in the Document Grid View that simultaneously can show similarity, sentiment, and summary analysis results. Furthermore, clusters are easy to select and thus inspect the member documents under other analysis perspectives and views. Given any set of documents resulting from a text analysis, one simple command allows those documents to be loaded into a Document View for further manual exploration.

2. Flexibly allow analysis output also to be used as input. Investigators using Jigsaw can select individual documents from any analysis view and can then request to see that document's text or see related documents. Jigsaw presents the results of similarity, clustering, and sentiment analyses visually

(output), but such results can be clicked on or selected by the analyst (input) to drive further exploration. This capability is pervasive throughout the system—any document or entity can be acted upon to drive further investigation. We believe this design, which helps to facilitate the core, iterative sensemaking cycle of visual analytics [36], enables smoother, more flexible interaction with the system, ultimately leading to deeper inquiry, exploration, and increased knowledge.

3. Integrate different, independent computational analysis measures through interactive visualization in order to extend functionality and power. A deep integration of automated analysis with interactive visualization results in capabilities beyond each of the two components ("the whole is greater than the sum of the parts"). For example, Jigsaw provides document-level sentiment analysis, but does not analytically provide sentiment with respect to specific terms, concepts, or features within a document. However, as illustrated in the car review scenario, by first performing content-based clustering that divides the car reviews into sets of documents discussing different car features, and then visualizing the sentiment on the resulting clusters, one achieves a type of feature-based sentiment. The scenario showed how the reviewers felt negatively about the car's suspension and transmission.

4. Provide computational support for both analysis directions: narrowing down as well as widening the scope of an investigation. Many investigations take the form of an hourglass: an analyst first confronts a large amount of data (top of the hourglass), iteratively filters and searches the data to discover a small number of interesting leads (middle of the hourglass), and then expands the data under investigation again by following connections from those identified leads (bottom of the hourglass). These new data points then represent the top of another hourglass and the analyst repeats the process. Cutting et al. [16] describe this narrowing-widening, iterative process as the Scatter/Gather method. To smoothly move through the different stages of an hourglass investigation, a visual analytics system should provide support for narrowing down as well as widening the scope of analysis. Jigsaw provides a variety of analysis support for both tasks. Document clustering and sentiment analysis help narrow down the scope by limiting it to one (or a few) clusters or taking only positive or negative documents into account; document similarity and recommending related entities help widen the scope by suggesting additional relevant documents; and identified entities help with both directions: they can be used to determine a germane subset of a document set (containing one or more identified entities) or to suggest other related documents (containing the entity of interest).

5. Expose algorithm parameters in an interactive user-accessible way. The effectiveness of many computational analyses depends on the choices of their parameters. Whenever possible, visual analytics tools should provide users intuitive access to the parameter space of the underlying analyses. In Jigsaw, we expose parameters in a number of different ways. For the k-means clustering, we expose the corresponding

parameters directly since they are quite intuitive. Users can either choose default values or define the number of clusters, specify whether the clustering should be based on the document text or only the entities connected to a document, and provide initial seed documents for the clusters. They then can display different clusterings from different parameter choices in multiple Document Cluster Views to compare and contrast them. We take a different approach for the cluster summarization algorithm. Instead of exposing the summarization parameters directly, we precompute a set of summarizations and let users explore the parameter space by selecting cluster summaries via an interactive slider (based on uniqueness vs. frequency of the summary words). Users have preferred this approach more than exposing the (not so intuitive) parameters of the summarization algorithm directly. For the dictionary-based entity identification and the sentiment analysis, users can provide their own domain-specific dictionaries. This flexibility has proven to be very useful in various domain-specific investigations that we and others have conducted with Jigsaw (e.g., investigating wine reviews, car reviews, scientific papers, and Java code). User requests for exposing additional algorithm parameters, such as regular expressions for the rule-based entity identification approach, confirm the importance of this lesson.

7 CONCLUSION

Helping investigators to explore a document collection is more than just retrieving the "right" set of documents. In fact, all the documents retrieved or examined may be important, and so the challenge becomes how to give the analyst fast and yet deep understanding of the contents of those documents.

In this article we have illustrated methods for integrating automated computational analysis with interactive visualization for text- and document-based investigative analysis. We implemented a suite of analysis operations into the Jigsaw system, demonstrating how to combine analysis results with interactive visualizations to provide a fluid, powerful exploration environment. Further, we provided two example sensemaking scenarios that show both the methodologies and the utility of these new capabilities. We included brief descriptions of the computational analysis algorithms we chose to help readers seeking to implement similar operations in their systems. Finally, we described our experiences in building the new system and the lessons we learned in doing so.

The contributions of the work are thus:

- Techniques for integrating computational analysis capabilities fluidly with different interactive visualizations, and realization of those techniques in the Jigsaw system.
- Illustrations of the benefits of this approach via two example sensemaking scenarios. These scenarios provide sample questions and tasks, methods to resolve them, and the analysis and insights that result.
- Guidance for HCI/visualization researchers about the implementation of practical, text-focused computational analysis algorithms.
- Design principles for the construction of future document analysis and visual analytics systems.

A particular strength of Jigsaw is its generality for analyses on different types of documents. Many other systems have been tailored to a specific style of document or content domain and thus provide sophisticated capabilities only in that area. Jigsaw has been applied in the domains here (academic research and consumer product reviews) and in other diverse areas such as aviation documents [49], understanding source code files for software analysis and engineering [53], genomics research based on PubMed articles [25], and investigations in fraud, law enforcement, and intelligence analysis [35]. The system is available for download.[5]

Many avenues remain for future research. We admittedly have not conducted formal evaluations or user studies of these new capabilities within Jigsaw. Determining the best methods to evaluate systems like this is a research challenge unto itself. Our earlier user study involving Jigsaw [34] identified the potential benefits of the system, so we believe that the addition of the new computational analysis capabilities will provide even further value. In particular, the new capabilities address analysis needs identified in the user study and determined through earlier trial use of the system by clients.

We also plan to explore newer, more powerful methods and algorithms for calculating analysis metrics. The areas of computational linguistics, dimensionality reduction, and text mining are ripe with analysis methods such as topic modeling [6] and multi-word expressions [5] that could be integrated into Jigsaw. Furthermore, allowing user-driven interactive feedback to modify and evolve the computational analyses would provide an even more flexible exploration environment.

Finally, we made a claim that to achieve its fullest potential within visual analytics, a system must deeply and seamlessly combine automated computational analysis with interactive visualization. Actually, according to the definition of visual analytics introduced in *Illuminating the Path* [58], we omitted the third key piece of the equation: integrated support for analytical reasoning. Systems such as Jigsaw seeking to provide comprehensive analytic value also should include facilities for supporting human investigators' analytic reasoning processes and goals.

We are encouraged that the vision of visual analytics is beginning to be realized. The system and experiences described in this paper illustrate the potential of such an approach: fluidly integrating computational data analysis algorithms with flexible, interactive visualizations provide investigators with powerful data exploration capabilities and systems.

ACKNOWLEDGMENTS

5. http://www.cc.gatech.edu/gvu/ii/jigsaw

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition)*.   ACM Press, 2011.

[2] C. Bartneck and J. Hu, "Scientometric analysis of the CHI proceedings," in *ACM CHI*, 2009, pp. 699–708.

[3] M. Berry and M. Castellanos, *Survey of Text Mining II: Clustering, Classification, and Retrieval*.   Springer, 2008, vol. XVI.

[4] E. A. Bier, S. K. Card, and J. W. Bodnar, "Principles and tools for collaborative entity-based intelligence analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 178–191, 2010.

[5] D. Blei and J. Lafferty, "Visualizing topics with multi-word expressions," arXiv:0907.1013v1, Tech. Rep., 2009.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[7] E. Braunstein, C. Görg, Z. Liu, and J. Stasko, "Jigsaw to Save Vastopolis - VAST 2011 Mini Challenge 3 Award: "Good Use of the Analytic Process"," in *IEEE VAST*, Oct 2011, pp. 323–324.

[8] N. Cao, J. Sun, Y.-R. Lin, D. Gotz, S. Liu, and H. Qu, "FacetAtlas: Multifaceted visualization for rich text corpora," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1172–1181, 2010.

[9] A. J. B. Chaney and D. M. Blei, "Visualizing topic models," in *AAAI ICWSM*, 2012, pp. 419–422.

[10] J.-K. Chou and C.-K. Yang, "PaperVis: Literature review made easy," *Computer Graphics Forum*, vol. 30, no. 3, pp. 721–730, 2011.

[11] J. Chuang, D. Ramage, C. D. Manning, and J. Heer, "Interpretation and trust: designing model-driven visualizations for text analysis," in *ACM CHI*, 2012, pp. 443–452.

[12] C. Collins, S. Carpendale, and G. Penn, "DocuBurst: Visualizing document content using language structure," *Computer Graphics Forum*, vol. 28, no. 3, pp. 1039–1046, 2008.

[13] C. Collins, F. B. Viegas, and M. Wattenberg, "Parallel tag clouds to explore and analyze faceted text corpora," in *IEEE VAST*, Oct. 2009, pp. 91–98.

[14] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. J. Gao, X. Tong, and H. Qu, "TextFlow: Towards better understanding of evolving topics in text," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2412–2421, 2011.

[15] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, N. Aswani, I. Roberts, G. Gorrell, A. Funk, A. Roberts, D. Damljanovic, T. Heitz, M. A. Greenwood, H. Saggion, J. Petrak, Y. Li, and W. Peters, *Text Processing with GATE (Version 6)*, 2011.

[16] D. R. Cutting, D. R. Karger, J. O. Pedersen, and J. W. Tukey, "Scatter/gather: a cluster-based approach to browsing large document collections," in *ACM SIGIR*, 1992, pp. 318–329.

[17] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. of the Society for information Science*, vol. 41, pp. 391–407, 1990.

[18] I. S. Dhillon and D. S. Modha, "Concept decompositions for large sparse text data using clustering," *Mach. Learn.*, vol. 42, no. 1/2, pp. 143–175, 2001.

[19] A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman, and C. Plaisant, "Discovering interesting usage patterns in text collections: integrating text mining with visualization," in *ACM CIKM*, 2007, pp. 213–222.

[20] W. Dou, X. Wang, R. Chang, and W. Ribarsky, "Parallel topics: A probabilistic approach to exploring document collections," in *IEEE VAST*, Oct 2011, pp. 229–238.

[21] S. G. Eick, "Graphically displaying text," *Journal of Computational and Graphical Statistics*, vol. 3, no. 2, pp. 127–142, 1994.

[22] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*.   Cambridge University Press, 2007.

[23] M. J. Gardner, J. Lutes, J. Lund, J. Hansen, D. Walker, E. Ringger, and K. Seppi, "The topic browser: An interactive tool for browsing topic models," in *NIPS Workshop on Challenges of Data Visualization*, 2010.

[24] C. Görg, Z. Liu, N. Parekh, K. Singhal, and J. Stasko, "Jigsaw meets Blue Iguanodon – The VAST 2007 Contest," in *IEEE VAST*, Oct. 2007, pp. 235–236.

[25] C. Görg, H. Tipney, K. Verspoor, W. Baumgartner, K. Cohen, J. Stasko, and L. Hunter, "Visualization and language processing for supporting analysis across the biomedical literature," in *Knowledge-Based and Intelligent Information and Engineering Systems*. LNCS Springer, 2010, vol. 6279, pp. 420–429.

[26] M. Gregory, N. Chinchor, P. Whitney, R. Carter, E. Hetzler, and A. Turner, "User-directed sentiment analysis: Visualizing the affective content of documents," in *Workshop on Sentiment and Subjectivity in Text*, 2006, pp. 23–30.

[27] B. Gretarsson, J. O'Donovan, S. Bostandjiev, T. Höllerer, A. U. Asuncion, D. Newman, and P. Smyth, "Topicnets: Visual analysis of large text corpora with topic modeling," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 2, pp. 23:1–23:26, 2012.

[28] S. Havre, B. Hetzler, and L. Nowell, "ThemeRiver: Visualizing theme changes over time," in *IEEE InfoVis*, Oct 2000, pp. 115–123.

[29] J. He, A.-H. Tan, C. L. Tan, and S. Y. Sung, "On quantitative evaluation of clustering systems," in *Clustering and Information Retrieval*, 2003, pp. 105–134.

[30] E. Hetzler and A. Turner, "Analysis experiences using information visualization," *IEEE Computer Graphics and Applications*, vol. 24, no. 5, pp. 22–26, 2004.

[31] "i2 - Analyst's Notebook," http://www.i2inc.com/.

[32] D. Jonker, W. Wright, D. Schroh, P. Proulx, and B. Cort, "Information triage with TRIST," in *International Conference on Intelligence Analysis*, May 2005.

[33] H. Kang, C. Plaisant, B. Lee, and B. B. Bederson, "NetLens: iterative exploration of content-actor network data," *Information Visualization*, vol. 6, no. 1, pp. 18–31, 2007.

[34] Y.-a. Kang, C. Görg, and J. Stasko, "How can visual analytics assist investigative analysis? Design implications from an evaluation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 5, pp. 570–583, May 2011.

[35] Y.-a. Kang and J. Stasko, "Examining the use of a visual analytics system for sensemaking tasks: Case studies with domain experts," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2869–2878, 2012.

[36] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," *Information Visualization: Human-Centered Issues and Perspectives*, pp. 154–175, 2008.

[37] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, Eds., *Mastering the Information Age – Solving Problems with Visual Analytics*. Eurographics Assoication, 2010.

[38] D. A. Keim and D. Oelke, "Literature fingerprinting: A new method for visual literary analysis," in *IEEE VAST*, 2007, pp. 115–122.

[39] G. Klein, B. Moon, and R. Hoffman, "Making sense of sensemaking 1: alternative perspectives," *IEEE Intelligent Systems*, vol. 21, pp. 70–73, 2006.

[40] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson, "Understanding research trends in conferences using PaperLens," in *ACM CHI: Extended Abstracts*, 2005, pp. 1969–1972.

[41] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian, "Tiara: Interactive, topic-based visual text summarization and analysis," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 2, pp. 25:1–25:28, Feb. 2012.

[42] Z. Liu, C. Görg, J. Kihm, H. Lee, J. Choo, H. Park, and J. Stasko, "Data ingestion and evidence marshalling in Jigsaw," in *IEEE VAST*, Oct. 2010, pp. 271–272.

[43] G. Marchionini, "Exploratory search: From finding to understanding," *Communications of the ACM*, vol. 49, no. 4, pp. 41–46, Apr. 2006.

[44] G. Marchionini and R. W. White, "Information-seeking support systems," *IEEE Computer*, vol. 42, no. 3, pp. 30–32, Mar. 2009.

[45] D. Oelke, P. Bak, D. Keim, M. Last, and G. Danon, "Visual evaluation of text features for document summarization and analysis," in *IEEE VAST*, Oct. 2008, pp. 75–82.

[46] D. Oelke, M. Hao, C. Rohrdantz, D. Keim, U. Dayal, L.-E. Haug, and H. Janetzko, "Visual opinion analysis of customer feedback data," in *IEEE VAST*, Oct. 2009, pp. 187–194.

[47] W. B. Paley, "TextArc: Showing word frequency and distribution in text," in *IEEE INFOVIS (Poster)*, 2002.

[48] B. Pang and L. Lee, "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of Association for Computational Linguistics*, 2004, pp. 271–278.

[49] O. J. Pinon, D. N. Mavris, and E. Garcia, "Harmonizing European and American aviation modernization efforts through visual analytics," *Journal of Aircraft*, vol. 48, pp. 1482–1494, Sept-Oct 2011.

[50] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," in *International Conference on Intelligence Analysis*, May 2005.

[51] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.

[52] L. Ratinov and D. Roth, "Design challenges and misconceptions in named entity recognition," in *CoNLL*, 2009, pp. 147–155.

[53] H. Ruan, C. Anslow, S. Marshall, and J. Noble, "Exploring the inventor's paradox: applying Jigsaw to software visualization," in *ACM SOFTVIS*, Oct 2010, pp. 83–92.

[54] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card, "The cost structure of sensemaking," in *ACM CHI*, 1993, pp. 269–276.

[55] J. Stasko, C. Görg, and Z. Liu, "Jigsaw: supporting investigative analysis through interactive visualization," *Information Visualization*, vol. 7, no. 2, pp. 118–132, 2008.

[56] A. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on web-page clustering," in *Workshop on Artificial Intelligence for Web Search (AAAI)*, 2000, pp. 58–64.

[57] V. Thai, P.-Y. Rouille, and S. Handschuh, "Visual abstraction and ordering in faceted browsing of text collections," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 2, pp. 21:1–21:24, Feb. 2012.

[58] J. J. Thomas and K. A. Cook, *Illuminating the Path*. IEEE Computer Society, 2005.

[59] F. van Ham, M. Wattenberg, and F. B. Viégas, "Mapping text with phrase nets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1169–1176, 2009.

[60] F. B. Viégas, S. Golder, and J. Donath, "Visualizing email content: portraying relationships from conversational histories," in *ACM CHI*, 2006, pp. 979–988.

[61] F. B. Viégas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1137–1144, 2009.

[62] R. Vuillemot, T. Clement, C. Plaisant, and A. Kumar, "What's being said near "Martha"? Exploring name entities in literary text collections," in *IEEE VAST*, Oct. 2009, pp. 107–114.

[63] M. Wattenberg, "Arc diagrams: Visualizing structure in strings," in *IEEE INFOVIS*, 2002, pp. 110–116.

[64] M. Wattenberg and F. B. Viégas, "The Word Tree, an Interactive Visual Concordance," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1221–1228, 2008.

[65] C. Weaver, "Cross-filtered views for multidimensional visual analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, pp. 192–204, March 2010.

[66] R. W. White, B. Kules, S. M. Drucker, and M. C. Schraefel, "Supporting exploratory search," *Communications of the ACM*, vol. 49, no. 4, pp. 36–39, Apr. 2006.

[67] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort, "The Sandbox for analysis: Concepts and methods," in *ACM CHI*, April 2006, pp. 801–810.

[68] H. Zha, "Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering," in *ACM SIGIR*, 2002, pp. 113–120.

Carsten Görg is an Instructor in the Computational Bioscience Program in the University of Colorado Medical School. He received a PhD in computer science from Saarland University, Germany in 2005. His research interests include visual analytics and information visualization with a focus on designing, developing, and evaluating visual analytics tools to support the analysis of biological and biomedical datasets. Dr. Görg is a member of the IEEE Computer Society.

Zhicheng Liu received the BS degree in computer science from the National University of Singapore in 2006 and the PhD degree in human-centered computing from the Georgia Institute of Technology in 2012. He is currently a postdoctoral scholar at Stanford University. His current research interests include visualizing big data and developing novel interaction mechanisms in visual analysis.

Jaeyeon Kihm is a PhD student in Information Science at Cornell University. He received the MS degree from the Georgia Institute of Technology in 2011 and the BS degree from the Illinois Institute of Technology in 2009. He is currently developing an energy-efficient user interface system for mobile information appliances.

Haesun Park is currently a Professor in the School of Computational Science and Engineering and director of the NSF/DHS FODAVA-Lead (Foundations of Data and Visual Analytics) Center at the Georgia Institute of Technology. Dr. Park has published over 150 peer reviewed papers in the areas of numerical algorithms, data analysis, visual analytics, bioinformatics, and parallel computing. She has served on numerous editorial boards including IEEE Transactions on Pattern Analysis and Machine Intelligence, SIAM Journal on Matrix Analysis and Applications, SIAM Journal on Scientific Computing, and has served as a conference co-chair for the SIAM International Conference on Data Mining in 2008 and 2009.

Jaegul Choo received the BS degree in electrical engineering from Seoul National University, Seoul, Korea in 2001 and the MS degree in electrical engineering from the Georgia Institute of Technology in 2009, where he is currently a research scientist as well as a PhD candidate in computational science and engineering. His research interests include visualization, data mining, and machine learning with a particular focus on dimension reduction and clustering methods.

John Stasko is a Professor and Associate Chair of the School of Interactive Computing at the Georgia Institute of Technology. He received a PhD in computer science from Brown University in 1989. His research interests are in human-computer interaction with a specific focus on information visualization and visual analytics. Dr. Stasko is a member of the IEEE and is on the Steering Committee for the IEEE Information Visualization Conference.

Examining the Use of a Visual Analytics System for Sensemaking Tasks: Case Studies with Domain Experts

Youn-ah Kang and John Stasko, *Senior Member, IEEE*

Abstract—While the formal evaluation of systems in visual analytics is still relatively uncommon, particularly rare are case studies of prolonged system use by domain analysts working with their own data. Conducting case studies can be challenging, but it can be a particularly effective way to examine whether visual analytics systems are truly helping expert users to accomplish their goals. We studied the use of a visual analytics system for sensemaking tasks on documents by six analysts from a variety of domains. We describe their application of the system along with the benefits, issues, and problems that we uncovered. Findings from the studies identify features that visual analytics systems should emphasize as well as missing capabilities that should be addressed. These findings inform design implications for future systems.

Index Terms—Visual analytics, case study, qualitative evaluation.

✦

INTRODUCTION

As visual analytics researchers, we ultimately hope that our technologies are making an impact and helping people gain value from their information. To do so, we need to evaluate such systems from multiple perspectives and reflect upon the findings [34,35]. Evaluation, however, is still not very common, probably because it is quite challenging to do [27]. While usability testing and controlled experiments remain crucial in the evaluation of visualization systems [8], particularly rare are actual case studies of prolonged visual analytics system use by analysts working in their domain with their own data.

Case studies can provide valuable findings and insights for visual analytics researchers. By detailing the use of a system, case studies yield a description of how a tool was used and where the users had problems. Until their particular challenges are understood, it also remains difficult to know how a visual analytic system helps expert users attain their goals. These findings are difficult to achieve through controlled lab studies.

The importance of case studies in information visualization has already been emphasized. Shneiderman and Plaisant [38] encouraged information visualization researchers to study users doing their own work in the process of achieving their goals. Perer and Shneiderman [25] also recognized the limitations of traditional controlled experiments in examining the process of exploratory data analysis.

Conducting case studies is challenging, however. First of all, it can be difficult to recruit appropriate people who are willing and able to use a particular system for their task on a regular basis. Case studies also often involve issues in the reliability, validity, and generalizability of results although these issues can be mitigated by scaling up the number of users. Nevertheless, it seems valuable to study domain experts working on complex problems over long time periods and learn how they employ systems. This paper examines extended field use of Jigsaw [39,40], a visual analytics system for helping analysts who work with large collections of documents. It complements an earlier comparative lab study of Jigsaw [20, 21] as well as descriptions of Jigsaw's use in solving problems within the VAST Challenge [5,13,28]. In this paper, we profile six investigators who have been using Jigsaw in their own work, including three

intelligence analysts, two academic researchers, and one business analyst.

The goals of this research include the following:

- To evaluate whether Jigsaw is helping analysts with their tasks and problems
- To understand its applicability to different types of documents and analyses
- To identify particularly useful features and capabilities of the system as well as missing or problematic ones
- To reflect on usage to inform the design of next generation tools for investigative analysis.

1 RELATED WORK

Komlodi, Sears, and Stanziola conducted a literature survey of about fifty user studies of information visualization systems [23] and found four main thematic areas of focus: controlled experiments comparing design elements, usability evaluations, controlled experiments comparing two or more tools, and case studies of tool use in realistic settings. Plaisant [27] believes that case studies are the least common type of studies.. She argues that case studies describe how users do real tasks in their natural environment, demonstrating feasibility and in-context usefulness.

Shneiderman and Plaisant [38] also encouraged information visualization researchers to assess the efficacy of tools by documenting usage through ethnographically-oriented and longitudinal participant observation. They suggested a research method called Multi-dimensional In-depth Long-term Case (MILC) studies as a technique that seems well adapted to study creative activities that users engage in. Encouraging information visualization researchers to study users doing their own work in the process of achieving their goals, the paper lists lessons from ethnography methods used in HCI [16,17,30] including observations and interviews and suggests evaluation methodology guidelines for information visualization researchers.

Lam et al. provided an in-depth discussion of evaluation scenarios after conducting an extensive literature survey of information visualization publications [24]. While the scenarios they suggested distinguish different study goals and types of research questions, the authors recommend that case studies can be effectively used to study "if and how a visualization tool supports the generation of actionable and relevant knowledge in a domain."

Despite of the significance of case studies in the field, there exist few evaluation projects based on case studies in visual analytics.

Chin et al. [9] conducted an observational case study with professional intelligence analysts in which participants worked on real-world scenarios, either as an individual analyst or as an

- *Youn-ah Kang is with Google Inc., e-mail: kang.younah@gmail.com.*
- *John Stasko is with School of Interactive Computing & GVU Center, Georgia Institute of Technology, e-mail: stasko@cc.gatech.edu.*

investigative team. The researchers revealed various characteristics of the analytical processes of intelligence analysts, such as the investigative methodologies they apply, how they collect and triage information, and how they identify patterns and trends.

Gotz et al. [14] also recognized the lack of public studies examining analyst behavior and conducted a user study with a few analysts to explore the ways in which they gather and process information. Through interview, observations, and written notes by analysts, they report important factors surrounding analyst behavior in information gathering and results processing, such as how they keep record and what their investigative style is like.

Perer and Shneiderman [25] recognized the limitations of traditional controlled experiments in examining the process of exploratory data analysis and developed an evaluation methodology for studying the effectiveness of their system, SocialAction. Consisting of a long-term case study [38] and in-depth interviews, the evaluation confirmed the core value of SocialAction–integrating statistics with visualization–and further provided guidance for redesign of the tool.

Saraiya et al. conducted a longitudinal study with biologists and bioinformaticians using real-life microarray data [33], following their insight-based controlled study [32]. The goal of the study was to gain basic understanding into the visual analytic process such as how different visualization tools are used to gain insight into data, what process is followed by people to acquire needed insights, and how insight is synthesized over time. In the study, participants were requested to keep a diary of the analysis process, the insights gained, the visualization and interaction techniques that led to the insights, and the successes and frustrations they experienced with the software tools.

Seo et al. also employed a case study approach to evaluate the HCE 3.0 interface, using interviews and participatory observations [37]. Their goals were to understand users' difficulties in learning the rank-by-feature framework and the HCE 3.0 interface so that they could improve the interface and their training methods for novice users. Through three case studies over eight weeks, the study demonstrated benefits of HCE 3.0 for knowledge discovery in research-level tasks, but yielded less guidance about improvements for the system.

A study by Bier et al. [3] assessed the suitability of their Entity Workspace System [2] in the context of design guidelines for collaborative intelligence analysis. The researchers modified their system based on five design guidelines and evaluated the system in both a laboratory study with intelligence analysts and a field study with an analysis team. Relying on analysts' subjective feedback in conjunction with quantitative logging data, they confirmed the positive effects of the tool on collaboration and the usefulness of the design guidelines for collaborative analysis.

In our research, we focused on use of the Jigsaw system in particular. Other existing systems such as Analyst's Notebook [18], Compendium [11,36], Sandbox [42], Entity Workspace [2], and Analysts Workspace [1] seek to assist analysts with similar document-related, sensemaking tasks.

2   METHODOLOGY

Jigsaw is not publicly available in general, but we distribute the system upon request. Approximately 150 people from a variety of domains including academics, government, law enforcement, intelligence, reporting, and fraud investigation have downloaded the system. However, we believe far fewer have used it extensively. We selected six analysts that we knew were using the system based on questions that they had sent to us about it in email. We asked if they would agree to tell us about their use of the system, and all agreed to conduct an interview with us and share their experiences.

The professionals include three intelligence analysts, two academic researchers, and one business analyst. They sought out Jigsaw after facing challenges in their own work and have been using Jigsaw for a range of 2-14 months. We conducted semi-structured interviews with each; two interviews were conducted face-to-face, and the other four were conducted over the phone.

Each interview lasted for about 45-60 minutes. While each was a semi-structured interview, we had a set of planned questions to make sure to cover important topics such as (1) What kind of tasks, data, and documents they used Jigsaw for, (2) To what extent and how Jigsaw helped their work compared to existing ways and methods, (3) What features were most/least useful, and (4) What barriers they encountered while using Jigsaw and how the tool can be improved. Sample questions include the following:

- For which tasks have you used Jigsaw? What kinds of documents are involved?
- What is the main purpose of using Jigsaw in analyzing those documents? What do you want to accomplish?
- Before using Jigsaw, how did you perform the tasks? What are advantages and disadvantages of the method?
- How do you typically work with Jigsaw and the documents?
- Which features do you use most? How does each of those features assist your task?
- What barriers did you encounter while using Jigsaw? How did you address the problem?
- How your usage has changed/evolved over time?
- What kind of features do you want to see in Jigsaw in the future?

While we took some notes during the interview sessions, all conversations were audio-recorded for further analysis. We also collected several screenshots whenever possible. The interviews were transcribed and analyzed using a general qualitative analysis technique, borrowing techniques from grounded theory [7,41]. After skimming through the transcribed texts, we determined initial coding schemes (i.e., core themes). We then carefully examined the transcript to find data fitting each coding scheme. Since the interview guide already had core concepts and themes, we focused more on disentangling phenomena and relationships behind users' experience with the tool. Once we went through the first round of analysis, we modified the coding schemes and repeated the process again. During the analysis, we also exchanged emails with participants as pertinent follow-up questions arose. Analysis was conducted by the first author as a component of her doctoral thesis.

3   CASE STUDIES

Throughout these studies, we found that the professionals have unique goals and consequently, different use cases of Jigsaw. This section describes each individual's particular background, objectives, and how they used the system.

3.1   P1: Aerospace Engineering Researcher

P1 is an Aerospace Engineering researcher at our university working on aerospace systems design. She was examining two air traffic control-related initiatives–the Next Generation Air Transportation System (NextGen) by the United States and The Single European Sky ATM Research (SESAR) by the European Union (SESAR). The two programs consist of new concepts, capabilities, and implementation plans over the next decade, pursuing a more efficient air traffic management.

While the objectives of SESAR and NextGen are similar, a number of differences exist between the two initiatives. In her field, the need for harmonization between the two has been recognized, and she wanted to analyze to what extent the two initiatives are compatible with each other. Particularly, she wanted to compare similarities and differences between the two programs–if a concept or capability suggested in one program also appears in the other program, and if so, how each program describes the same concept. To do so, she needed to examine components, roadmaps, terminologies, and definitions in each program thoroughly. Each program has seven huge volumes of documentation, and each
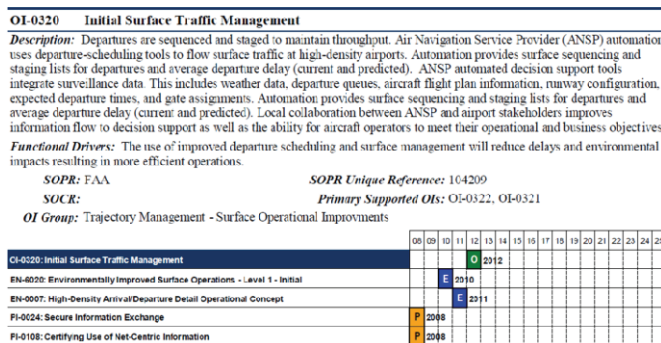
Fig 1. Original document: Operational Improvements-0320 in NextGen Initiative.
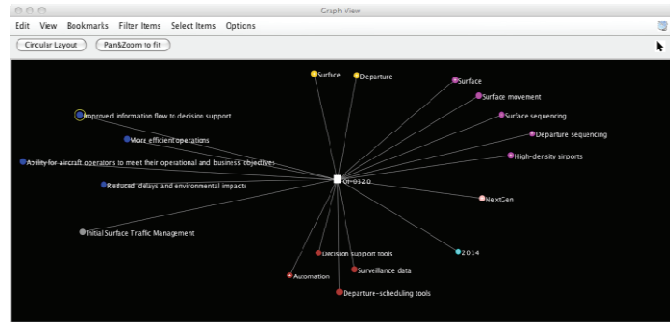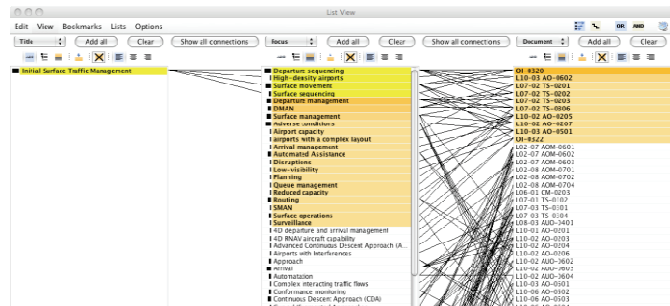

Fig 2. NextGen OI-320 is a node and its relevant "focus" entities such as "surface movement," "surface sequencing," "departure sequencing," and "high-density airports."

volume has dozens of documents in it. Each program has hundreds of Operational Improvements (OI) and enablers such as policy, technology, and procedures. Fig 1 illustrates one example of the many Operational Improvements in NextGen.

Her goal in this project was to create a mapping between the two programs by identifying similarities and relationships between operational improvements of each (Table 1).

Originally, she did the comparison manually using Microsoft Word and a search function. That is, she searched for descriptions of NextGen and identified keywords, and then she reviewed descriptions of SESAR containing matching keywords one by one, which was lengthy and cumbersome. Given the high number of descriptions and concepts, it became increasingly difficult to form a clear understanding of the underlying relationships and similarities between the two programs. At that point, she searched for a more analytically efficient way of reviewing the information and found Jigsaw. She had been using it for about 12 months when we interviewed her.

Table 1. Example of a Mapping between NextGen and SESAR Improvements

NextGen Operational Improvements	SESAR Operational Improvements and Steps
OI-0320: Initial Surface Traffic	L07-02 TS-0201: Basic Departure Management (DMAN)
	L07-02 TS-0202: Departure Management Synchronized with Pre-Departure Sequencing
	L07-02 TS-0203: Integration of Surface Management Constraint into Departure Management
	L07-02 TS-0306: Optimized Departure Management in the Queue Management Process
	L10-02 AO-0205: Automated Assistance to Controller for Surface Movement Planning and Routing
	L10-03 AO-0501: Improved Operations in Adverse Conditions through Airport Collaborative Decision Making
	L10-03 AO-0602: Collaborative Pre Departure Sequencing

In order to import documents into Jigsaw, she modified the original document (Fig 1) into text files so that it could be readable. She created entity types including 'Title," "Initial Operational Capability (IOC) indicators," "focuses," "benefits," as well as "the procedures, concepts and systems," relevant to each operational improvement.

She performed analysis mainly using the Graph View and the List View of Jigsaw. In the Graph view, she searched for any OI of interest and the document associated with the OI appeared as a node. She then further expanded the node to reveal the different entities relevant to the document of interest. After filtering out all but the "Focus" entities and expanding all nodes, all connections between relevant documents are represented. An example of a Graph View representation resulting from querying one of the NextGen OI is shown in Fig 2.

She used the List View to obtain similar connections, as illustrated in Fig 3. She set the first, second, and third columns to display the document's title, focus, and ID number, respectively. Then she selected the title corresponding to one of the NextGen OIs so that Jigsaw can provide a list of focuses associated with the

Fig 3. The Relationships between NextGen OI-0320 and relevant SESAR Operational Improvements.

document in the second column. By further selecting the different focuses of that particular OI, she could also see the ID number of all other relevant documents in SESAR. Following this process, she was able to map all NextGen OIs to focus-related SESAR OIs more efficiently, which would have been impossible with the manual approach. She pointed out that the complexity of each program's structure made it difficult to rigorously investigate and identify similarities and differences between both programs, but using visual analytics, she was able to review and analyze the information in an analytically efficient way. Her work using Jigsaw led to a publication in American Institute of Aeronautics and Astronautics [26]. After the research was done, she and her team continued to use Jigsaw for other research work.

In addition to current functionalities in Jigsaw, she wished compatibility with other query databases so that she could import documents directly from other databases and statistical capabilities that can count and measure connection strength.

3.2 P2: Business Analyst at Management Services

P2 is an analyst of an accounting firm in Malaysia. While his company provides a variety of services related to business management, he specializes in financial fraud and forensic investigation. He usually receives large amounts of both structured and unstructured data from his client. While his team has several tools that can effectively analyze structured data such as transactional data, they did not have an appropriate tool that can help analyze unstructured data such as emails or text files.

His main task is to examine unstructured data from financial databases of clients and to identify any linkages between people or companies relevant to financial fraud such as fictitious suppliers' invoices (i.e., bloated expenses to minimise tax), systematic deletion of suppliers' invoices, or fictitious customers' invoices to boost revenue. Before using Jigsaw, he would put all of the text documents into an Excel database, search for specific keywords within the database, and start investigation by reading all the returned documents containing that keyword. Obviously, this process required manpower to make the database and make it searchable with appropriate keywords.

He had been using Jigsaw for about 14 months. First he converts all documents into text files and imports them into Jigsaw. Then he identifies entities such as organizations, people, dates, locations, description, and zip codes. He starts an investigation with the Wordtree View, in which he searches for names of interest, for example "ABC" company, simply to see the context of the person or company. Next he examines the connections more carefully in the List View to observe what documents link the two people/companies together and who is connected most. Once he sees a potential connection between an entity and a company, he searches for the company and further investigates if other entities are linked to the company. He sometimes uses the Document Cluster or Timeline View to check the amount of documents within a certain topic or time frame.

Through this repetitive process, he can reveal connections between entities and use it as evidence for financial fraud. Thus, Jigsaw provides support for his task by making it easier to find linkages between entities in emails. In one case involving 4.5 million transactions, his team identified approximately 100,000 transactions as fictitious supplier invoices over a period of 10 years using data mining software. They suspected "John Doe" as being the prime culprit, but they needed evidence for that. They asked the HR personnel to seize his notebook and cloned his hard disk drive. After indexing all the documents on his notebook, they imported about 100,000 emails from the past 10 years into Jigsaw to find the motivation for the fraud. After analyzing the documents, they finally found that the theft of funds occurred because the suspect needed to support his children's education costs overseas.

Because his data comes as different formats such as pdfs, docs, and emails, he wanted to be able to import documents directly into Jigsaw instead of having to convert them to .txt format manually (Jigsaw's import of pdf and MS word files is sometimes problematic). Since he mainly looks for evidence, he also seeks the ability to statistically compute closeness or correlation between connections.

3.3 P3: PhD Candidate in Industrial and Systems Engineering

P3 is a PhD student at our university. Her research is about enterprise transformation, in which she tries build mathematical models of how firms can evolve over the years. In her previous research, she formulated mathematical models about company transformation, and now, she wants to validate the models by combining them with historical data of several companies. The company data, which includes 5,000+ company announcements and new articles of nine IT companies for 10 years, contains critical information about firms such as new product releases, executive/board changes, business expansion, strategic alliances, etc. By measuring how often those events have occurred in the past, she aims to combine the quantitative information with her model to see if the model is valid or not. That is, she is ultimately trying to transform qualitative information about the IT companies into a quantitative form that can be incorporated in her model. But she was in her initial stage of the research, and she first wanted to understand the documents and generate keywords based on the understanding for the next steps. After actively searching for software, she decided to use Jigsaw for her research and used it for 2 months.

Again, her goal in using Jigsaw is to obtain an overview of the huge document collection and extract keywords from those documents. Her documents were stored in Excel spreadsheets, which is an appropriate format for Jigsaw. She added entity types such as event type, company name, capitalIQ, and date, so that she could understand key events of each company. While she tried all the system views, she ended up using two views: The List View (most frequently) and the Calendar View when she was focusing on a specific time period, e.g., if something is occurring in a certain period.

While Jigsaw helps her research primarily at the initial stage, she thinks that it is very helpful in making sense of the documents in a relatively short amount of time. For a more detailed analysis and the ultimate output, she is using other software such as NorthernLight and statistical tools in conjunction with Jigsaw:

Jigsaw is ... for understanding. If I need to talk to my advisor about something, I'd go back to Jigsaw and import some documents, and then I can talk [about] what's really going on in this company. For more formats, I need to do statistical analysis, and I have to use other software that have better output format. For this particular project, I'm using more than 5 software [tools].

One difficulty she encountered was working with entities because the system did not identify those she really wanted. She did not find the people and organizations identified by the system very helpful and had to create her own lists. Because her purpose of using text analytics software was to finally create a statistical analysis instead of getting to know about the data in detail, she sought more functionality in terms of output such as a timeline table or word count results.

3.4 P4: Intelligence Analyst at a Police Department

P4 is an intelligence analyst at a police department in a city of close to 70,000 population. His work includes making sense of incident/crime reports everyday and discovering patterns, trends, and any top issues in the city. Particularly, he seeks to make better connections between individuals and other information collected in the incident reports. Because the amount of reports increases day by day, he has been trying to find ways to better analyze the narrative text data from the incident reports in their records management system.

Before using Jigsaw, he did not have any ways to systematically work with the information. Basically he could not do anything but read and remember. He read all the reports individually and tried to remember different connections between people, and then recognized names and locations that were outstanding. In order to know who is connected to whom, and in what documents, he printed a copy of the documents, put all the printed reports together, and tried to see the relationships.

When he discovered Jigsaw, he found it very helpful because he wanted to connect people, narrative text, subjects, and concepts in the same system. His goal in using Jigsaw was therefore to make sense of the crime reports and to find connections, patterns, trends, and associated names/places/other incidents. He had been using the system for 12 months when we spoke with him.

In order to work with Jigsaw, he reads in crime reports and puts them into an Excel spreadsheet, in which he adds labels for each column such as "Case number," "Name," "Person involved," "Incident address," "Home address," "Report date," and "Description." Then he imports the spreadsheet into Jigsaw and starts the investigation. The number of documents (e.g., the number of rows in the spreadsheet) ranges from 600 to 40,000 crime reports, depending on the time frame the investigator is interested in, for example, 1 year or 2-3 months of crime incidents. He mainly uses the List View, Graph View, and Document View. In the List view, he normally employs several lists such as persons, addresses and crime types, and conducts a search on a person and examines what addresses and crimes they are connected with. He sorts the lists by connection strength to get a quick sense of relationships between persons, addresses, and/or crimes. He also likes the visual aspect of Graph View in that he can look at connections through link analysis. He generally starts with one individual and then expands out from that person to see what documents, individuals, and addresses that person is connected to. While many times he starts with one person, after expansion of entities, he starts looking at other individuals and their relationships. In the Document View, he takes the information from the List View–such as a suspect or victim, selects the person, and then reads all the crime reports that person has been involved in and looks for any patterns or trends related to that person, involving crimes. Sometimes he uses the Calendar View by selecting an

individual in the List View. Then he finds a strong connection with another individual and proceeds to look at those two individuals together in the Calendar View, in order to identify when they are associated with each other, on what dates.

He has already experienced the utility of Jigsaw in his work by helping the police to arrest a criminal. The police were trying to find a criminal, and he searched for the name of another related person in the document collection and examined connections between the two, finally identifying an address where the criminal might be. He liked both visual and investigative support by Jigsaw:

I think Jigsaw's strength is its visual support, and investigative support. It would have been impossible without it...When I showed the results and connections to other colleagues, it was easy for them to understand how a certain person is connected to others, that is, providing the context.

One of the issues he encountered is determining the amount of data to import. If he imports documents from the last two years, it would be easier for him to see long-term trends and links between associates. However, it will take significant time to import the documents and clean up the entities (i.e., manually fixing incorrect entity identification). If he imports documents of only several months, it will be faster to import and handle, but he will be able to see short-term trends only. Considering the trade-off, he must spend significant time considering the optimal point.

3.5   P5: Intelligence Analyst at a National Lab

P5 is an intelligence analyst at a national laboratory. His department receives a number of resumes for post-docs and researchers who are applying to the lab throughout a year. Among the applicants, he is interested in finding someone who has expertise in a specific area, and being an intelligence analyst, he utilizes his analysis skills in finding candidates. To identify who has the specialty the laboratory requires, he looks at not only the technology/specialization an applicant explicitly expressed, but also publications, co-authors and collaborators, and previous institutions of an applicant.

Before using Jigsaw, he performed the task using Analyst's Notebook, which he felt was limited because he had to manually type in all the data in resumes to the Analyst's Notebook and create connections:

I was having to do it one at a time and tying them together manually, really. I mean I was using Analyst's Notebook, but pretty much you have to put the data in by yourself. There's not a lot of ways to pull in data, so it's really a lot of work, especially when there's a lot of resumes in our system.

When he was introduced to Jigsaw, he found that it might be a good fit for his task–finding connections between people and technologies (specialties). Since then, he had been using the system for 7 months. He usually works with 10-12 resumes and creates entity types such as institutions, organizations, technologies (specific types of technologies), publications, co-publications, employment history, dates, and emails. Especially, he tries to find who is connected with whom within a community. By investigating the connections, he ultimately seeks to find an expert in a specialized area, for example, an energy expert.

Working with resumes, he found the Document View really helpful. Interestingly, he uses the view for "identifying what views to use," as well as for simply reading the documents. He first reads a couple of documents in Document View and determines which other views would be appropriate and effective for analyzing those documents. That is, by getting a brief overview of what each document looks like, he decides which views to utilize for investigation. Among other views, the List View helps him clearly visualize who is connected to what technology or organization. Particularly, the view is useful when an applicant does not explicitly mention a certain technology as specialization but still has

background or experience relevant to the technology in the past. Using the List View, P5 could see possible connections and find a good candidate who is knowledgeable about a technology, which would have been much more difficult otherwise. He also often uses the Document Cluster View when he wants to see how the documents can be categorized. He then would select a specific document cluster to read some of the documents in that category.

He mentioned that entity identification and being able to focus on the inter-connectedness of ideas between people and technologies were especially beneficial. Due to these features of Jigsaw, the process of investigating resumes has become more efficient and effective, as it helps him bring connected people together that he might not have been able to see otherwise.

While entity identification is a benefit, it also seemed to be a barrier to him. Because Jigsaw does not always recognize all the entities as he wants, he has to go through the documents and clean up entities after the initial import.

3.6   P6: Intelligence Analyst at Air Force

P6 is an intelligence analyst at the Air Force. He was working on a project in which they examined the Research and Development Descriptive Summaries, which are budget documents for R&D programs in the Department of Defense [29]. It is a large document collection (>10,000) from 20+ agencies such as Air Force, Navy, DARPA, etc., and each document contains a one-page budget summary including description and justification.

By analyzing these documents, he sought to identify common themes, what programs are similar, what makes them similar, and who are working on similar topics. Because it was a large document collection, he had no idea of how they are related in the beginning. So he searched for a visual analytics tool that can help his analysis, and finally found Jigsaw.

He had been using Jigsaw for 14 months when we interviewed him, and his goal was to find related tools, topics, technology, and people working on a similar topic in the documents and to discover clusters of data that he might not notice. Instead of deeply analyzing the document collection, he wanted to highlight similarities and connections among the documents so that he could narrow down to specific entities to further investigate.

For this task, he first wanted to find entities that had a similar function. That is, he used Jigsaw for a similar tool search and a synonym search. For example, if tool A forecasts certain type of data, then he tried to find other similar tools and examine their functionalities. Whenever he found a tool of interest, he queried it in the Jigsaw control panel and read returned documents that contained the tool. He also did the same process for a verb that expresses specific functionality such as "predict."

He imported all 10,000 text documents into the system and added entity types such as agency, name of technology, and description. Usually he started with the circular layout in the Graph View, in which entities appearing within multiple documents are shown inside a circle. The stronger the connection is, the closer to the center the entity is shown. That is, he sought to learn the most common themes (Fig 4) among the document collection. From there, he searched for interesting terms and looked for the documents that came up. Then he opened the List View to further explore the connections. Sometimes he would examine immediate clusters–a group of documents towards the center–in the circular layout in the Graph View and highlight those entities so that he could explore them in the List View. By doing this process repeatedly, he was able to find what he wanted.

With Jigsaw, he could effectively search for similar tools and technologies that required further investigation. Through the circular layout, he was able to easily identify where to start his investigation when he did not have a clue where to begin. Even when he had some idea of what he would investigate, Jigsaw helped by showing other interesting documents and keywords so that he could investigate further, which led to a better set of documents instantly. Once he got a set of documents of interest, then he could see important
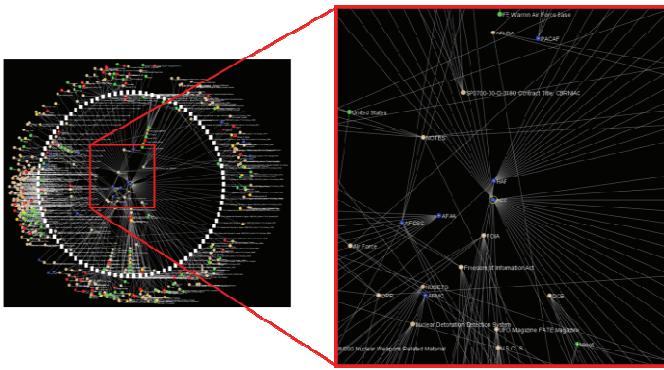
Fig 4. Critical research areas identified by the circular layout in the Graph View

connections such as what are related topics, what kind of programs are related to the topics, and who are working on the programs. He emphasized that Jigsaw was particularly helpful when convincing people because the visualization itself drew other people's attention to his work:

..and it was pretty. People who received the briefs with that picture in there, they loved it. They said that the coolest picture was the Graph View in Jigsaw. That's a sign [that] it's a good analytic tool. But having that graphic that you are able to show the most central themes in this set of documents and say that's because of this and this…It's definitely nice to look at that stuff.

Visualization helps convince people. People pay attention a lot more than if I just told them. It proves itself.

In terms of problems, he pointed out that there was no way to easily select a subset of data in Jigsaw. Because he had to select each document one by one, he found it difficult to look at a group of documents at a time. He also wanted some mathematical measures such as centrality between documents so that it can give a more objective sense.

4 FINDINGS AND DISCUSSION – HOW THEY USED JIGSAW

Reflecting on our interviews and discussions with analysts, a number of common themes emerged. Ahead of time, we cared about how an interactive visual system for investigative analysis assisted document sensemaking in various domains, and what kind of issues emerged upon the use of a system. We also hoped to see if professionals used the tool in unexpected ways. We characterize four dimensions in this section.

4.1 Types of Tasks

While all individuals in the study were from different domains and had unique problems, we could classify their tasks into a few categories, described below.

- *Relationship/connection between entities*: P2 (business analyst), P4 (analyst at a national lab), and P5 (analyst at a police department) searched for a tool that could help them make connections and find complex relationships between entities that were not apparent simply by reading documents. They were investigating emails to detect financial fraud, crime reports to make linkages, and resumes to find a candidate with specific expertise, respectively. Rather than seeing the big picture and understanding the entire story, they did a more targeted investigation. For this type of task, it seemed that Jigsaw's model of connection was sufficient and the analysts felt it highly useful and beneficial to their task.

- *Search/Comparison*: P1 (Aerospace engineering researcher) and P6 (analyst at Air Force) used the visual analytics system to compare documents and search if the documents contain specific keywords. P1 explicitly compared two sets of documents, examining whether a set of documents contain similar concepts identified in the other set of documents. P6 tried to find if certain tools or technologies have similar functionalities within the document collection, using the system for a similar tool search and synonym search.

- *Understanding*: P3 actively looked for a tool that can help her understand the huge collection of documents, and thus she used the system to attain a better, clear understanding of the documents. By "understanding," we mean gaining an overview of the documents. She did not conduct a detailed analysis using the system. Instead, based on the overall understanding she gained from the system, she set the basis for a further analysis, which she performed using other software.

In addition to these three types of tasks, some of the analysts found the system useful as a communication aid as well.

- *As a communication aid/shared understanding of data*: P2, P4, and P6 commented that through the visualizations created by the system, they were able to effectively share findings and connections with colleagues. While they did not initially expect that effect, it seemed clear that the visualization system had a persuasive power and added value in communicating with others.

4.2 Learning the System

Jigsaw is a relatively complex system and has a number of features that may not be intuitive at first. All the professionals we interviewed had technical knowledge enough to learn and utilize the system. To learn about the system, every person watched the video tutorials available on the web [19] and gained a general idea of how the system works before they started using it. While most of the users also read the tutorial document and found the tutorial very helpful, they admitted that mostly they went through by themselves and interacted with the second author to ask questions and solve issues that arose.

A few of the analysts told us that they did not have any problems in learning to use Jigsaw and the system was pretty intuitive and easy to use. Still, many of the users seemed to encounter a learning curve. This was more about making sense of "how to better analyze my data using this tool," rather than about learning how to use the system itself. Even after they became familiar with the system and its features, they tried to find the best way to analyze their own data among a number of views and ways to display the data in each view, thus "constructing a frame" [22]. They had questions such as "which views are most appropriate for my data and task?" or "what entity types do I want to put in this column?" Once they found the optimal approach in their own way, they seemed to settle down with it; their usage pattern did not change much.

4.3 Unexpected Use of the System

In the study, we recognized that the professionals sometimes used the system in unexpected ways, which may provide some insights for design. The first one is *using the views for evidence/output generation*, rather than for exploration. Jigsaw was originally designed for investigative analysis; it helps a person determine which document to read next. But often, people used it as a search tool with a visual aid; after they found specific connections by searching for a keyword, they created a representation of these connections. For example, in the case of P1, she wanted to create a mapping between two documents and used the List View to more effectively generate the mapping, which formerly was done manually (Fig 5). In these cases, it seemed that people missed the investigative power but instead used the system as a presentation aid.
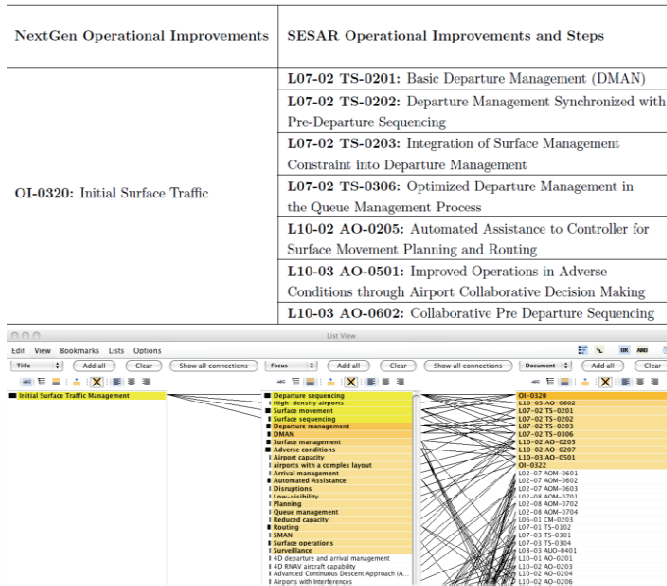
NextGen Operational Improvements	SESAR Operational Improvements and Steps
OI-0320: Initial Surface Traffic	L07-02 TS-0201: Basic Departure Management (DMAN)
	L07-02 TS-0202: Departure Management Synchronized with Pre-Departure Sequencing
	L07-02 TS-0203: Integration of Surface Management Constraint into Departure Management
	L07-02 TS-0306: Optimized Departure Management in the Queue Management Process
	L10-02 AO-0205: Automated Assistance to Controller for Surface Movement Planning and Routing
	L10-03 AO-0501: Improved Operations in Adverse Conditions through Airport Collaborative Decision Making
	L10-03 AO-0602: Collaborative Pre Departure Sequencing

Fig 5. A mapping between two sets of documents created manually vs. by Jigsaw.

One of Jigsaw's goals is to help analysts with a large number of documents. In the study, however, we found a few experts using Jigsaw for a relatively small number of documents. In those cases, they worked with information-dense documents and did not want to be overwhelmed by the information shown in the system. Thus, they separated documents into several projects, making each project manageable. P5 said that he usually imported only about 10 documents into the system for his analysis:

Sometimes it's as low as five. I tried a couple of hundred at first, but it was really too much information. Now I try stick to under 10. For most of the time, I've done 12. I don't like putting too much into one project because it becomes too complicated.

Some might say it's too few [documents] to use Jigsaw, but it's not that easy. Resumes have condensed [content] in a few pages. You're looking at, in addition to everywhere the person's been, people who they've worked with. Typically you have a list of publications that have 5 or 6 names, a couple publishers per line. Using those ten documents to compare to another ten documents, it begins to become more complex. Ten documents...doesn't sound [like] a lot, but it is quite a bit of information.

In contrast, we found some analysts using Jigsaw itself as a database. Those people wanted to merge new incoming documents with an existing Jigsaw project and build a historical dataset so that later they can explore it in a single project file. Three professionals emphasized that they wanted to accumulate new reports to the existing project so that they do not need to re-run all the computations and start over the entity clean-up process. Several users commented:

P4: Analysis is ongoing, it's never done. I want to build on previous data...I'm trying to figure out how much data should I import. The more data I import...I can see long-term trends and make long-term connections between associates better. But the issues would be time to import and clean up entities on a bigger set of data.

P3: I have about 30 Jigsaw projects. An issue is that 2000-3000 is the maximum for Jigsaw to handle. I mean, the processing time is acceptable for that amount. This is HP documents from 2009 to 2011. I can't do it from 2002 to 2011 because it's [going to] be more than

10,000. So I just do it like from 2002 to 2004, something like that. If it was doable, I'd definitely import them all at once.

This notion of "file management" or "project management" could have an important design implication for analytics systems. We will discuss this matter more in detail in the next section.

4.4 Issues and Problems

Some issues and problems in using the system have been identified through the study, at various levels. Here, we want to highlight a few prominent issues.

One of the initial barriers in working with Jigsaw was technical issues in the preparation stage such as importing data into Jigsaw and identifying entities. Technically, Jigsaw can read in documents in a variety of file formats including text, html, pdf, Word, and Excel files. However, plain text files or Excel files are the most reliable type of file to import, and users are recommended to transform their documents into text or Excel files if possible. Because people often have documents as pdf or Word files with complex formatting and images, importing these files directly into Jigsaw is less reliable, and therefore, users need to put extra effort to convert their documents into plain text or Excel files. Identifying and working with entities is another similar issue. While Jigsaw provides automated entity identification via third party libraries, which attracts many users, it is not perfect and many false positives and negatives can occur. In order to fix incorrect entity identification, users have to manually choose each word to add, remove, or modify. Creating a new entity type is common because users have their own interest when working with documents. Actually, all of the professionals in the study created their own entity types applicable to all documents in the collection. Users also have to go through the process of entity aliasing, which creates aliases for entities that are identical but worded differently.

While this grounding process–both importing documents and cleaning up entities–does not seem to be a serious issue in terms of the analysis process, it turned out that most people considered it as one of the biggest difficulties in using Jigsaw. Without addressing these issues in the beginning, they are not even able to see their data properly displayed in the views. When they encounter any problem in the process, they typically needed to contact the developers and get instructions, which could be cumbersome and even daunting to someone without technical background. All the professionals in the study mentioned that the initial processing required a lot of time and effort. Once they undergo this stage, however, they became easily engaged in working with the views.

P5: One of the biggest difficulties that I encountered was entity identification. When importing data, because a resume is not a type of data that Jigsaw is designed to read, I still have issues with entity identification. Well, I have to go through it one by one. Have to do a lot of cleaning after the initial importing of data.

Another issue that the analysts faced was that Jigsaw has very limited filtering options and users are not able to easily select a subset of data in the views. Currently, once Jigsaw reads in documents, all the operations and computations are run upon the entire set of documents. That is, once users have ingested a collection of documents into Jigsaw, all the document and entities are "active." If they want to temporarily exclude some documents and explore connections only for another set of documents, the only way to do it is to start over in the data importing process. They have to decide which documents in the collection to examine, create another collection of those selected documents, and import the documents into the system. In other words, there is no easy way to select a subset of documents while working with views. Users wished that they had a better, flexible way to have a certain set of documents, as expressed in the quote below. P6 compared IN-SPIRE [15], a visual analytics system for text analysis which provides an overview of the key themes and trends across a document collection,

to Jigsaw when discussing this feature. In IN-SPIRE, he was able to make a selection of documents even after all the dataset was displayed:

I started IN-SPIRE about at the same time I used Jigsaw. The thing I liked best in IN-SPIRE, which Jigsaw doesn't have, was that you have all the dataset up there on the screen (the galaxy view), and I could easily select across all data, make the selection and make the rest of them outliers, and have just the ones I have selected. In Jigsaw, I have to have all sets of data.

We assume that if selecting and working with a subset of data was easy enough, some of the professionals might not have had to segment their data into several Jigsaw project files, since they would not have had the information overload issue.

Finally, there was an issue of trust in the system. While people favored the automatic power of the visual analytics system, they did not seem to solely rely on the system as in this quote, which is a common behavioral pattern of analysts [22].

P2: I'm the only one who's using it in our team. They don't think it's reliable enough.

It seems that this mistrust is raised when the analysis process does not flow smoothly. When the system fails to import documents or identify entities that the analysts want to see, they tend to attribute it to the lack of system reliability. This tendency is more likely to appear to people with less technical capability, those who are not willing to put extra efforts into troubleshooting. Or simply, some experts think that the system assists part of their work more efficiently, but ultimately, they believe that they can do the job more accurately. For example, after working with Jigsaw, P1 double-checked its findings with those from a manual process in order to validate her analysis:

P1: Finally, we carefully reviewed descriptions of OIs for which one or many counterparts were identified with experts, in order to ensure that the themes and ideas behind these concepts were indeed analogous. It was found that the mappings obtained through Jigsaw were similar to the ones obtained manually, and thus we could say that Jigsaw offers a valuable alternative to our manual approach.

5 DESIGN IMPLICATIONS

Based on findings discussed in the previous section, we derive a series of design implications for next generation systems.

5.1 Supplement Automatic Entity Identification

While there exist a number of entity identification systems [4,10,12] and visual analytics systems that incorporate entity identification [2,3,6,31], the process typically is not perfect. Some entities may not be identified at all, some may have an incorrect entity type assigned, and some identified ones may not be entities. Systems should provide ways to correct such errors, and the process needs to be intuitive and efficient. While Jigsaw allows users to modify, remove, add, and alias entities, professionals pointed out that it is still not a simple, easy process, as mentioned in the previous section. For capabilities such as entity aliasing, the process is not automatically going forward. That is, when new documents are imported, the analyst must manually create the aliases again.

Another issue is that although Jigsaw allows users to create a new entity type and specify the instances of that entity, our analysts seemed to be unaware of the feature. For example, once they create a new entity type "Company name," they could create a text file that has each different possible entity value such as "HP," "Apple," "IBM," etc. While every user in the study created their own entity types, most of them did not know about this feature but specified each entity every time they opened a new project, which took

significant time. Four users suggested a feature that Jigsaw already provides:

P5: I suggest an entity library you can draw on for every project. Then you wouldn't have to keep creating new entities…It will be nice to have an entity list that you can apply to each project you do and not have to recreate them. For example, a list of universities that would be identified every time, a list of technologies that would be identified every time, so you only have to make the list once.

We suspect that the way to create a new entity type was not intuitive or salient enough to users. The feature could have been more nicely incorporated with the entity identification work flow, for example, by asking them to type a list of entities instead of importing a text file, so that users do not need to create an extra file outside the system.

5.2 Allow Flexible Data (document) Management

Previously, we discussed that some analysts worked with multiple sets of a small number of documents while others wanted to accumulate documents into one project and build a database. In most cases, analysts needed to experiment with the "right" number of documents to import in order to optimize their analysis. One of the reasons for this issue of data size is information overload (e.g., not wanting to be inundated by information shown), which can be addressed by providing flexible data management. Currently, once a user imports a document collection into Jigsaw, all the documents and entities are active, which can be overwhelming. If a user wants to investigate only part of the document collection, there is no easy way to do so except to create another subset of the collection and import it. This is inefficient especially when users want to examine different subsets of documents in a single document collection. Ultimately, users desire to be able to flexibly work with documents within a single database, and a system should provide the ability to easily select a subset of documents to investigate once users import a document collection. For example, a system could provide a way to choose a subset of documents and run analysis only for the selected documents. Or a system could allow users to temporarily exclude a set of documents so that they can work with the remaining documents only. We assume that if selecting and working with a subset of data was easy enough, some of the professionals would not have had to segment their data into several Jigsaw project files since they would not have had the information overload issue.

Systems also need to provide a way to easily accumulate documents into an existing project. In many cases, users may want to build a database over time, especially when they receive documents incoming regularly. Currently in Jigsaw, if users want to add only one or two documents to the existing Jigsaw file, they can but they must repeat the process of computational analysis on the entire document collection. This is inefficient because the analyses not only must be re-run, but also users have to perform the entity clean-up process again. Often, users do not have a complete set of documents prior to investigation or they receive new documents continuously. They would want to simply "merge" new documents into the existing file, upon which entities are already cleaned up and computational analysis is done.

5.3 Empower with Numbers

Jigsaw was developed for unstructured text data and does not provide wide-ranging statistical analysis per se. For example, in order to show connection strength, the system uses colors (darkness) or list order. However, most of the analysts in the study strongly expressed that statistical functionality would be really desirable. Depending on the domain and task, analysts often need to convert results from investigative analysis into evidence, which is better supported with quantified information such as descriptive statistics or counts. In the study, several users wished to have statistical importance metrics such as degree centrality, betweenness,

closeness, or others so that they could have more accurate metrics of the connections between entities and documents. Even for investigative analysis systems that deal with unstructured data such as text, it seems important to have simple statistics and measures.

5.4 Consider Allowing Visualization Modification

The professionals in this study wanted to have more control and flexibility over the visualizations. They sometimes wanted to be able to annotate, mark, and change the representations. Such changes may not be feasible or desirable from the point of view of the system, however. For example, the visualizations presented by a system may communicate analysis metrics or results computed about the data. Allowing the user to modify the visualization would be, in this case, inappropriate because it could make the visualization present the analysis data inaccurately. Conversely, allowing the analyst to simply highlight or augment the visualizations would not violate the fundamental data-to-representation mapping. Presently, Jigsaw allows no view augmentation. Should it? It is important that system designers and developers carefully consider the style of changes, if any, that viewers can make to a system's visualizations.

5.5 Invest in Tutorial

Usually, visual analytics systems for investigative analysis tend to have a large number of features and interaction techniques, which makes it difficult to become familiar with a tool without any external aids such as one-on-one training or written instructions. In many cases, tutorials seem to be quite important and helpful for learning visual analytics systems. While some people may argue that users do not pay much attention to tutorials, all of the analysts we interviewed said that they put considerable time and effort in reading the tutorial document and watching video tutorials.

Another reason for the importance of a tutorial is intermittent use of a system. Many professionals pointed that they do not use the system on a regular basis. Instead, they used the system when they have enough time, when they receive new data, or when they need to prepare a brief. Consequently, they often forgot about some functions and operations and had to revisit tutorials. Thus, it is desirable to provide an intensive but still easy-to-understand tutorial. For example, breaking down the tutorial into subtopics with use-cases and examples would be really helpful, as the users commented:

P5: For learning, I mainly used the video tutorial. It was very useful actually. They are good because they're broken down into topics and you can pick what you need help with. I like it a lot.

P4: I wished [to have] a better tutorial though. I want to see more examples about each view so that I can find the best way to analyze my own data.

5.6 Jigsaw-specific recommendations

The study helped identify issues and future work for Jigsaw:

- *Focus on useful views*: While different users have different preferences of views, it was clear that the List View was most useful. We suggest that future development focus on improving the features and interface in the List View, as it will definitely benefit real world users. The Document View, the Document Cluster View, and the Graph View were also used by several analysts. Multiple analysts mentioned that they did not find the Timeline useful, and the Scatterplot View was not even used at all. Those views may need significant changes or be removed.
- *Give them power to control*: When working with their own data, users want to actively interact with the system because they have their own goals and expectations from the system. While Jigsaw is very good at "showing" documents and entities in different ways, professionals wanted to be able to annotate and manually alter visual representations.

6 CONCLUSION

In order to evaluate long-term, field use of Jigsaw, we conducted in-depth case studies with analysts from a variety of domains. We interviewed six investigators from the intelligence, academic, and law enforcement communities who had been using the system for a period of 2-14 months. We asked them about their use of Jigsaw, the types of data they were working on, and difficulties they encountered. Analysts used Jigsaw for finding relationships, comparing documents, getting an overview, and sharing analytical products with others. Their primary difficulties included importing data into Jigsaw, identifying entities in the preparation stage, and selecting a subset of data during data exploration.

The contributions of this work thus include:

- Identification of real-world cases of how an interactive visual system for investigative analysis assisted document sensemaking in various domains and tasks;
- Discussion of issues and findings that emerged upon the use of the visual analytic system;
- Development of design recommendations for the system and future visual analytics tools.

A growing number of visual analytics systems are being developed and used in practice. Assessing the utility and value of a system is essential for improving it, and we recommend the case study as a useful evaluation method. It helps to understand the types of tasks and problems a system can address and to identify strengths and weaknesses of a system in real world settings.

REFERENCES

[1] C. Andrews. Space to think: Sensemaking and large, high-resolution displays. PhD Dissertation, 2011. http://scholar.lib.vt.edu/theses/available/etd-08232011-160350/

[2] E. Bier, E. Ishak, and E. Chi. Entity Workspace: an evidence file that aids memory, inference, and reading. *Proceedings of the IEEE International Conference on Intelligence and Security Analysis*, pp. 466–472, May 2006.

[3] E. Bier, S. Card, and J. Bodnar, Entity-based collaboration tools for intelligence analysis. *Proceedings of IEEE VAST*, pp. 99–106, Oct. 2008.

[4] Balie. http://balie.sourceforge.net/ (accessed 01 March 2012).

[5] E. Braunstein, C. Görg, Z. Liu, and J. Stasko. Jigsaw to Save Vastopolis - VAST 2011 Mini Challenge 3 Award: "Good Use of the Analytic Process", *Proceedings of IEEE VAST '11*, Providence, RI, pp. 323-324, October 2011.

[6] A.R. Chappell, A.J. Cowell, D.A. Thurman, and J.R. Thomson. Supporting mutual understanding in a visual dialogue between analyst and computer. *Proceedings of Human Factors and Ergonomic Society 48th Annual Meeting* (New Orleans, LA), Human Factors and Ergonomics Society: Santa Monica, CA, pp. 376–380, September 2004.

[7] K. Charmaz. Qualitative interviewing and grounded theory analysis. In *Handbook of Interview Research: Context and Method*, (J. F. Gubrium and J. A. Holstein, eds.), CA: Sage Publications, 2002.

[8] C. Chen and M. Czerwinski (Eds.) Introduction to the Special Issue on Empirical evaluation of information visualizations. *International Journal of Human-Computer Studies*, vol. 53, no. 5, pp. 631-635, 2000.

[9] G. Chin, O.A. Kuchar, and K.E. Wolf. Exploring the analytical processes of intelligence analysts. *Proceedings of ACM CHI*, pp. 11–20, April 2009.

[10] Cicero. Language Computer Corp. http://www.language computer.com/ (accessed 01 March 2012).

[11] Compendium. http://www.compendiuminstitute.org (accessed 11 July 2012).

[12] GATE − General Architecture for Text Engineering. http://gate.ac.uk/ (accessed 01 March 2012).

[13] C. Görg, Z. Liu, N. Parekh, K. Singhal, and J. Stasko. Jigsaw meets Blue Iguanodon - The VAST 2007 Contest, (Contest Paper), *Proceedings of IEEE VAST '07*, Sacramento, CA, pp. 235-236, October 2007.

[14] D. Gotz, M. X. Zhou, and Z. Wen. A study of information gathering and result processing in intelligence analysis. In IUI 2006 Workshop on IUI for Intelligence Analysis, 2006.

[15] E. Hetzler and A. Turner. Analysis experiences using information visualization. *IEEE Computer Graphics and Applications,* pp. 22-26, 2004.

[16] J. Hughes, V. King, T. Rodden, and H. Andersen. The role of ethnography in interactive systems design, *ACM interactions*, vol. 2, no. 2, pp. 56-65, April 1995.

[17] J. Hughes, J. O'Brien, T. Rodden, and M. Rouncefield. Design with ethnography: A presentation framework for design, *Proc. of Design of Interaction Systems '97*, ACM Press, New York, pp. 147–159, 1997.

[18] i2 – Analyst's Notebook. http://www.i2group.com/us (accessed 11 July 2012).

[19] Jigsaw tutorial. http://www.cc.gatech.edu/gvu/ii/jigsaw/tutorial (accessed 11 March 2012).

[20] Y. Kang, C. Görg, and J. Stasko. The evaluation of visual analytics systems for investigative analysis: Deriving design principles from a case study. *Proceedings of IEEE VAST,* pp. 139–146, October 2009.

[21] Y. Kang, C. Görg, and J. Stasko. How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 5, pp. 570-583, May 2011.

[22] Y. Kang and J. Stasko. Characterizing the Intelligence Analysis Process: Informing Visual Analytics Design through a Longitudinal Field Study. *Proceedings of IEEE VAST '11*, Providence, RI, pp. 21-30, October 2011.

[23] A. Komlodi, A. Sears, and E. Stanziola. Information Visualization Evaluation Review. ISRC Tech. Report, Dept. of Information Systems, UMBC.UMBC-ISRC-2004-1.

[24] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 2012. Appeared online: 30 November 2011.

[25] A. Perer and B. Shneiderman, Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. *Proceedings of ACM CHI*, pp. 265–274, April 2008.

[26] O. J. Pinon, D. N. Mavris, and E. Garcia. A Visual Analytics Approach to the Qualitative Comparison of the SESAR and NextGen Efforts. *9th AIAA Aviation Technology Integration and Operations ATIO Conference*, 2009.

[27] C. Plaisant. The challenge of information visualization evaluation. *Proceedings of AVI*, pp. 109–116, May 2004.

[28] C. Plaisant, G. Grinstein, J. Scholtz, M. Whiting, T. O'Connell, S. Laskowski, L. Chien, A. Tat, W. Wright, C. Görg, Z. Liu, N. Parekh, K. Singhal, and J. Stasko. Evaluating Visual Analytics at the 2007 VAST Symposium Contest, *IEEE Computer Graphics and Applications*, Vol. 28, no. 2, pp. 12-21, March 2008.

[29] Research and Development Descriptive Summaries (RDDS). http://www.dtic.mil/dtic/stresources/budgetNplanning/rdds_desc.html (accessed February 8, 2012).

[30] A. Rose, C. Plaisant, and B. Shneiderman. Using ethnographic methods in user interface re-engineering, *Proceedings of DIS '95: Symposium on Designing Interactive Systems*, ACM Press, New York, pp. 115–122, 1995.

[31] A. Sanfilippo, B. Baddeley, A.J. Cowell, M.L. Gregory, R. Hohimer, and S. Tratz. Building a human information discourse interface to uncover scenario content. *Proceedings of the 2005 International Conference on Intelligence Analysis* (McLean, VA), May 2005.

[32] P. Saraiya, C. North, and K. Duca. An evaluation of microarray visualization tools for biological insight. *Proceedings of the Symposium on Information Visualization '04* (Los Alamitos, USA), pp. 1−8, 2004.

[33] P. Saraiya , C. North , V. Lam, and K.A. Duca. An Insight-Based Longitudinal Study of Visual Analytics, *IEEE Transactions on Visualization and Computer Graphics,* vol. 12, no. 6, pp. 1511-1522, November 2006.

[34] J. Scholtz. Beyond usability: Evaluation aspects of visual analytic environments. *IEEE VAST,* pp. 145–150, October 2006.

[35] J. Scholtz. Developing guidelines for assessing visual analytics environments. *Information Visualization*, vol. 10, no. 3, pp. 212–231, 2011.

[36] A. Selvin and S. B. Shum. Narrative, sensemaking, and improvisation in participatory hypermedia construction. *CHI '08 Workshop on Sensemaking*, Florence, Italy, 2008.

[37] J. Seo and B. Shneiderman. Knowledge discovery in high dimensional data: Case studies and a user survey for the rank-by feature framework. *IEEE Transactions on Visualization and Computer Graphics*, 12(3), pp. 311−322, 2006.

[38] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. *Proceedings of ACM BELIV*, pp. 1–7, May 2006.

[39] J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, vol. 7, no. 2, pp. 118–132, 2008.

[40] J. Stasko, C. Görg, and Z. Liu. Sensemaking across Text Documents: Human-Centered Visual Exploration with Jigsaw, *CHI '08 Workshop on Sensemaking*, Florence, Italy, April 2008.

[41] A. Strauss and J. Corbin. *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications: Newbury Park, Calif, 1990.

[42] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis: Concepts and Methods. *Proceedings of ACM CHI '06* (Montreal, Quebec), pp. 801–810, April 2006.

Visual Analytics Support for Intelligence Analysis

Carsten Görg*
University of Colorado

Youn-ah Kang†
Google Inc.

Zhicheng Liu‡
Stanford University

John Stasko§
Georgia Institute of Technology

ABSTRACT

Intelligence analysis challenges investigators to examine large collections of data and documents and come to a deeper understanding of the information and events contained within them. Visual analytics technologies hold great promise as potential aids for intelligence analysis professionals. We describe our research to better understand intelligence analysis processes and analysts, learn how visual analytics can help investigators, and design visual analytics systems to serve in this role. To illustrate these ideas, we present a hypothetical intelligence analysis scenario that explores a collection of text documents using the Jigsaw system that we have created. The system combines computational analysis of document text with interactive visualizations of the document contents and analysis results. Evaluating such systems is very challenging and the article concludes by discussing potential evaluation methodologies for these types of systems.

Keywords: Visual analytics, investigative analysis, intelligence analysis, information visualization, knowledge acquisition, data exploration, case study, qualitative user study.

1 INTRODUCTION

Visual analytics is a relatively new research field that integrates the interactive visualization and exploration of data with computational data analyses [8]. Intelligence analysis has been one of the key application domains of visual analytics since the area's inception in 2004, facilitated by the creation of the National Visualization and Analytics Center by the Department of Homeland Security. An initial research roadmap [11] described challenges and goals of the new field and identified tasks, data, and analytical scenarios focused on homeland security and prevention of terrorism.

Enabling insights through the analysis of large amounts of diverse and dynamic data was the underlying grand challenge in the research agenda. As stated in [11], "The analysis of overwhelming amounts of disparate, conflicting, and dynamic information is central to identifying and preventing emerging threats, protecting our borders, and responding in the event of an attack or other disaster. This analysis process requires human judgment to make the best possible evaluation of incomplete, inconsistent, and potentially deceptive information in the face of rapidly changing situations to both detect the expected and discover the unexpected."

Intelligence analysis requires investigators to gather as much available data as possible in order to better understand a situation and then make judgments about the appropriate next steps to take. Two fundamental types of investigative scenarios exist within the intelligence domain: (1) targeted analysis scenarios, in which analysts are tasked with examining specific people, organizations, or incidents, as well as locations and dates, in order to either investigate past events or uncover an imminent threat, and (2) open ended,

*e-mail: Carsten.Goerg@ucdenver.edu

†e-mail: ykang@google.com

‡e-mail: zcliu@cs.stanford.edu

§e-mail: stasko@cc.gatech.edu

strategic analysis scenarios, in which analysts are tasked with learning as much as possible about a person, organization, country, or situation in order to gain a deeper understanding, conduct an accurate assessment, and possibly make a prediction on the likely chain of events that will occur at a later point in time. Examining and understanding large collections of textual documents plays an important role in both types of these scenarios. Analysts must gather nuggets of information within textual documents from diverse sources, ranging from reports from field agents to open source news articles. Examining textual documents is fundamentally a slow process (due to the sequential nature of reading) and it is challenging for the analysts to keep track of what they discovered and form an internal mental model that represents a coherent picture of the events, people, places, and organizations discussed in the documents. Uncovering and understanding the connections between those entities across a large collection of documents is one of the key challenges they face.

Based on a cognitive task analysis of working analysts, Pirolli and Card [9] identified a number of "pain points" in the intelligence process that are particularly challenging to human analysts. These pain points include the costs of scanning, recognizing (assessing), and selecting items for further attention; the costs of shifting attention and control; the limited span of attention for evidence and hypotheses; and the difficulty of generating alternative hypotheses. All these challenges are exacerbated when the amount of data to examine grows larger and larger. Today's "big data" technologies often make the acquisition of data easier, but they present increasing challenges to analysts who must review and investigate all that data. In this article we highlight a number of our research projects on intelligence analysis from the last five years, including an observational study to gain a better understanding of the intelligence analysis process and its characteristics, the development of a visual analytics system that integrates computational text analyses with interactive visualization in order to explore collections of documents, and an evaluation of the utility of the system via a controlled laboratory experiment as well as observational case studies of extended use of the system in the field.

2 INTELLIGENCE ANALYSIS PROCESS

Analyzing and understanding end-users needs and tasks is one of the fundamental requirements for creating useful computational tools. To better understand intelligence analysis, it is important to explore the mindset and methodologies of analysts as well as the fundamental processes they conduct. Heuer [4] examined the psychology of intelligence analysis and the types of mental reasoning analysts must engage in. In particular, he identified a number of challenges analysts must confront in the analytical reasoning process. For example, when people encounter a situation of uncertainty, they typically will develop a single hypothesis explaining the situation and will work to gather evidence confirming the hypothesis. Intelligence analysts, however, are trained to develop multiple hypotheses and seek out information that can discredit many of the hypotheses.

Many other researchers have studied the intelligence analysis process in order to construct abstract models of it. While a number of process models exist, most involve some form of iterative cycle of exploration, including steps such as data collection, pro-

cessing, analysis and production, dissemination, and planning and direction [1].

Pirolli and Card's notional model of the sensemaking loop for intelligence analysis [9] has been widely cited and adopted by researchers within the visual analytics community. It consists of a linear set of states characterizing both data and process flow in an investigation. Analysts iterate through this process over the course of an investigation. At a high level, the model contains two primary loops: a foraging loop in which analysts collect data and evidence, and a sensemaking loop in which analysts reflect on the data in order to generate schema and hypotheses about the situation and ultimately construct a presentation of the findings. Each loop contains three stages that further refine the process and both loops are connected through an overarching reality/policy loop.

This model broadly characterizes the workflow used for analysis activities and it has guided the development of a number of computational tools. However, it abstracts substantially from analysts' work in the real world and does not provide an adequate level of detail necessary to develop tools that analysts can integrate seamlessly with their existing workflow. Furthermore, not all analysts agree that the linear structure of the model captures the way they work. Dr. Kristan Wheaton, Professor at the Department of Intelligence Studies at Mercyhurst College, proposed an alternative model in which modeling, collection, analysis, and production stages take place in parallel, just with different emphases over the course of an investigation. At the beginning, emphasis focuses on modeling and then throughout the investigation it shifts to collection, analysis, and finally production.

To better understand the analytical process and its requirements in the intelligence domain, we conducted our own qualitative user study [6]. Professor Wheaton provided us with the opportunity to observe three teams of intelligence analysts in training within the intelligence program at Mercyhurst College. The student teams each conducted an intelligence analysis project throughout an entire academic term (ten weeks). One team consisted of four undergraduate students and performed analysis on a project for which we served as a "client"; the other two teams consisted of graduate students and conducted a structured analysis on projects provided by external clients.

We found that four processes dominated the overall workflow: construction of a conceptual model, collection, analysis, and production. The study helped us to better understand some misconceptions that visual analytics researchers may harbor about intelligence analysis. For instance, analysis is typically not about finding an answer to a specific problem and it does not evolve in a sequential process. Instead, analysis is often about determining how to answer a question, what to research, what to collect, and what criteria to use. The process is often organic and parallel. Another misconception is that intelligence analysts typically operate as lone investigators, researching some problem. We, conversely, found that collaboration is commonplace and crucial, frequently being asynchronous. Also, the student analysts we observed did not seek grand, monolithic computational analysis tools. Instead, the teams used a variety of computational tools with many being small applications used for one specific purpose. They sought ways to integrate existing tools and easy-to-use new tools that leveraged existing analysis methods.

Finally, our study surfaced a number of recommendations for visual analytics technology developers:

- Externalize the thinking process - Help analysts continuously build a conceptual model

- Support source management - Enable managing both pushed and pulled information and organizing sources meaningfully

- Support analysis with constantly changing information - Integrate collection and analysis in a single system and help analysts use structured methods during collection

- Help analysts create convincing production - Support insight provenance and sanity checks of analytical products

- Support asynchronous collaboration rather than synchronous collaboration for exploratory analysis

3 EXAMPLE SCENARIO EMPLOYING VISUAL ANALYTICS

In order to better demonstrate how visual analytics can aid intelligence analysis, we present an example scenario. "The 9/11 Commission Report" is a publicly available report about the 9/11 terrorist attacks on the World Trade Center in New York. One version of the report is stored as a pdf document with 585 pages. In order to better simulate a larger collection of short intelligence reports, we split this document into 585 pages and consider each as a separate document. We use the page breaks as separators since the report does not have a natural structure that would lend itself to being split into short documents of a few paragraphs.

To illustrate this scenario, we employ the Jigsaw visual analytics system [10]. Jigsaw combines automated text analyses with interactive visualizations for exploring and analyzing collections of unstructured and semi-structured text documents. It automatically identifies entities of interest in the documents, such as people, places, and organizations, and then shows connections between those entities across the entire collection, as well as connections between documents and entities. Connections are defined by co-occurrence: if two entities co-occur in the same document, they are connected to each other as well as to that document. If entities co-occur in many documents they have a stronger connection. Even though this untyped connection model based on co-occurrence is very simple, it has turned out to be a powerful tool for investigative analyses. It works best if the documents are not too large, as it is often the case for news articles or case reports that usually span a few paragraphs.

We present the scenario from the point of view of a hypothetical intelligence analyst who is examining the document collection. To begin, the analyst imports the 585 single-page documents and runs an automatic entity identification. She uses the integrated Open-Calais webservice to identify people, locations, and organizations; the integrated GATE package to identify money entities, and built-in regular expression matching algorithms for identifying date entities. The analyst removes all entities that occur in only one document (they would not contribute to any connections) and performs a basic entity clean-up process, including removing wrongly identified entities and aliasing entities with multiple representations such as "George Bush" and "George W. Bush". The entire process results in a document collection with 369 people, 200 location, 252 organization, 12 money, and 464 date entities across the 585 documents.

The analyst begins the investigation seeking an overview of the entities. She uses the List View to display lists for Location, Person, Organization, and Money and change the list ordering from alphabetic to frequency-of-occurrence to see the most frequent entities in the document collection (Figure 1). The small bar to the left of each entity indicates the number of documents in which it occurs. Entities with aliases are shown in italic font and the aliases are displayed as tool tips, as shown for *Usama Bin Ladin*. The most frequent locations are *United States* (364 occurrences), *Afghanistan* (184), *Pakistan* (98), *New York* (77), and *Saudi Arabia* (71); the most frequent persons are *Bill Clinton* (65), *George W. Bush* (59), *Usama Bin Ladin* (59), *Richard Clarke* (50), and *George Tenet* (36); and the most frequent organizations are *al-Qeda* (233), *Central Intelligence Agency* (214), *Federal Bureau of Investigation* (181), *White House* (93), and *Federal Aviation Administration* (91).

The analyst next selects *Usama Bin Ladin* in the person list and reorders the other lists by strength of connection to the selection in order to see the entities most common with him (Figure 2, left).
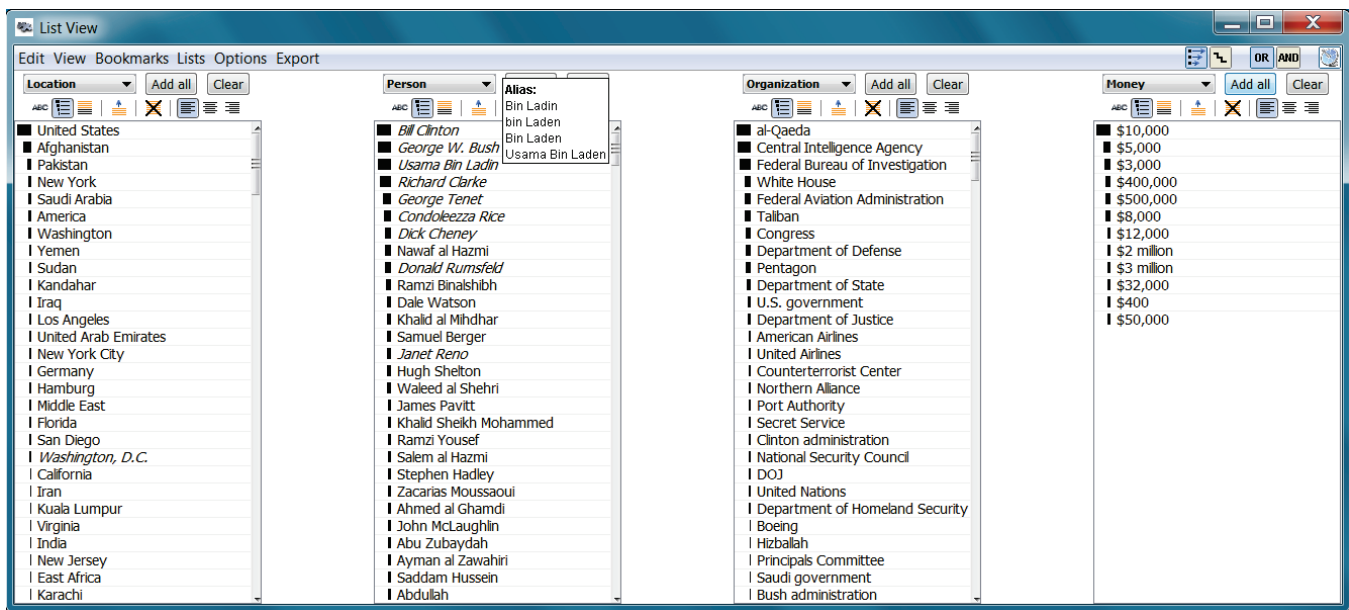
2

Figure 1: List View showing an overview of the 9/11 Commission Report, focusing on the Location, Person, Organization, and Money entities. All entities are sorted by frequency.
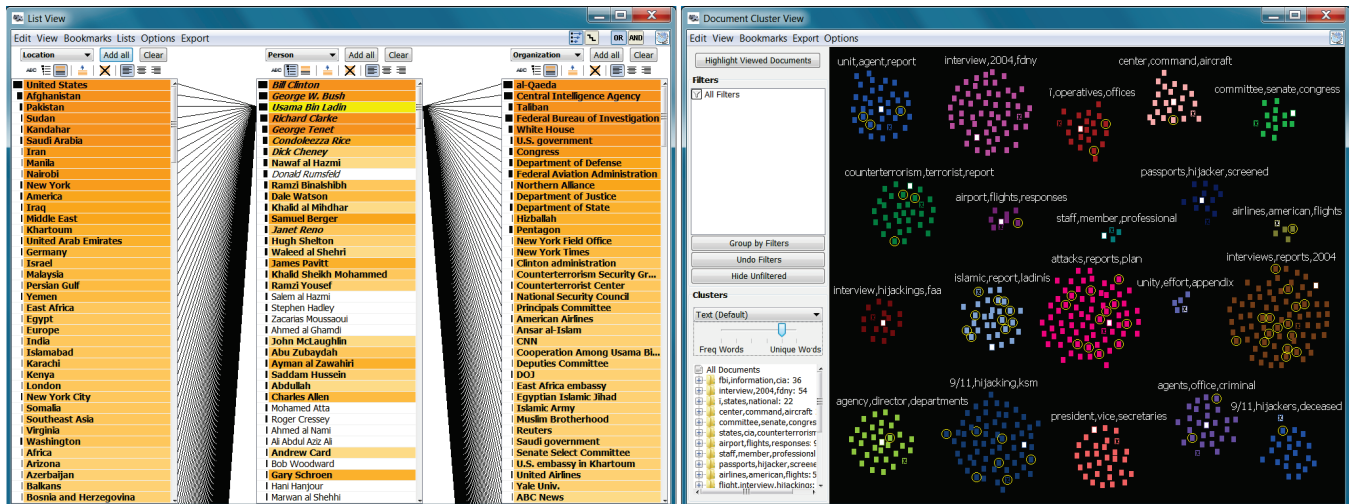

Figure 2: List View (left) showing locations, persons, and organizations connected to *Usama Bin Ladin*. Document Cluster View (right) showing different clusters of related documents (small rectangles in different colors). Documents connected to *Usama Bin Ladin* are selected (surrounded by a yellow circle).

The List View highlights entities connected to a selected entity (yellow) via an orange background. Darker shades of orange indicate or stronger (more frequent) connection. Entities that are not directly connected have a white background. *United States*, *Afghanistan*, *Pakistan*, *Sudan*, and *Kandahar* are the most connected locations to *Usama Bin Ladin*; *al-Qaeda*, *Central Intelligence Agency*, *Taliban*, *Federal Bureau of Investigation*, and *White House* are the most connected organizations. He is also strongly connected to the people *Bill Clinton*, *George W. Bush*, *Richard Clarke*, and *George Tenet*.

To better understand the themes and topics in the report, and in particular those in which *Usama Bin Laden* is mentioned, the analyst opens the Document Cluster View and displays the documents clustered by text similarity (Figure 2, right). Each document is displayed as a small rectangle and each cluster is labeled with three keywords. Text analysis algorithms integrated in Jigsaw automatically compute the clusters and summaries. The cluster summaries represent important topics in the report, including counterterrorism, hijackings, attacks, interviews, and president.

Cross-view selection and filtering are important capabilities in visual analytics systems. Since *Usama Bin Ladin* is still selected in the List View (Figure 2, left), the documents he appears in are also selected in the Document Cluster View (Figure 2, right), indicated by a yellow circle. He is connected to more than ten documents in the "islamic, report, ladin's", "attacks, reports, plan", and "interviews, reports, 2004" clusters and to seven documents in the "9/11, hijacking, ksm" cluster. The analyst also could use cross-view selection in the opposite direction for a different kind of exploration: when she selects the "president, vice, secretaries" cluster, she observes in the List View that *George W. Bush*, *Dick Cheney*, *Condoleezza Rice*, and *Donald Rumsfeld* are the most connected people to that cluster. Interestingly, *Donald Rumsfeld* is not connected to *Usama Bin Ladin* (Figure 2, left).

To learn more about Rumsfeld, the analyst opens the Graph View and explores the people and organizations connected to him using a "circular layout" approach (Figure 3, left). This approach positions the documents that mention *Donald Rumsfeld* on a circle (white

3
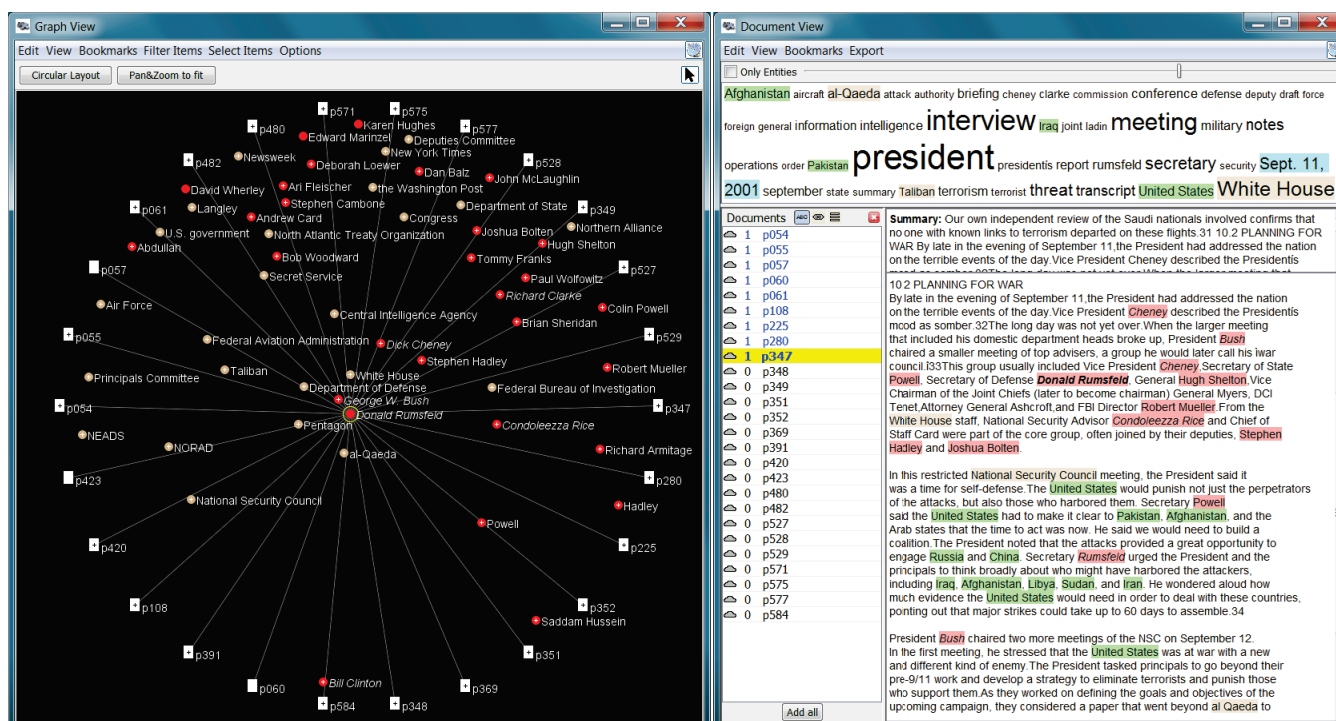
Figure 3: Graph View (left) showing a Circular Layout with documents (white rectangles) that mention *Donald Rumsfeld*. Persons (red circles) and organizations (tan circles) are positioned inside the circle of documents, the more connected they are the closer they are positioned to the center of the circle. Document View (right) showing all documents mentioning *Donald Rumsfeld*. Above the selected document's text (right) is a one sentence summary and below are the affiliated entities. The word cloud (top) summarizes all documents loaded in the view.

rectangles) and the related entities within that circle (red circles for people, tan circles for organizations). More highly connected entities are placed closer to the center of the circle. The layout shows that *Donald Rumsfeld* is strongly connected to *George W. Bush*, *Dick Cheney*, and *Stephen Hadley*, as well as to the organizations *White House*, *Department of Defense*, and *Pentagon*. The Graph View also supports interactive exploration of the connection network via expand and collapse operations. A double click on a document or an entity expands that item and brings in all other items that are connected to it. Items having additional connections that are currently not shown are indicated with a plus sign. A double click on an item that already shows all its connected items (e.g. *Donald Rumsfeld*) collapses that item and hides all its connected items.

The analyst next wants to read the documents about Donald Rumsfeld, so she opens them in the Document View (Figure 3, right). The document list (left) shows the 26 documents (pages) that mention him; documents shown in blue have already been examined, and the number in front of the document indicates how often it was displayed. The word cloud (top) summarizes the currently loaded document set using the most frequent words in those documents. The selected (yellow) document in the list is presented on the right. Above the documented text is a one-sentence summary of the document computed by a text summary analysis. To support quick scanning of documents, entities in the word cloud and in the document itself are highlighted: people in red, locations in green, dates in blue, and organizations in tan. This document (p347) talks about a restricted National Security Meeting on the night of the attacks in which "Rumsfeld urged the President and the principals to think broadly about who might have harbored the attackers, including Iraq, Afghanistan, Libya, Sudan, and Iran." (The numbers in some sentences in the document are footnote references.)

To investigate if similar documents exist in the collection, the analyst opens the Document Grid View and sorts and colors the

documents by their similarity to document p347 (Figure 4, top). The shading of blue indicates document similarity: dark blue indicates similar documents, light blue indicates documents that do not have much in common with the reference document. A tooltip provides the one sentence summary of a document. A similar document p215 mentions "Bonk told Bush that Americans would die from terrorism during the next four years. During the long contest after election day, the CIA set up an office in Crawford to pass intelligence to Bush and some of his key advisors." It seems that there might have been some miscommunication in the post election transition. To understand the role of the former president *Bill Clinton*, the analyst displays his name in the Word Tree View (Figure 4, bottom). A Word Tree [12] shows all occurrences of a word or phrase across all documents in the context of the words that follow it. Each word can be explored further by a click. The Word Tree View for *Bill Clinton* shows that his "administration effectively relied on the CIA to take the lead in preparing long-term offensive plans" and that "One of the great regrets of my presidency is that I didn't get him [Bin Ladin] for you".

SIDEBAR: VISUAL ANALYTICS TOOLS FOR INTELLIGENCE ANALYSIS

A few commercial tools for intelligence analysis employ visual analytics techniques including Analyst's Notebook from IBM i2 (http://www.ibm.com/software/industry/i2software), nSpace from Oculus (http://www.oculusinfo.com/nspace), and Palantir's suite of systems (http://www.palantir.com). An extensive discussion of academic research projects employing visual analytics for understanding text and document collections can be found in [2].

The scenario exploring the 9/11 Commission Report and the images used in this article were produced using the Jigsaw visual analytics system. Jigsaw was designed to help investigators ex-
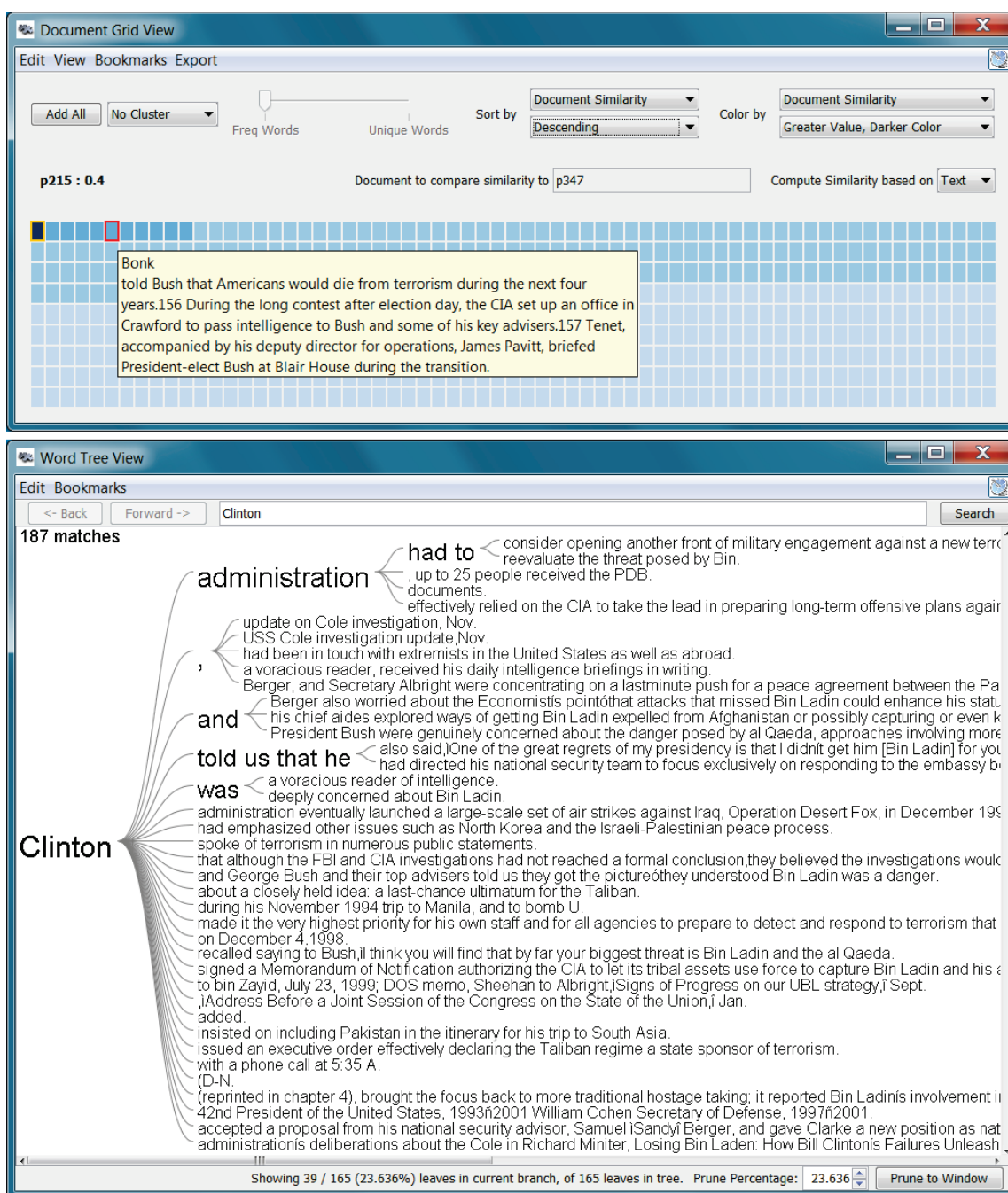
Figure 4: Document Grid View (top) with the document (small rectangle) order and shading set to correspond to the document's similarity to the selected document p347. Word Tree View (bottom) showing occurrences of the person *Clinton* and the most common phrases that follow him in sentences within the 9/11 Commission Report.

plore and understand collections of text documents, and in particular, to follow trails of ideas embedded across the documents. The system's name comes from the notion of "putting the pieces together." Early versions of Jigsaw emphasized a suite of interactive visualizations portraying the documents' contents and connections between entities in the documents [10]. More recently, we have integrated computational text analysis capabilities [2]. Beyond intelligence analysis, Jigsaw has been used to explore consumer review, academic research, fraud, investigative reporting, law enforcement, business intelligence, and email document collections. The Jigsaw system itself, as well as example datasets (including the one used in this article), tutorials, videos, and re-

lated articles are available for download on the webpage (`http://www.cc.gatech.edu/gvu/ii/jigsaw`).

4 EVALUATING SYSTEMS FOR INTELLIGENCE ANALYSIS

The evaluation of visual analytics systems is a challenging research area in itself: there is no consensus among researchers on how to effectively and objectively measure the contributions of a system to an analyst's generation of insights. This is especially true when intelligence analysis is the domain being studied. In our research, we have used multiple techniques to evaluate the effectiveness of the systems we have built.

In one evaluation study [5], we employed a small synthetic dataset with embedded ground truth, consisting of 50 imaginary short reports with a hidden threat. We then recruited sixteen students, divided them into four groups, and asked them to conduct an analysis of the documents and identify the hidden threat. Participants in the first group only worked with pencil and paper. They received a printout of all the reports and some blank sheets for note taking. Participants in the second group received an electronic copy of the reports and could use basic text editing software for reading and searching the documents. Participants in the third group used only the Document View of the Jigsaw system to read and analyze the document collection. This setup was similar to the previous one, providing functionality for reading and searching; however, the Document View also highlighted identified entities within the documents. Participants in the fourth group used the entire suite of visualizations in Jigsaw to conduct the analysis.

The study participants worked with the documents for 90 minutes and then wrote debriefing statements, which we compared to the ground truth and then graded for accuracy. We also conducted follow-up interviews and collected their notes. Additionally, we videotaped all the sessions and used screen capture software in the settings where participants worked on a computer. We started our analysis with an inductive approach to examine the qualitative data in order to unveil potential concepts and themes and to understand the influence of the tools in the different settings. At a later stage of our analysis we combined inductive and deductive approaches and supplemented them with observations from the video logs, screen captures, and other quantifiable data.

We found that overall the participants using the full Jigsaw system outperformed all other groups on average. Because of the small subject population this result was not statistically significant, however. We did observe four particular strategies that participants employed in their investigations. These strategies ranged from first reading all the documents very carefully to finding an initial clue and following a trail from it. The participants using Jigsaw applied three out of the four strategies and performed well using any of those strategies.

We also used Jigsaw for our own investigations and participated in a number of IEEE VAST Conference Challenges and Contests over the past few years. These contests provide synthetic document collections with embedded ground truth. Participating teams are tasked with finding a hidden threat in the documents. Working with Jigsaw on the contest datasets helped us to gain practical experience in these types of intelligence investigations, improve the system, and develop additional functionality. We further describe the influence of our participation in the VAST contests on the design and development of the Jigsaw system in another article [3].

Because we have made Jigsaw available to anyone to use in their own work, examining real world use by other people is another type of evaluation that we have employed. In order to better understand how professional analysts have been using the system and to determine its benefits and limitations in practice, we interviewed six investigators who had been using Jigsaw for an extended period of time [7]. These individuals included an aerospace engineering researcher, a business analyst investigating fraud, a doctoral candidate in Industrial and Systems Engineering studying enterprise transformation, and intelligence analysts at a national lab, the Air Force, and a police department. The goal of this study was to evaluate whether Jigsaw is helping analysts with their tasks, to understand its application to different types of documents and domains, and to identify useful features and capabilities of the system as well as missing or problematic features.

We identified a number of applications of the system across more than one participant. Many used Jigsaw to find connections and relationships between entities, one of the core goals of the system. Some used it as a search and comparison tool to more conveniently

work with text documents, and many used it to gain a broader understanding or overview of their documents. Surprisingly to us, some of the participants also used Jigsaw as a communication aid to share their understanding with others. We originally created the system as an analysis tool and that application is always how we have thought of it. It was interesting to note that some of the study participants also were using it to present findings and tell a story to their colleagues.

The investigators in the study identified a number of limitations and issues with the system as well. Some of the participants wanted better ways to work with only subsets of their document collections. They wanted to be able to dynamically filter out documents in an investigation, but also maintain the ability to reintroduce filtered documents as desired. Document import was another particular challenge and often required manipulating and translating their original documents into a form that Jigsaw could better analyze. Furthermore, problems that arose in document import or in any other use of the system raised questions in the investigators' and their colleagues' minds about the accuracy of the system. They commented how any kind of issue or usability problem eroded their trust.

Our study identified a number of future objectives for Jigsaw and other visual analytics systems for document analysis. The investigators all believed that entity identification is crucial and they wanted easier and more reliable mechanisms to perform it and correct/modify it. They also sought to have more flexible mechanisms for document management activities such as import, storing, filtering, and maintaining. A number of the users wanted more quantitative and statistical analysis capabilities. For instance, they expressed a desire for more network analysis and modeling metrics. In terms of the user interface, some of the investigators wanted to be able to annotate the system views, highlight particular items, and add notes and comments on top of the visual representations.

5 CONCLUSION

Intelligence analysis requires people and organizations to review and assess large collections of information in order to better understand current situations and take the appropriate next steps. The sheer scale, diversity, and complexity of the information to be explored often makes such analysis cognitively demanding. Furthermore, the information is often recorded as narrative text, not quantitative data, and thus it is not as amenable to automated analysis techniques.

Visual analytics technologies that combine computational text analysis with interactive visualization provide a powerful new paradigm for helping intelligence analysts in their work. While current visual analytics systems have illustrated the potential of the field, many challenges remain. Visual analytics systems must scale to increasingly larger collections of data in order to keep up with our growing ability to log and record information. Additionally, visual analytics systems should assist investigators in the complex processes of analytical reasoning, hypothesis formulation, and decision making.

6 ACKNOWLEDGEMENTS

REFERENCES

[1] Central Intelligence Agency. *A Consumer's Guide to Intelligence*. Diane Pub Co, 1999.

6

[2] C. Görg, Z. Liu, J. Kihm, J. Choo, P. Haesun, and J. Stasko. Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, 2013. to appear.

[3] C. Görg, Z. Liu, and J. Stasko. Reflections on the Evolution of the Jigsaw Visual Analytics System. *Information Visualization*, 2013. to appear.

[4] R. Heuer. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency, 1999.

[5] Y.-a. Kang, C. Görg, and J. Stasko. How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 17(5):570–583, May 2011.

[6] Y.-a. Kang and J. Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *IEEE VAST*, pages 21–30, Oct. 2011.

[7] Y.-a. Kang and J. Stasko. Examining the use of a visual analytics system for sensemaking tasks: Case studies with domain experts. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2869–

2878, 2012.

[8] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. *Information Visualization: Human-Centered Issues and Perspectives*, pages 154–175, 2008.

[9] P. Pirolli and S. Card. Sensemaking Processes of Intelligence Analysts and Possible Leverage Points as Identified Through Cognitive Task Analysis. In *International Conference on Intelligence Analysis*, May 2005.

[10] J. Stasko, C. Görg, and Z. Liu. Jigsaw: Supporting Investigative Analysis through Interactive Visualization. *Information Visualization*, 7(2):118–132, 2008.

[11] J. J. Thomas and K. A. Cook. *Illuminating the Path*. IEEE Computer Society, 2005.

[12] M. Wattenberg and F. B. Viégas. The Word Tree, an Interactive Visual Concordance. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1221–1228, 2008.

Morgan State University

VACCINE Annual Report – Year 4

iLaw Enforcement App Assistance Program for Students (iLEAPS)

Supporting Documentation

Surveys

SESSION 1

I. Process

iLeaps team arranged for several cross disciplinary students to participate in the test procedure a week before the test event. The day prior to the test event, several participants reported that they would not be able to make the test. The iLeaps team quickly arranged for alternative student participants, but had to limit them to those from the School of Engineering.

II. Procedure

The participants were trained on the part of the software that they were testing (User/Officer/Dispatch). During the test iLeaps team members assisted the participants that needed further instruction.

III. How Statements Were Recorded

All participants were surveyed using survey monkey

IV. Orientation

Participants were trained on the parts of the software that they were testing, and given a quick explanation on how the pieces work together

V. What Went Wrong

The process for choosing participants needs to be improved. This would have given the iLEAPS team more time for orientation and more time for the participant to answer surveys directly after testing.

VI. What Went Right

The software worked the way it had been tested previously in the lab

VII. Analysis of Survey Results

The participants were asked to rate the following questions on a scale of 1-5, where 1 is the least desirable and 5 is the most:

Question	1	2	3	4	5
Rate ease of use	0%	0%	0%	33.3%	66.7%
Rate the functionality	0%	0%	0%	50%	50%
Rate the overall design	0%	0%	16.7%	16.7%	66.7%
Rate the ease of naviagation	0%	0%	0%	33.3%	66.7%
Rate the likelihood of using such an application in emergency situations	0%	0%	16.7%	0%	83.3%

The participants were also asked to indicate (1) any general comments that they had, (2) any behaviors/functionality that needed to be modified, and (3) any features they would like to see incorporated in to the system. A synopsis of the responses is as follows:

- Do not require user to refresh content manually

- Add an option to take a picture of an incident

- Tutorial video for first time users

- Minimize confusion when selecting an incident to report by allowing users to see all incidents before making a choice

SESSION 2

I. Process
The iLEAPS team recruited 10 cross discipline undergraduate students from Morgan State University and 10 incoming freshman to participate in testing the iLEAPS software. We contacted participants a week in advance and verified via text the day before testing. 7 out of the 10 undergraduate students were in attendance for the testing and 10 out of 10 incoming freshman were in attendance.

II. Procedure
All students were given an overview of the apps. Each participant was trained on the part of the software they were testing. Participants still needed further instructions on using the application as they interacted with all its features.

III. How Statements Were Recorded
All participants were surveyed using survey monkey

IV. Orientation
All Participants were given a brief overview of all three components; they were then broken into groups of three and given a more detailed description on how the software operates. Each group consisted of one dispatch, one officer and one user.

V. What Went Wrong
The orientation took too long and was inconvenient for participants.

VI. What Went Right
The software worked the way it had been tested previously in the lab

VII.   Comparison with Previous Test
Our first test had only 10 undergraduate participants in total. They were not formally instructed on the nature of the program and what role they played in testing its capabilities. The 10 students were interchanged throughout the testing period due to schedule conflicts, which caused confusion and many opportunities for misinformation. All of the students involved provided vague responses to using the software. During the second test, we had nearly 20 participants. We recruited 10 new undergraduates and 10 incoming freshman. All students were briefed before and on the day of testing, on what they would be doing. Students were given a demonstration of the function of the software. Each student actively engaged with instructors to gain further knowledge of the software as well as verbally contribute suggestions on how to improve it. The feedback that was received was much greater and more detailed than from our first test.

VIII.   Survey Results
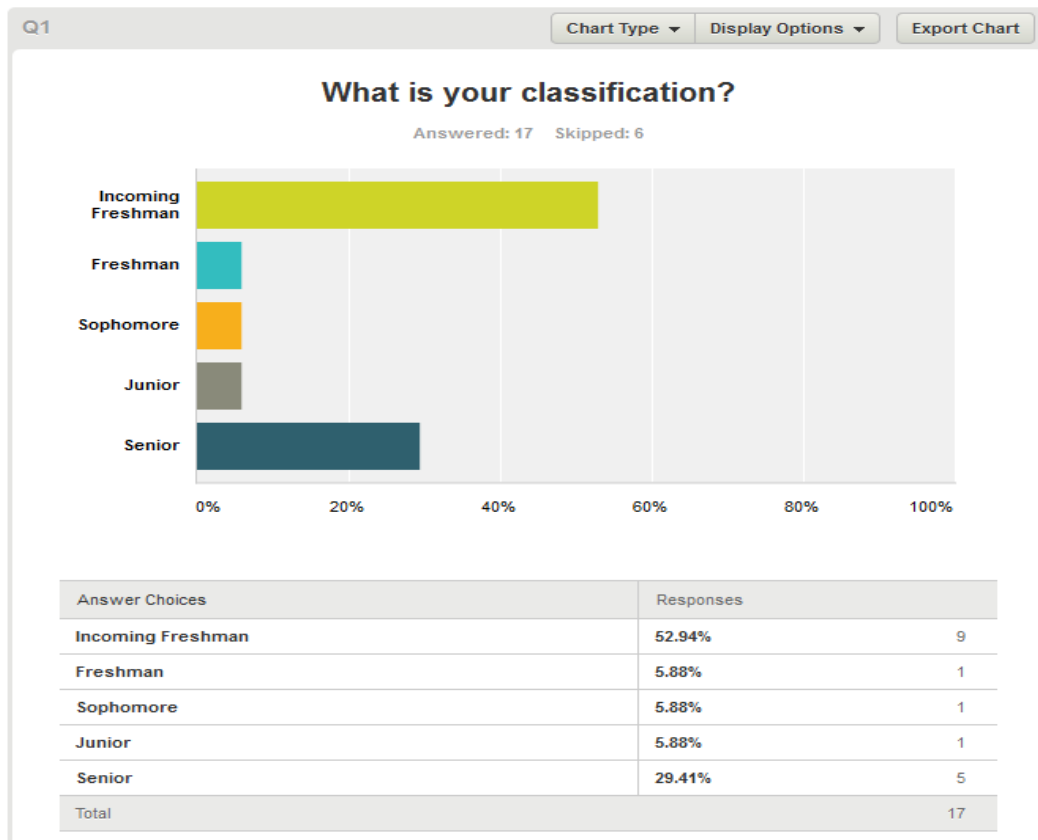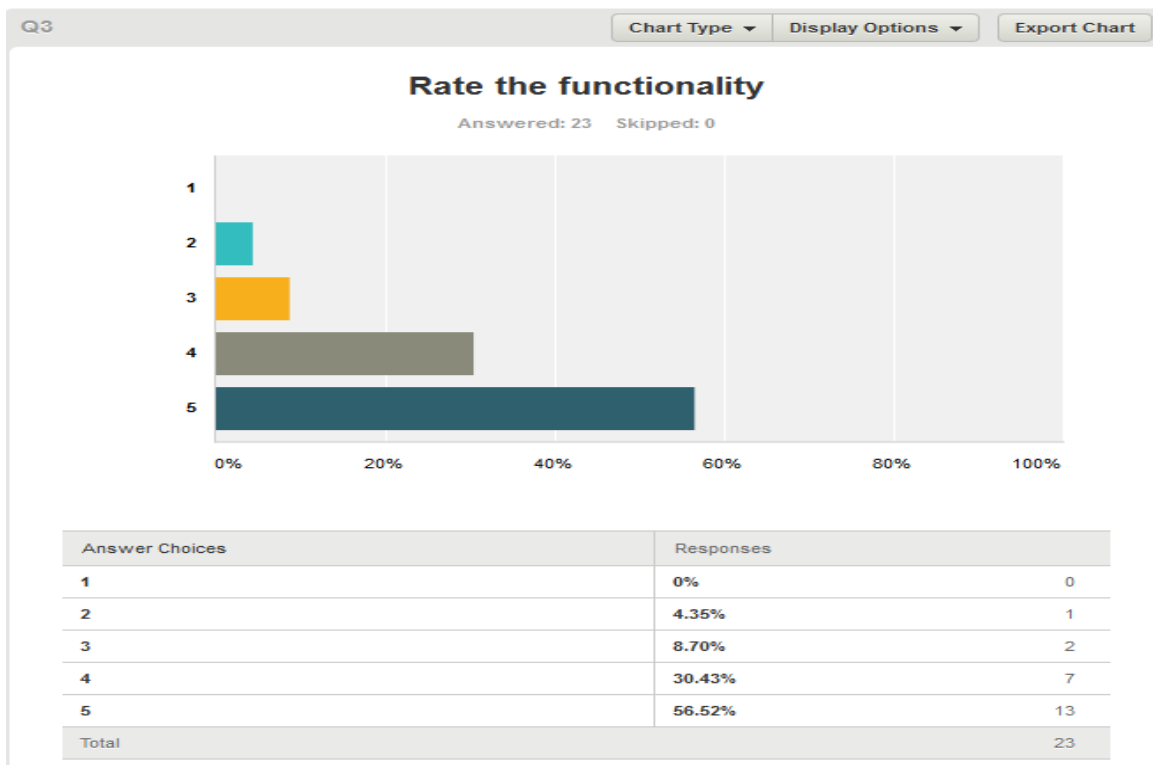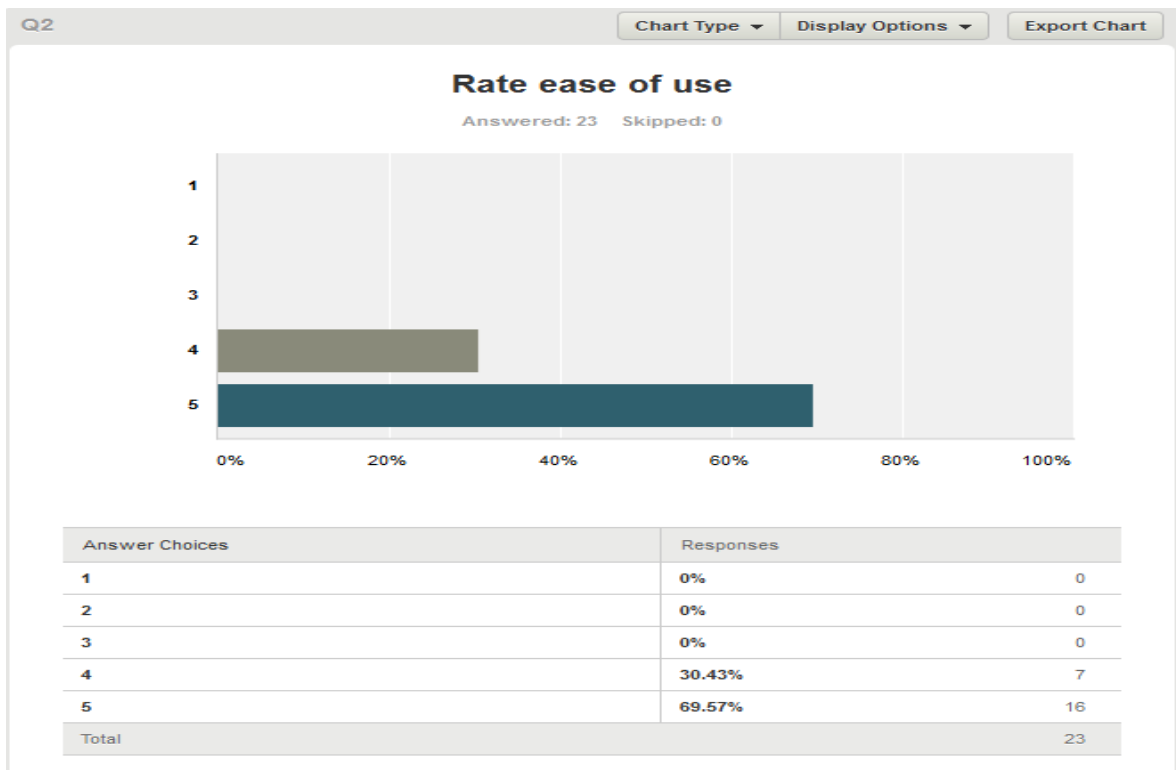
| Q1 | | Chart Type ▾ | Display Options ▾ | Export Chart |

What is your classification?

Answered: 17   Skipped: 6

Answer Choices	Responses	
Incoming Freshman	52.94%	9
Freshman	5.88%	1
Sophomore	5.88%	1
Junior	5.88%	1
Senior	29.41%	5
Total		17

Q2

Rate ease of use

Answered: 23   Skipped: 0

Answer Choices	Responses	
1	0%	0
2	0%	0
3	0%	0
4	30.43%	7
5	69.57%	16
Total		23

Q3

Rate the functionality

Answered: 23   Skipped: 0

Answer Choices	Responses	
1	0%	0
2	4.35%	1
3	8.70%	2
4	30.43%	7
5	56.52%	13
Total		23

Rate the ease of the navigation

Answered: 23   Skipped: 0
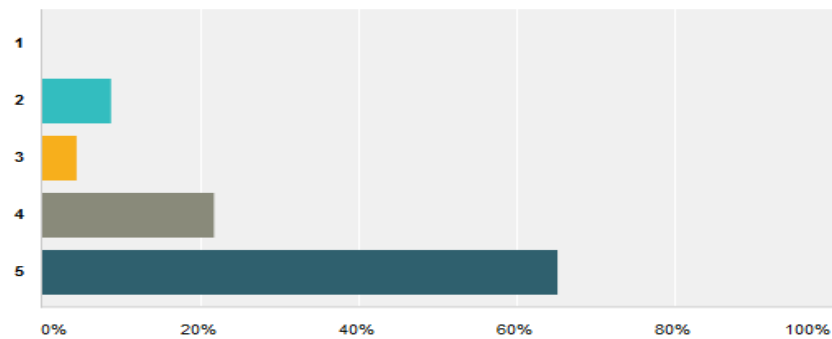
Answer Choices	Responses	
1	0%	0
2	8.70%	2
3	4.35%	1
4	21.74%	5
5	65.22%	15
Total		23

Rate the likelihood of using such an application in emergency situations

Answered: 23   Skipped: 0

Answer Choices	Responses	
1	0%	0
2	8.70%	2
3	13.04%	3
4	26.09%	6
5	52.17%	12
Total		23

Rate the likelihood of using such an application in non-emergency situations

Answered: 23   Skipped: 0
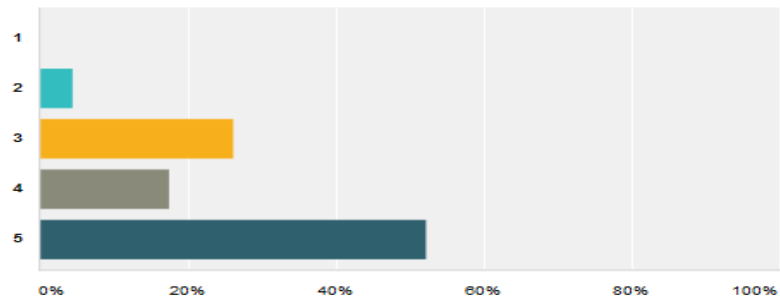
Answer Choices	Responses	
1	0%	0
2	4.35%	1
3	26.09%	6
4	17.39%	4
5	52.17%	12
Total		23

Rate the overall design

Answered: 23   Skipped: 0
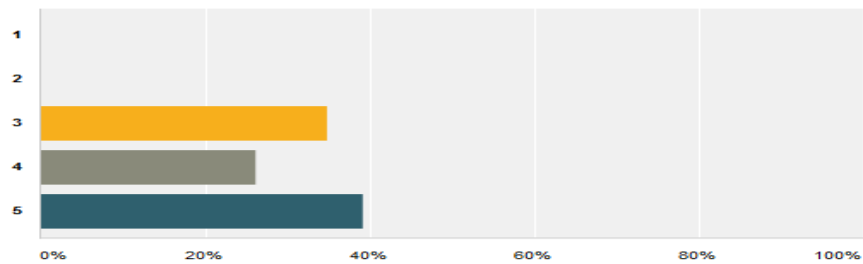
Answer Choices	Responses	
1	0%	0
2	0%	0
3	34.78%	8
4	26.09%	6
5	39.13%	9
Total		23

Pennsylvania State University

Designing a Web Service to Geo-Locate Subjects of Volunteered, Textual Geographic Information

Ryan Mullins[1], Frank Hardisty[1], Scott Pezanowski[1], Sujatha Das[2], Alexander Savelyev[1], Alan MacEachren[1], Prasenjit Mitra[3], Anuj Jaiswal[4]

[1]The Pennsylvania State University Department of Geography, 302 Walker Building, University Park, PA 16802, United States
[2]The Pennsylvania State University Department of Computer Science and Engineering, 313F Information Science and Technology Building, University Park, PA 16802, United States
[3]The Pennsylvania State University College of Information Science and Technology, 313F Information Science and Technology Building, University Park, PA 16802, United States
[4]Reunify, LLC. Record Linkage Research, 2043 Colorado Avenue, Suite #3, Santa Monica, CA 90404, United States

In recent years, the amount of publicly available spatial, or spatially-enable, data has grown tremendously, due in large part to the proliferation of GPS-enabled technologies in mobile devices and in-car navigation systems, and from the location information integrated into web applications, especially social networking services. Networks like Twitter, Four-Square, Facebook, and others allow users to provide insights into current events in real time via short form textual updates or statuses. Parallel to the availability of this type of citizen-produced data, there has been a growing interest in analyzing this data to examine sentiment or track how information disperses through networks. Many modern social networks provide a means to locate the contributor of status updates. The location of a contributor is typically given as geographic coordinates, latitude and longitude, that is accurate to the position provided by the web-enabled device used to submit the status update. This spatial information, along with the temporal information inherent to status updates, enables spatial and temporal analysis of contributor patterns. However, although some updates include information on the location of a contributor, little capacity is provided for geographically locating the subject, or subjects, that contributors are referencing.

In this paper we describe a web service, in development at the Pennsylvania State University, that enables the geolocation of people, places, and events described in common status updates from online social networks. We describe the use of techniques from a wide array of research areas – applied linguistics, natural language processing, search engine optimization, and geographic information science – to parse out people, places, and events explicitly or implicitly mentioned in status updates, and then analyze and contextualize these entities to locate them in geographic space. Finally, we outline how this service can be integrated into the development of dynamic, map-based, visual analytical interfaces, specifically in the context of crisis management and emergency response.

Keywords: geocoding, geospatial web services, natural language analysis, volunteered geographic information

Designing Map Symbols for Mobile Devices: Challenges, Best Practices, and the Utilization of Skeuomorphism

Joshua E. Stevens, Anthony C. Robinson, Alan M. MacEachren

GeoVISTA Center, Department of Geography, The Pennsylvania State University

Abstract. In this work we make three contributions to the design and use of map symbology on mobile devices. First, we present an overview of the current state of mobile symbology and best practices based on previous empirical findings. Second, we demonstrate the design of a new set of map symbols for mobile devices based on these guidelines and proposed design strategies. These new symbols were developed for comparison with an existing standard used by emergency management and disaster relief professionals. Lastly, we discuss the role of skeuomorphism in the context of affording interaction in map symbol design. We believe this work advances the science of mobile symbology and demonstrates a practical application of skeuomorphic design in modern mapping applications.

Keywords: Design, Symbology, Interaction, Skeuomorphism

1. Introduction

A number of recent technologies and tools have demonstrated the utility of mobile mapping devices in a wide range of settings. The potential utility of mobile maps is particularly clear in mission-critical environments such as law enforcement or emergency response and recovery. Within such settings, mobile devices must not only be reliable in both form and function, they must have utility to the mission. This requirement places particular emphasis on the visual interface of the device, as it connects the user to the application where information must be interpreted and acted upon quickly. With a growing number of location-based applications and support tools being

developed for these purposes, great care must be taken in the way events are symbolized.

In addition to mission-critical usage by professional users, mobile maps have permeated consumer environments and are ubiquitously employed to aid navigation, enhance tourism, and enable a wide array of location-based services that range from tracking a lost phone to local dating services (Lee, Zhu and Hu 2005). Whether for profession or pleasure, the success of many mobile maps relies on interactive functionality that transforms geographic data into actionable information.

At the same time, most maps are still largely static; base maps rarely need updating, key points of interest (POI) tend to reflect a constant location and information (e.g., a name and address), and many elements of cartographic design are typically not interactive (e.g., highway labels, POI). For these reason, we argue that an effective mobile map is one that successfully distinguishes interactive and non-interactive symbology.

This paper reviews research on mobile mapping with an emphasis on map symbology for and interaction with mobile maps. Based on the review, we elucidate some of the design challenges and considerations required for providing meaningful maps on a range of mobile devices reliant on interactive symbology. Emphasis is given to applications in emergency management and law enforcement where decisions must be made quickly and accurately, though we extend our findings to the broader range of mobile map uses. We conclude the paper with preliminary guidelines derived from the literature for design of mobile maps and then outline some key research and development challenges focused on leveraging technological advances to achieve effective mobile mapping applications.

1.1. Mobile Devices: Scope, Limitations, and Considerations

Mobile devices are considered here to be ones that are essentially portable and can be used remotely. These include laptop computers, smartphones, tablet computers, wearable devices, Personal Digital Assistants (PDAs), and in-vehicle units that include but are not limited to navigation systems. For the purpose of this paper, we focus on the subset of mobile devices that are held in or used by one's hands while mobile. This focus excludes laptop computers as well as wearable devices used primarily for augmented reality applications. Accordingly, the remainder of this article deals primarily with smartphones and tablet computers.

Mobile devices defined in this way have many potential benefits, including: collecting information in the field or during a police patrol (Bragdon et al. 2011), obtaining live updates remotely when in the field (Weakliam et al.

2005), communicating with natural/manmade disaster first responders (Erharuyi and Fairbairn 2003, MacEachren et al. 2006), and visual collaboration among large teams during urban emergencies (Monares et al. 2009).

Despite the advantages suggested above, mobile devices are not without their limitations. Most notable of these are reduced screen size and display resolution (Burigat and Chittaro 2011), reduced capability for input and interaction while the user is in motion (Bragdon et al. 2011), and limited processing power and memory (Follin and Bouju 2008). Each of these limitations influences how symbology can be applied and whether or not a particular approach is effective. These limitations can often be compounding. For example, limited screen space will dictate limits on the number and size of symbols that can be displayed legibly on the screen and the extent of territory that can be displayed. A desktop environment, in addition to supporting display of larger territories and more symbols can alleviate the constraints it does have through zooming, panning, and other interactions. However, the mobile utility of smaller devices is hamstrung by the increased effort similar interactive behaviors require when using the device in concurrence with another common task, like walking or driving. If this increase in effort to interact and the attention that must be directed to that interaction is too onerous, it can impede the device's advantages as a mobile aid (Willis et al. 2009).

A number of attempts have been made to understand and combat these limitations. The remainder of this paper will discuss several of these efforts that are most relevant to mobile mapping with an emphasis on issues of map symbology and interaction with the map display for mobile devices. Through this discussion, the latest directions in the literature and key challenges that remain for future research are also identified.

The existing literature on mobile map symbology is diverse and addresses many specific goals and interests. It is therefore useful to identify common themes that are prevalent. We have identified three themes that appear within the broader context of map symbology and interface design for mobile devices; specifically, the literature can be categorized as emphasizing:

1. **Symbology.** This work addresses use-specific symbol design, the salience of the symbols due to display parameters, figure-ground relationships, semantics, and the performance of symbols in task-driven evaluations.
2. **Interaction.** Research in this area emphasizes how users interact with devices, operate menus, allocate attention, and respond to displays that use symbols as part of a greater suite of functions. Examples include graphical hints that alert the user to symbols existing outside of the currently active view, symbols that aggregate when

they are too numerous for the screen, and the role of touch, gestures, and buttons to manipulate the display.

3. **Remote information access and collaboration.** Mobile devices are frequently used to report, update, collect, and communicate information remotely. A growing body of literature is centered on the use of map symbology and interface design to facilitate these collaborative tasks

There is considerable overlap between these themes and they are by no means absolute. We continue in the following sections with key lessons learned, emphasizing those most relevant to items 1 and 2: symbology and interaction.

2. Previous Research: Symbol Design and User Experience

Map symbols have been categorized in multiple ways and there is a wide range of terminology used to discuss map symbology in the literature. To create consistency across the studies that will be cited below, this paper will use the terminology described by MacEachren (1995), focusing on the relative abstractness or iconicity of symbols as it related to three categories of positional symbols that each exhibit a range on this continuum: pictorial, associative, and geometric.

Many mobile devices in current production utilize touch-screen displays, thus it is crucial that new symbol designs consider existing work in this area. Morrison and Forrest (1995) conducted one of the earliest studies evaluating pictorial symbols on touch-screen devices within the context of tourist maps. Their work highlights the need to consider design not only from the standpoint of variables affecting individual symbols (e.g., size and hue), but also semantic relationships across and between multiple symbols. For example, their results show that for many symbols, size does not influence the accuracy of visual search tasks but may greatly affect how quickly symbols are found. This relationship is moderated by the semantic context suggested by other symbols on the map. A telephone symbol for example may be interpreted as the location of a pay phone when used in isolation or in tourism maps, while the same symbol might be interpreted as a service for calling help if nearby symbols reflect first-aid and medical care. In other words, how users interpret a symbol is influenced as much by context and nearby symbols as by the symbol's own design.

In addition to suggesting semantic context, the design of nearby symbols can also influence the effectiveness of individual symbols by affecting the

overall salience of a particular symbol. Kuo-Chen refers to this as complexity contrast and it greatly influences the time required to identify symbols (2008). Related work suggests that associative symbols, such as the simple monochrome pictorial symbols that are typically part of standards-compliant recommendations, are not as strongly affected by changes in size as are realistic, multi-colored, sketch-based, or 3-dimensional pictorial symbols (Elias and Paelke 2008). Although this limitation is important to consider for map symbols in general, it is especially important for mobile devices with limited screen space. Such a limitation encourages the use of simple, abstract symbols over complex symbols with more detail and realism (Lee, Forlizzi and Hudson 2008).

While the interpretation of abstract symbols is less affected by changes in symbol size, abstract symbols (particularly geometric symbols) are subject to misidentification since the relationship between the symbol and what it refers to is often arbitrary. A trade-off therefore exists between accommodating a limited screen size through smaller, abstract symbols and maintaining semantic clarity.

Isolating symbology from the total experience of using a mobile device is at first an attractive idea since it removes complexities of human-computer interaction and a potentially limitless variety of application contexts. However, the appropriate design of symbols for mobile devices requires a complete understanding of how existing symbols are used, which symbol design traits and interactions between the devices and symbols improve or impede performance, and under which conditions or scenarios certain symbol types or designs are most useful. The factors influencing user experience are subject to varying abilities of individual users and an ever-changing range of device capabilities (Baus, Cheverst and Kray 2005, Meng 2005).

On par with the contentions of the previously mentioned work, Apple asks designers of their mobile applications to "embrace simplicity," (2011, p. 152-153). In addition to generic advice, the human interface guidelines provided by Apple impose specific requirements for size and quality of icon designs to ensure that the designs are effective visually and tactically. The requirements derive from limitations on the user's ability to see a symbol and the device's ability to recognize the user's fingertip when touched. Similar recommendations are provided by the interface guidelines from Google, which offers less specific advice but again reiterates the need to avoid complex, highly detailed, and realistic icons (2011). Additionally, both firms insist that designers consider icon design in the full context of the other elements of the interface and the purpose for which they will be used.

3. Design Recommendations and Suggested Guidelines

Together, the studies reviewed allow us to make the following suggestions:

1. Well-designed symbols should utilize black and white figure-ground relationships or be based on mixed colors that have established meanings, such as the U.S. interstate shield's use of red and blue, and have high contrast against the base map. Since the majority of vector base maps are by default light in color, this suggests that symbols for such maps should maximize figure-ground with a dark frame and light symbol.

2. Symbols with strong semantic relationships with their referent can be identified more quickly on smaller screens than those that are arbitrary and geometric, sketch-based, or based on a 3-dimensional rendition of the referent. However, increasing levels of abstraction may impede the accuracy of symbol identification, and even readily identifiable symbols can be misidentified if their purpose is ambiguous (e.g., the telephone symbol in Morrison and Forrest's research (1995)). Thus it is important to balance the *speed* with which users identify semantically strong pictorial symbols and the *accuracy* with which simpler abstract symbols can be located on the map.

3. Symbols should be smaller when displayed in large numbers - potentially removing the symbol frame if necessary as long as figure-ground is maintained and the symbols remain touchable (for symbols that require interactivity).

4. Symbols intended for concurrent use should be similar in complexity to avoid a large contrast gradient in symbol design, except in the case where greater salience is desired for particular symbols (e.g., a hospital symbol). High complexity contrast between regular symbols and those deemed important would then be preferable (e.g., a hospital symbol made with an "H" will stand out amongst other symbols made from simple geometric shapes, like squares and circles).

Three additional factors that should be taken into account by symbol designers are:

1. ***The capabilities of the target device(s).*** Touch-screen devices have additional symbol size requirements not present on devices that use other input methods. If the symbol requires tactile interaction, it must be large enough to be touched by a fingertip. This complicates the task of accommodating small screens and avoiding clutter.

2. ***The purpose of the device or application should influence the method of interaction***. Symbols used alongside other tasks that are cognitively engaging (like driving) should require as little interaction as possible. If interaction is required, it should be in the form of gestures, spoken commands, or make use of hard buttons that do not require the user to look at the screen.

3. ***Symbols that can be interacted with should be visually distinguishable from those that cannot***. How this is achieved will depend on the other variables employed in the symbol design. A bold frame may alert the user that a symbol is an aggregate and can be clicked for more information, however this approach would be less effective if symbol frames were employed for other purposes (e.g., event status or the degree of damage due to a disaster).

There is a complex relationship between each of these design decisions. It is unlikely that a symbol will be optimal for every setting, to every user, on every device. The application developer, cartographer, or designer must weigh the costs against the benefits and evaluate the performance of their symbology whenever possible.

3.1. Symbol Shape and the Map Symbol Box Model

In addition to the position of icon shape on the abstract-pictorial continuum, the shape of the frame surrounding the map icon is a vital consideration. The shape of the frame dictates how much space exists for the icon, and thus the actual size of the interpretable icon depends on the symbol frame's overall shape and border thickness (Figure 1, right). In general, symbols with square frames (or rounded rectangles as in Figure 1) provide more internal space than other frame shapes to maximize the icon size, which is beneficial in visual search tasks (Morrison and Forrest 1995). Moreover, the additional space afforded by maximizing the space around an icon can be utilized for other cues, like indicating interactivity as discussed in 3.2.

To aid in design decisions and the specification of design variables, we propose the *map symbol box model* (Figure 1, left). Similar to the box model specified for cascading style sheets (CSS) for HTML documents[1], the map symbol box model clarifies the foundational elements upon which a symbol is constructed. Defined in this way, individual design variables can be isolated and discussed explicitly, greatly enhancing the design process for new symbols and allowing precise critique of existing symbology.

---

[1] http://www.w3schools.com/css/css_boxmodel.asp
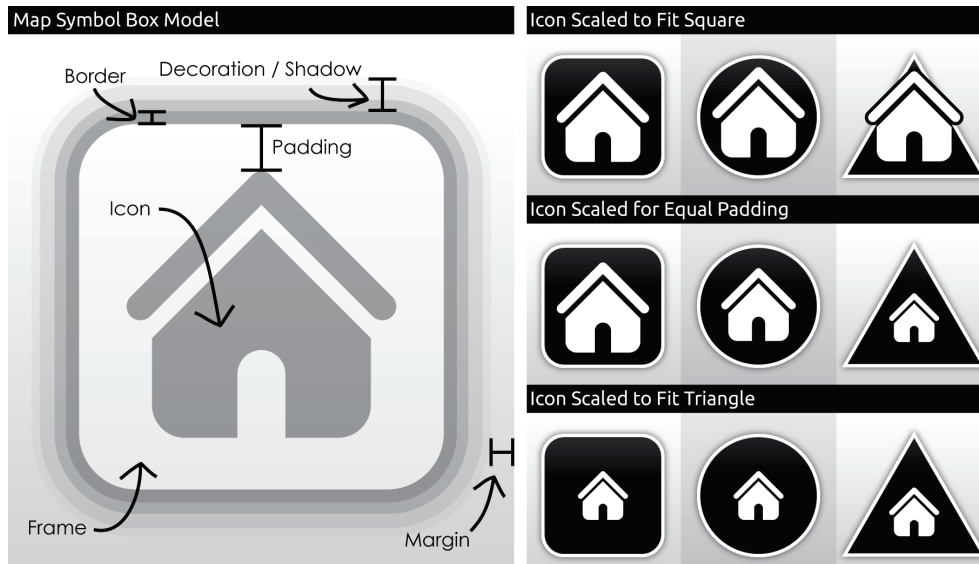
Figure 1. The proposed map symbol box model (left) and an example of specifying symbol size relationships using the padding element of this model (right).

By deconstructing map symbols into 'boxes within boxes,' each with a specific attribute, the design process can more accurately reflect the relationships between the various features of a complete map symbol.

3.2. Interactive Cues via Skeuomorphism

Skeuomorphs, typically defined as design elements that reflect ornamental references to previous (and potentially obsolete) analogs, are ubiquitous in current user interfaces. The uses of skeuomorphism in digital interfaces are typically aesthetic, such as realistic textures (leather and stitching) used in the design of mobile applications. Skeuomorphism can also be used to suggest semantic relationships. For example, the 'save file' feature in most software applications is represented by a 3.5" floppy disk, even though these disks are rarely used and new computers no longer come with the ability to read floppy disks. Despite a lack of academic literature on skeuomorphic interface designs and criticism for their form-over-function nature, we contend that when used sparingly, skeuomorphic designs have the potential to enhance the performance of map symbols.

We hypothesize that an effective use of skeuomorphism is to indicate that a map symbol is clickable (or touch-able). By evoking a heuristic response akin to what we experience in the real world, map symbols – like elevator buttons and doorbells – can indicate interactivity by appearing to have a touchable surface different than their surroundings.

Figure 2 provides an example of a skeuomorphic interactive cue:

Figure 2. A subtle, 3-dimensional knurled skeuomorphic cue gives the interactive symbol (right) the appearance of being touchable.

4.   GeoVISTA Symbology and Future Work

Within a project supported by the Department of Homeland Security, the Penn State GeoVISTA Center designed mobile symbology as a possible alternative to the existing Homeland Security Working Group (HSWG) symbology[2]. The GeoVISTA symbology (Figure 3) was designed in compliance with the guidelines presented in this paper. In follow up research, we plan to compare the symbol sets in a user study with several tasks completed on a mobile device.

Figure 3. The DHS HSWG symbology (left) and their redesigned GeoVISTA counterparts (right). The GeoVISTA symbols also have an interactive, skeuomorphic version using the cue in Figure 4.

Notable design decisions that distinguish the GeoVISTA symbology from the HSWG set are outlined below.

---

[2] http://www.fgdc.gov/HSWG/index.html

1. The GeoVISTA symbols are arranged in a square frame. This provides space within the symbol frame to increase the size, and therefore legibility, of the icon.
2. To have high figure-ground contrast with the greatest number of base maps, which are typically light in color, the GeoVISTA symbology uses a light icon within a dark frame.
3. The GeoVISTA symbology is intended to have a strong semantic relationship with the events or places being depicted.
4. The GeoVISTA symbology is designed as an entire set to have consistent visual complexity from one symbol to the next.
5. To further promote figure-ground relationships, the GeoVISTA symbology has a slight shadow and raised appearance, which helps separate the symbols from the base map.

Symbols intended to be interactive with feature a skeuomorphic cue that distinguishes them from non- interactive counterparts.

The user study evaluation will be based on four tasks that cover visual search, semantic relationships, interactivity, and preference. Results are forthcoming.

References

Apple Inc. 2011. iOS Human Interface Guidelines: User Experience. 1-170.

Baus, J., K. Cheverst & C. Kray. 2005. A Survey of Map-based Mobile Guides. In *Map-based Mobile Services: Theories, Methods and Implementations,* eds. L. Meng, A. Zipf & T. Reichenbacher, 193-209. Berlin: Springer.

Bragdon, A., E. Nelson, Y. Li & K. Hinckley. 2011. Experimental analysis of touch-screen gesture designs in mobile environments. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, 403-412. Vancouver, BC, Canada: ACM.

Burigat, S. & L. Chittaro (2011) Visualizing references to off-screen content on mobile devices: A comparison of Arrows, Wedge, and Overview + Detail. *Interacting with Computers,* 23**,** 156-166.

Elias, B. & V. Paelke. 2008. User-Centered Design of Landmark Visualizations. In *Map-based Mobile Services,* eds. L. Meng, A. Zipf & S. Winter, 33-56. Springer Berlin Heidelberg.

Erharuyi, N. & D. Fairbairn (2003) Mobile geographic information handling technologies to support disaster management. *Geography,* 88**,** 312-318.

Follin, J.-M. & A. Bouju. 2008. An Incremental Strategy for Fast Transmission of Multi-Resolution Data in a Mobile System. In *Map-based Mobile Services,* eds. L. Meng, A. Zipf & S. Winter, 57-79. Springer Berlin Heidelberg.

Google Inc. (2011) User Interface Guidelines. http://developer.android.com/guide/practices/ui_guidelines/index.html (last accessed Nov. 4, 2011).

Kuo-Chen, H. (2008) Effects of computer icons and figure/background area ratios and color combinations on visual search performance on an LCD monitor. *Displays,* 29**,** 237-242.

Lee, D. L., M. Zhu & H. Hu (2005) When location-based services meet databases. *Mobile Information Systems,* 1**,** 81-90.

Lee, J., J. Forlizzi & S. E. Hudson (2008) Iterative design of MOVE: A situationally appropriate vehicle navigation system. *International Journal of Human-Computer Studies,* 66**,** 198-215.

MacEachren, A. M. 1995. *How maps work : representation, visualization, and design.* New York :: Guilford Press.

MacEachren, A. M., G. Cai, M. McNeese, R. Sharma & S. Fuhrmann. 2006. GeoCollaborative Crisis Management: Designing Technologies to Meet Real-World Needs. In *7th Annual National Conference on Digital Government Research: Integrating Information Technology and Social Science Research for Effective Government*, 71-72. San Diego, CA.

Meng, L. (2005) Egocentric Design of Map-Based Mobile Services. *Cartographic Journal,* 42**,** 5-13.

Monares, A., S. F. Ochoa, J. A. Pino, V. Herskovic & A. Neyem. 2009. MobileMap: A collaborative application to support emergency situations in urban areas. In *Computer Supported Cooperative Work in Design, 2009. CSCWD 2009. 13th International Conference on*, 432-437.

Morrison, C. & D. Forrest (1995) A study of point symbol design for computer based large scale tourist mapping. *The Cartographic Journal,* 32**,** 126-136.

Weakliam, J., D. Lynch, J. Doyle, M. Bertolotto & D. Wilson. 2005. Delivering Personalized Context-Aware Spatial Information to Mobile Devices. In *Proceedings, 5th International Workshop, W2GIS 20052005*, 194-205. Lausanne, Switzerland.

Willis, K. S., C. Hölscher, G. Wilbertz & C. Li (2009) A comparison of spatial knowledge acquisition with maps and mobile maps. *Computers, Environment and Urban Systems,* 33**,** 100-110.

Developing map symbol standards through an iterative collaboration process

Anthony C Robinson, Robert E Roth ¶, Justine Blanford, Scott Pezanowski, Alan M MacEachren

GeoVISTA Center, Department of Geography, 206 Walker Building, The Pennsylvania State University, University Park, PA 16802, USA; e-mail: arobinson@psu.edu, reroth@wisc.edu, jib18@psu.edu, pezanowski@psu.edu, maceachren@psu.edu

Abstract. Geographic information is commonly disseminated and consumed via visual representations of features and their environmental context on maps. Map design inherently involves generalizing reality, and one method by which mapmakers do so is through the use of symbols to represent features. Here we focus on the challenges associated with supporting mapmakers who need to work together to reach consensus on standardizing their map symbols. On the basis of a needs assessment study with mapmakers at the US Department of Homeland Security, we designed a new, mixed-method symbol standardization process that takes place through a web-based, asynchronous platform. A study to test this new standardization process with mapmakers at DHS revealed that our process allowed participants to identify many issues related to symbol design, meaning, and categorization. The approach elicited sustained, iterative engagement and critical thinking from participants, and results from a poststudy survey indicate that participants found it to be useful and usable. Results from our study and user feedback allow us to suggest multiple ways in which our approach and platform can be improved for future applications.

Introduction

Consumers of geographic information often develop their understanding of geographic phenomena through the use of visual representations of the phenomena and their surrounding environment on maps. To create maps, cartographic designers wield a wide range of graphical and nongraphical generalization operators to simplify reality and communicate a purpose or afford a particular function (Robinson et al, 1995). How these decisions are made depends on a few key concerns, including the desired output format, the map audience, and the message the map should convey or function it should support (Brewer, 2005). The nature of the problem for cartographic designers is such that there is never a single perfect design solution (Monmonier, 1991).

One of the key mechanisms by which cartographic designers can communicate geographic knowledge is through the use of graphic symbols to represent features on a map. Symbols use graphic sign-vehicles to stand for their real-world referents, and the way in which sign-vehicles can be manipulated to support map interpretation has been a focus for decades of research in academic cartography (Bertin, 1983; MacEachren, 1995). Much of this research has focused on characterizing how changes to sign-vehicles may influence the ways in which users perceive and understand symbols on a map (Petchenik, 1977). Somewhat less attention in recent years has focused on the collection, evaluation, and standardization of existing

¶Also Department of Geography, 550 North Park Street, University of Wisconsin-Madison, Madison, WI 53706, USA.

symbols to develop functional symbols sets for application in real-world mapping contexts. This topic is the focus of our present research. Among other things, the development of map symbol standards requires collaboration among multiple cartographers to agree upon symbol sign-vehicles, the definition of referent features, and categories by which symbols may be organized for intuitive application and reuse.

In this paper we present a collaborative, mixed-method approach for tackling these challenges for groups of cartographers who need to develop standardized sets of map symbols. The development of coherent and refined symbol sets among subgroups of cartographers at large agencies like the US Department of Homeland Security (DHS) is a necessary step before one can accomplish the wider goal of sharing symbols across mission areas and ultimately across agencies and other larger entities. Our iterative symbol standardization process uses a web-based, distributed, and asynchronous collaboration platform to deliver round-based activities to help cartographers revise and refine their map symbols, including symbol definitions and symbol categories as well as their basic graphical depiction. To test our approach, we recruited a group of cartographers from the DHS to iteratively audit, refine, and categorize their map symbols.

The following sections describe common symbol standards and processes for their development, the design and development of our new symbol standardization process, results from our case-study application of the process with cartographers at DHS, and what we have learned from this case study to suggest refinements to our collaborative symbol standardization process. We conclude with ideas for future work that emerge from our results.

Symbol standards and their development
Symbol standards

Map symbol standardization received early attention from academic and practicing cartographers over 150 years ago. Funkhouser (1937) highlights a series of proceedings from the 1853–76 meetings of the International Statistical Congress (ISC) as the first printed discussion of map symbol standardization. Proponents argued that the primary advantage of standardization is that the resulting maps can be made directly comparable with one another. Despite these efforts, the standards developed by the ISC were not widely adopted, and practicing cartographers considered them to be impractical.

Interest in map symbol standards was renewed with the rise of the *communication paradigm* in cartography (Board, 1967; Koláčny, 1969). This paradigm specifies that the map is a medium through which the cartographer delivers a message to the map user. To ensure this message is delivered effectively, symbols must be selected by the cartographer to represent geographic features. Standardized symbols can improve cartographic communication by establishing a consistent set of sign-vehicles (Kostelnick et al, 2008).

In subsequent years, some progress on symbol standardization was reported for economic maps (Nikishov and Preobrazhensky, 1971; Ratajski, 1971), topographic reference maps (Joly, 1971; Komkov, 1971), and transportation maps (Rado and Dudar, 1971). Robinson (1973) also noted existing conventions for geologic, hydrologic, and soils maps that were nearing standardization.

Robinson (1973) identified key advantages and disadvantages to implementing symbol standards for thematic maps. Four advantages include: (1) the meaning of a symbol can remain consistent, (2) map users would not need to rely on a legend once a standard has been learned, (3) symbol standards would make map reading easier to teach, and (4) maps are easier to create if symbolization is already prescribed. Today, we would suggest that additional advantages include: (5) the ability to compare multiple maps directly, and (6) improved ease in sharing information within and between organizations. Robinson identified three disadvantages to symbol standardization: (1) resistance from cartographers who are

already employing their own symbolization, (2) inability to adapt the symbol standard to a specific objective or task, and (3) the inability to compensate for map user preferences. We also see disadvantages related to: (4) the inability to reconcile competing conceptualizations of the symbolized geographic phenomenon (Harvey and Chrisman, 1998), (5) the inability of a single graphical standard to reproduce consistently and clearly on different types of media, and (6) inability to enforce the use of a standard once it has been developed.

Our research focus on supporting collaboration and iterative work to develop symbol standards for cartographic design speaks to a broader shift in cartography in recent years from the traditional communication paradigm, where cartographers develop maps as a one-way dissemination of knowledge to map readers (Robinson, 1952), to a notion of cartography as a process of developing knowledge, where representations constantly change and interpretations may change over time as well (Dodge et al, 2009; Kitchin and Dodge, 2007). This latter conceptualization of cartography is reflected in the many transient forms of mapping available today that can be created by end-users, initiated with the purpose of discovering knowledge, and intended to elicit multiple interpretations of reality. The connection to map symbol design is clear—where once it may have been rational to consider the development of a single map symbol design convention that everyone might use (fitting the communication paradigm of cartography), now it is necessary to accommodate the evolving nature of mapping through the continual refinement and redevelopment of multiple symbol sets that may be called into use in different circumstances by different stakeholders.

Emergency management is one of the few application domains in which symbol standards have received a lot of recent attention. Maps quite often provide visual common ground for teams of collaborators who must focus on establishing and maintaining situational awareness in an emergency situation (Tomaszewski and MacEachren, 2006). To be effective, maps for emergency management contexts must be readily interpretable by decision makers, analysts, first responders, and, in many cases, map users in the general public. Developing standard sets of map symbols is one mechanism by which it is possible for mapmakers and map users alike to engage geographic information from emergency contexts in an effective manner (Dymon, 2003; Kostelnick et al, 2008).

Multiple symbol standards designed to support emergency management are in use today. Examples include standards for demining (GICHD, 2005), military operations (Department of Defense, 2008), and emergency response (ANSI, 2006; Spatial Vision Ltd, 2007). The focus of most of these symbol standards is on point symbols, although some recent standardization efforts have also focused attention on symbolizing area features (Kostelnick et al, 2008).

Existing processes for developing map symbol standards

Current methods for developing map symbol standards typically feature multiple phases that include collecting existing symbols, defining features that must be symbolized, and evaluating the resulting symbol standard. Here we describe the specific processes used to develop several recent map symbol standards designed to support emergency management activities.

The ANSI (American National Standards Institute) 415-2006 INCITS Homeland Security Map Symbol Standard is a point symbol standard designed for use during domestic crisis response efforts (figure 1). Development of the ANSI standard featured five steps: (1) create definitions for desired feature types, (2) collect existing symbols, (3) classify those symbols by thematic similarity, (4) produce a matrix showing a recommended symbol for each feature, and (5) logically define each symbol in the matrix (Dymon and Mbobi, 2005). The symbols were then evaluated using an online survey by emergency responders. Symbols not meeting a 75% approval rating were either deleted or modified (22 of 214 symbols failed). A recent

(a)   (b)
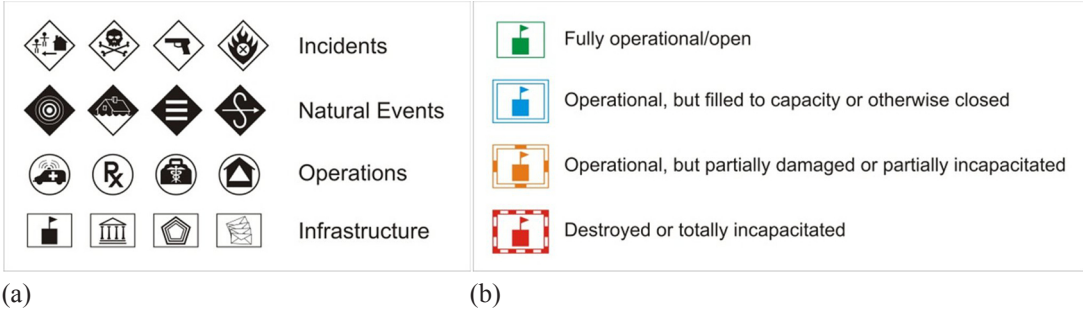
Figure 1. [In color online.] Example symbols from the four symbol categories prescribed by the ANSI (American National Standards Institute) 415-2006 INCITS Homeland Security Map Symbol Standard (a). The ANSI visual method for showing levels of operational status with symbols is also depicted (b).

study of the ANSI symbols conducted with firefighters yielded different results, with only 7 of the 28 fire-related symbols yielding a comprehension rating above 75% (Akella, 2009).

Another symbol standard focused on supporting humanitarian demining operations, the Information Management System for Mine Action, was developed in five steps: (1) survey existing symbols, (2) develop criteria for the design of symbols, (3) design an initial draft of the symbols, (4) qualitatively evaluate the draft symbol set, and (5) revise the symbols according to expert feedback (Kostelnick et al, 2008). Twenty-one domain experts reviewed symbols and their definitions, noting those that should be modified with written comments and suggestions.

The Australian All-Hazards symbol standard extended the Australian Inter-service Incident Management System standard developed to serve a range of emergency response agencies in the Pacific Rim region. The All-Hazards symbol standard includes for point, line, and area features. Its development was completed in three stages: (1) project planning to define tasks, deliverables, and deadlines, (2) consultation and audits to identify existing symbols and their usage, and (3) creation and evaluation of draft and final symbol sets (Martin and Black, 2007).

Characterizing user needs for a new symbol standardization process
Our research focuses specifically on the point symbol needs of the DHS, a domestic security organization that includes twenty-two agencies that focus on a wide range of mission areas, each of which has specialized geographic information requirements.

In preliminary work we focused attention on the ANSI 415-2006 INCITS Homeland Security Map Symbol Standard (ANSI, 2006). We conducted fourteen interviews with mapmakers at seven DHS agencies to characterize the adoption of the ANSI standard, to identify the other map symbol standards and ad hoc symbol sets, to describe critical incidents related to symbology, to identify technical and organizational constraints on symbol standard development and implementation, and to gather feedback on new and improved processes for developing symbol standards. Here we briefly summarize our findings from this study; full details on this research are available elsewhere (Robinson et al, 2011).

DHS mapmakers are responsible for a very wide range of map products. Situation assessment and DHS asset locator maps are among the most commonly developed products. Some DHS mapmakers are engaged in operations centers where maps are requested throughout the development and response to a major event, and maps generated in this context are designed primarily to support basic situational awareness. Other DHS mapmakers (for example, those who work on infrastructure protection and response-planning tasks) focus on developing large collections of reference maps to show critical infrastructure of various types.

Some DHS mapmakers are responsible for creating comprehensive atlases of infrastructure for use by other government agencies to plan security for major social and political events. Still others at DHS are engaged in managing infrastructure used and owned by DHS, since it is a very large government agency with a great deal of property and buildings under its responsibility. Most DHS mapping products include sensitive or classified information, and therefore we are unable to share specific examples here.

Our interview results revealed key issues associated with the adoption of the ANSI standard. The ANSI standard is not used in whole by any of our participants, and is used in part by only a few. Participants state that it does not match their mission-specific needs. The ANSI symbols are also seen as hard to parse, too intricate, and problematic when applied across a range of common map scales. The ANSI standard was intended to play the role of a one-size-fits-all symbol set for DHS use, and participants felt it failed to adequately suit the unique aspects of crisis management activities that the wide range of DHS missions can involve. Participants describe no significant technical issues related to symbol standard development and implementation, but they describe significant organizational challenges that suggest new policies are needed to ensure standards are used.

Participants indicate that they currently use ad hoc, informal symbol standards in lieu of the ANSI standard. These symbol collections typically are developed on a one-time basis by a few cartographers at each DHS agency. Furthermore, our participants suggest that formalizing, refining, and sharing these ad hoc map symbol standards is a way forward toward the development of new, useful symbol standards.

A mixed-method process and platform for standardizing symbols
On the basis of our needs assessment research with cartographers at DHS, we designed a new symbol standardization process intended to formalize, refine, and share existing ad hoc standards. The standardization process we developed relies on a distributed, asynchronous platform so that busy cartographers can participate in standard development without being in the same place at the same time. Our approach makes use of flexible open-source web tools to support and capture the process of standard development. This strategy enables repeatability, ensures that we document key decision points and their rationale, and encourages participants to view symbols from a variety of vantage points.

Our iterative, mixed-method approach is inspired by early work by Suchan and Brewer (2000) who proposed a wide variety of means by which maps can be studied through the use of qualitative methods. Specifically, we focused on the use of ethnographic and survey approaches for eliciting knowledge about the process of mapmaking and map use at DHS. While Suchan and Brewer did not explicitly recommend an iterative approach, they do highlight the ability to triangulate results through the use of mixed-method approaches. We chose to build on Suchan and Brewer's guidance with methodological approaches used to study the usability and utility of geographic information tools and methods that make use of iterative, user-centered design principles (Haklay, 2010; Robinson et al, 2005; Van Elzakker and Wealands, 2006).

Mixed-method standardization process
Four rounds constitute our symbol standard development process. The first round focuses on needs assessment to identify and collect current symbols and map examples as well as to discuss problems with existing symbols and symbol categories. An important component of this stage is the identification of ambiguous or misleading symbols as well as symbols that are poorly designed graphically or do not work well for all required mapping contexts.

In the second round, participants begin developing categories for the symbol set by completing a card sorting activity, a knowledge elicitation technique requiring participants

to assign individual symbols (ie, cards) to one in a set of multiple categories (Cooke, 1994). A description of the utility of card sorting method for map symbol design is provided by Roth et al (2010), which includes a discussion of different card sorting variations that may be employed given various stages of map symbol set design. Following these guidelines, participants complete two sets of card sorting, beginning with an 'open' sort, in which they are able to establish their own set of categories (the second sort is completed as part of the third round). Following the open card sort, participants discuss the sorting results and vote on an initial set of categories for structuring the symbol set. Throughout the second round, participants discuss and vote on how to handle redundant and/or poorly designed symbols identified through the open sort and on ideas for new symbols not included in the sort.

In the third round, participants complete a second, 'closed' card sorting activity in which they assign the revised symbol set to the categories identified and agreed upon in round 2; while participants are not able to create their own categories during this sort, they can make use of an 'other' category. This activity helps ensure that the final standard reflects an agreed-upon structure that has been iteratively refined. This round also includes discussion and voting on topics related to evaluating the new symbol standard.

In the fourth and final round, the symbols are redesigned according to the feedback collected from the prior rounds. The revised symbol set then goes under an external review of the new standard by cartographers and map users for quality control, as well as an evaluation through a tabletop exercise or other scenario-based approach.

A web platform for symbol standardization

Our platform, which we call the e-Symbology Portal (figure 2), is a customized Drupal (http://drupal.org/) application that facilitates the creation of asynchronous, round-based activities for interactive refinement and formalization of a map symbol standard. Activities supported by the e-Symbology Portal include threaded discussions and polls (figure 3), and a wide range of multimedia content can be presented to users in the portal, including text, images, and videos.

Each round has a text-based instruction page that introduces the goals of the round and provides an explanation of and links to the activities included in the round. Each round of participation is opened for a specified timeframe (1–5 days, depending on the activity). Contributions in each round are moderated by a member of our research team to distill

Figure 2. [In color online.] An example of an instruction page for round 1 in the e-Symbology Portal.

Figure 3. [In color online.] An example of an e-Symbology Portal poll, with follow-up discussion.

feedback into key issues for further discussion and voting. In addition, we have implemented a procedure designed to anonymize participation to promote diversity of opinions—similar to the way in which a Delphi exercise (Dalkey, 1969; Linstone and Turoff, 1975) functions.

For the card sorting activities, the process makes use of WebSort (http://www.websort.net), a web-based application that provides graphic and text card sorts through a straightforward drag-and-drop interface (figure 4) (Chaparro et al, 2008; Wood and Wood, 2008). WebSort features analytical tools to help identify clusters in category assignments for cards, which in turn can be used as feedback to participants to help inform iterative development of symbol categories.
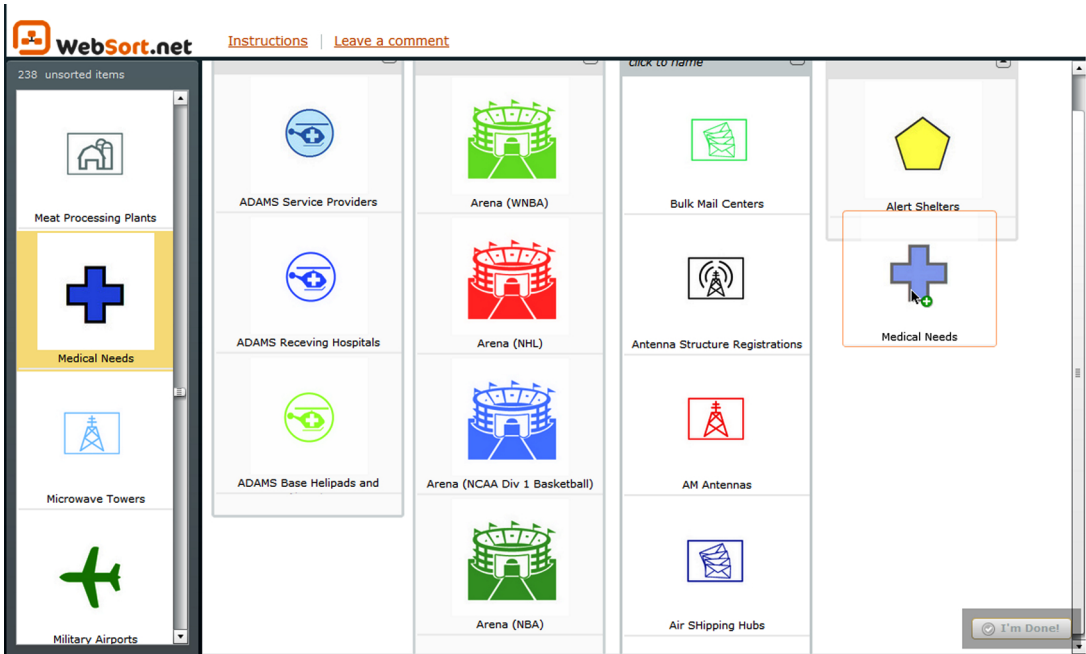
Figure 4. [In color online.] The WebSort tool allows users to develop symbol categories by dragging and dropping symbol 'cards' into user-defined category groups.

It is important to note that we anticipate different user groups to require somewhat different activities in each round of standard development and we have crafted a configurable process and platform to suit different map symbol needs. While the key goals listed above may remain the same, some groups who already have large symbol sets may not need to spend much time developing *new* symbols, and instead may focus on categorization and definition issues. Other groups with more nascent map symbol sets may require a deeper focus on both types of problems.

Developing standard symbology for US Customs and Border Patrol

To evaluate our process for symbol standard development, we worked with seven participants at DHS's Customs and Border Patrol (CBP) division to formalize and refine their ad hoc map symbol standard, a collection of 168 point symbols. CBP participants took part in a three-week study focused on completing the first three rounds of our symbol standardization process (the final round was omitted because it involves external review and evaluation activities). The following sections describe the results we gathered from each round.

Round 1 results

Round 1 focused on identifying problems with the current CBP symbol set, suggesting new symbols that should be included in a refined symbol set, and discussing general issues with respect to the categorization of CBP symbology.

A number of symbols were identified that need improvement. These include symbols that appear too similar (16 examples), are graphically complex (8 examples), difficult to interpret (25), or redundant symbols that represent the same feature (1 example). Participants also identified symbols that need to be added to the current symbol set (5 examples).

In terms of symbol categorization, participants suggested that categories should be kept at a relatively high level rather than too specific. One participant suggested that using an alphabetical matrix was a good idea since this format made it easy to look up symbols.

To prompt further discussion, we asked participants whether symbols in the CBP standard should be categorized at all: four voted yes, one voted no, two did not vote. We also asked participants whether or not the categories they had applied in their ad hoc standard (before starting through our standardization process) should be used in their new standard: three voted yes, one voted no, one voted no categories should be used at all, and two did not vote.

Round 2 results

In round 2, participants completed an open card sorting exercise to develop a range of possible categories for CBP symbols. Using WebSort, participants were presented with a set of cards, each showing a single symbol. Participants were asked to sort these cards into groups of their choosing based upon their similarity. We did not instruct participants in various definitions of similarity; rather, this round was focused on eliciting the diverse range of individual conceptions of categories in order to stimulate and sustain further iterative refinement in later rounds of the process.

Results from participant card sorting in this round show a wide range of possible category options for the CBP symbol set. As noted above, WebSort provides visual and interactive analysis techniques to explore the agreement of symbol groupings across participants. Figure 5 shows a screenshot of the WebSort dendrogram visualization, which uses hierarchical clustering to order the cards according to how often they were placed in the same category by participants. Categorization structures can be explored by interactively changing the number of desired groupings using a slider control.

Using these analysis tools, we were able to identify four general categories that had substantial agreement across all participants: agency facilities, infrastructure, assets, and events. We presented this information to participants and asked them to discuss
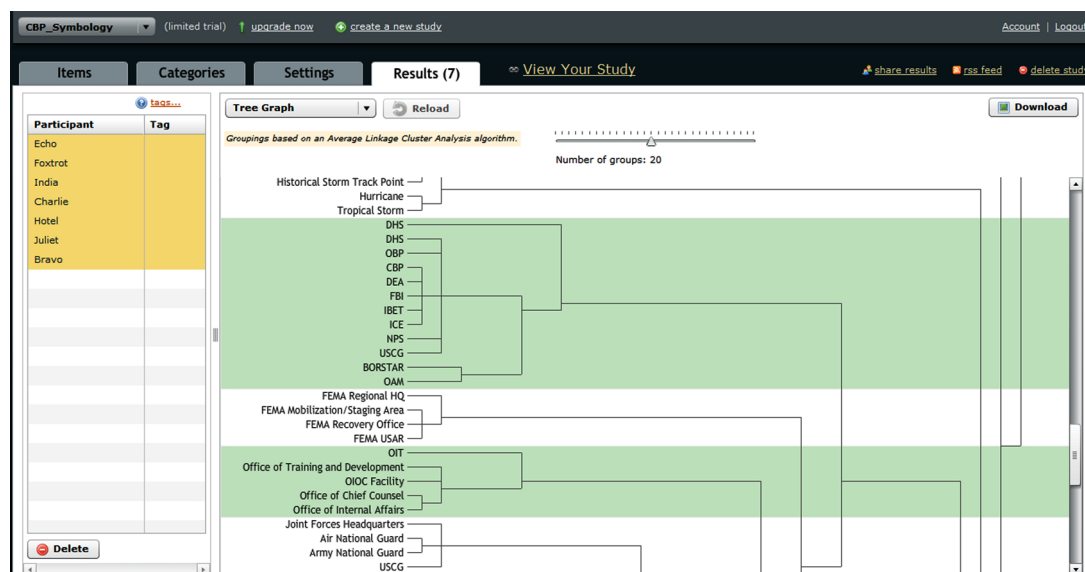
Figure 5. [In color online.] Interactive exploration of card sorting results using the dendrogram visualization in WebSort.

these results and reflect on this category structure. In the discussion, participants stated that the four categories were a good starting point, but too broad to be very useful. Based on further discussion, participants suggested nine more possible categories: CBP, events, assets, picture symbols, DHS, miscellaneous, general not Office of Border Protection (OBP) specific, intelligence, and landmark not OBP Specific. In a subsequent poll, participants voted on which categories to carry over into the next round (round 3) of symbol standard development. The agreed-upon categories for the new standard included: CBP, events, assets, DHS, and miscellaneous and picture symbols (figure 6).

In addition to the card sorting exercise, round 2 asked participants to continue refining the symbol set based on the issues identified during round 1. Among the issues addressed in this activity were ambiguity problems with several symbols, the deletion of one symbol from the CBP standard, and the addition of four new symbols. During this discussion, participants also indicated that a general design improvement for event features was necessary to identify the individual event symbols as part of the same higher level category, such as through the use of a common background color or shape.

Round 3

In the third round, participants completed a closed card sorting activity. Unlike round 2, where participants created their own categories, this time participants were asked to place symbols (including symbol additions/deletions from round 2) into the categories chosen in round 2. Five of the six categories were included in the closed sorting activity: CBP, events, assets, DHS, and miscellaneous. The picture symbols category was not included for logical consistency (ie, the distinction is based not on the feature type, but on the type of symbol representing the feature type) and, after discussions with participants, the miscellaneous category was included to provide an 'other' category for symbols that were not easy to categorize.

The closed card sorting activity was important to the standard development in two ways. First, we were able to identify nineteen symbols that were not placed in any of the categories a majority of the time. Discussions on these ambiguous symbols revealed that a sixth category called general government or external entities was needed to collect the majority of these

Which of the following categories should be included in the new standard?

| view | edit | outline | track |

Submitted by **reroth** on Mon, 03/01/2010 - 00:24

CBP

88% (7 votes)

Events

88% (7 votes)

Assets

88% (7 votes)

Picture Symbols

75% (6 votes)

DHS

63% (5 votes)

Miscellaneous

50% (4 votes)

General not OBP specific

50% (4 votes)

Intelligence

50% (4 votes)

Landmark not OBP specific

0% (0 votes)

Total voters: *8*

» **Add new comment**

Figure 6. [In color online.] Results from the round 2 symbol category poll.

symbols; and a follow-up poll determined that the first term was a most appropriate label for this category.

Second, the closed card sorting activity spurred a discussion among participants about the possibility of including a hierarchical categorization for the symbols. Participants generally felt that the category structure they had developed so far, while valid, was still too vague to be maximally useful. Participants suggested creating subcategories in some instances to provide a hierarchy within the symbol set. Discussion focused on the CBP category in which six subcategories were identified and adopted.

These round 3 activities led to multiple rounds of discussion and voting on which new categories/subcategories to add, what they should include, and general guidelines for what should constitute a reasonable symbol category (eg, maximum number of symbols, and whether or not picture symbols should exist separately as their own category). Three separate discussion threads and ten polls were used during this round. From these activities, participants reached consensus on a final set of categories on which to vote for adoption in
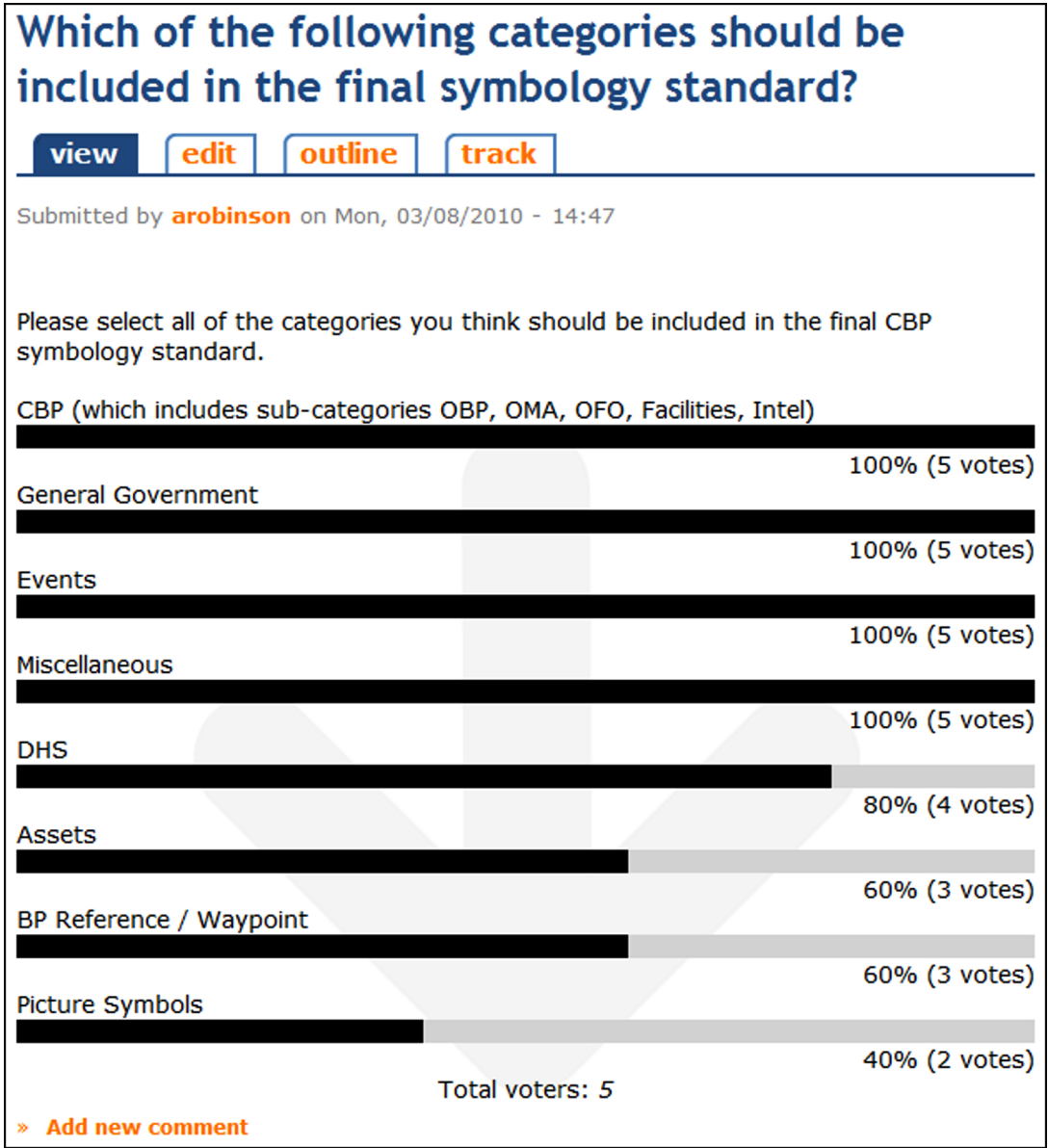
Figure 7. [In color online.] Results from the round 3 symbol category poll.

the CBP standard (figure 7). Six categories were approved: CBP [with subcategories OBP (Office of Border Patrol), OAM (Office of Air and Marine), OFO (Office of Field Operations), and Intel], general government, events, miscellaneous, DHS, and assets. The BP (Border Patrol) reference/waypoint, although receiving a majority of votes, was later determined in discussions to be a subcategory of CBP. The picture symbols category, which did not receive a majority of votes, was included in the final standard because these symbols need to be maintained in a separate ESRI style sheet.

User feedback and process refinement recommendations
In this section we characterize feedback from our study participants as well as the issues we encountered (and recommendations for handling those issues) while conducting and moderating the trial of our standardization process and platform.

Participant survey results

As outlined above, our study resulted in significant changes and refinements to the CBP symbol standard. To further gauge the effectiveness of our process, and to suggest possible improvements, we created a short online survey for participants to complete. Survey results (survey questions and full results available in supplement A at http://www.personal.psu.edu/acr181/Survey.pdf) indicate that most participants were satisfied with the outcome of this study, that the methods we used were helpful toward refining their symbology, that the time commitment required was acceptable, that the materials they received were useful, and that the interactions they had with moderators were positive. This survey also revealed that voting was particularly useful for moving the process beyond back-and-forth discussions, and that a card sorting activity to begin the study in round 1 would have helped to suggest symbol problems/issues to kick start the overall standardization process.

Process issues and recommendations

While executing the case study with CBP, we were able to identify modifications to our approach to improve participation and feedback. First, maintaining consistent participation from busy professional mapmakers remains a challenge. We expected participants to spend roughly 60 minutes over the course of each week-long round to complete the activities. Participants were generally split into two groups: (1) a highly active group that completed all activities and spent a longer than expected time contributing to the message boards and (2) a less active group that missed substantial portions of some activities. Our approach to encouraging participation from the latter group was to send reminder e-mails once every two days. While this strategy was effective in getting passive participants to complete activities that could be completed in a single sitting (ie, the card sorting and polling activities), it was not effective in generating continued contributions on the discussion boards. To overcome this issue, we would like to explore the possibility of adding tangible incentives for participation. In addition, we suggest leaving discussion boards open for a longer period than 5 business days to allow extra time for less active participants to contribute before moving to the next topic.

A second issue, also related to time constraints, was a notable difficulty in transitioning between rounds. A key component of the round-based approach is to have moderators summarize the feedback collected in each round and then use these summaries to tailor activities in the following rounds. Because of constraints on participant availability, each round was opened on a Tuesday and closed on the subsequent Monday, meaning that each round needed to be summarized in a single evening with new content posted by early Tuesday morning. This was difficult for moderators to complete. In the future, we suggest building in 2–3 days between rounds for moderators to summarize the prior round and post the next round's content. This would also help to combat participant fatigue, giving them a break between the time-intensive final voting at the end of one round and the equally time-intensive opening exercise at the beginning of the next round.

A third issue we noted was the high reading load given to participants at the start of each round. Part of our strategy to encourage participation was to supply a document at the start of each round that provided instructions for round activities; mirroring the content that was shown on the e-Symbology site. Several of these guides were quite long, particularly in the earlier rounds when participants were less familiar with the e-Symbology interface. In the future, we would recommend alternative media, such as video demonstrations, to assist in communicating the instructions associated with each round. We have already begun developing several videos to use in the next trial of our standardization process.

Finally, we found that concluding the symbol standard process development requires an additional round in which we present a summary of the standardization results to

the participants. This helps participants evaluate how successful their efforts were and provides the opportunity to hold a concluding vote to approve the final symbol set and its categories. While we conducted both activities in our test with CBP, we had not anticipated the need for these steps in our original process methodology.

Conclusions and future research

In this work we have highlighted the need to support groups of mapmakers in their efforts to standardize map symbols. Previous processes to help define symbol standards have had mixed results. Some standards have been widely adopted, while others have not. Based on prior work and our own needs assessment study with mapmakers at DHS, we designed a new symbol standardization process that blends together multiple methods of knowledge elicitation in a web-based, asynchronous platform.

In the first trial of this new standardization process and platform with mapmakers at CBP, participants identified a large number of issues related to symbol design, symbol meaning, and symbol categorization. Our approach was successful at eliciting sustained, iterative engagement from participants, and feedback from a postparticipation survey indicates that participants were pleased with the outcome. In testing our process and evaluating participant experiences with the process we also learned a variety of ways in which we can improve upon our approach and platform.

The results from our research suggest a wide range of possible new directions for subsequent work. An obvious next step is to refine our symbol standardization process further and to apply it with other groups of mapmakers. A long-term goal is to generalize our approach to the point at which it can be used by a wide variety of mapmakers engaged in topics beyond emergency management. It will be especially vital to focus on supporting teams of collaborators. For example, in crisis mapping, it is common for mapmakers to come from a wide variety of constituent groups, including local, state, federal, and nongovernmental agencies. We can expect each group to come to a situation with a particular set of representational norms in mind, and a process like ours could be used to complete collaborative standard development tasks like those recently completed to develop humanitarian demining map symbols, for example (Kostelnick et al, 2008). New ways of interacting with and collaborating on geospatial problems will also require iterative approaches for refining representation conventions. For example, recent work by Cai and Yu (2009) has focused on supporting collaborative deliberation using maps as devices for supporting argumentation and discussion. A collaborative geodeliberation environment of the future may include mobile interfaces for contributing feedback as well as standard web interfaces and, in addition to group-associated conventions for representing features, it will also be necessary to discuss and refine symbols to ensure their utility and usability across multiple output formats.

Once a standard has been developed, there are not good mechanisms for mapmakers to discover and share symbol sets. One possible solution would be a web-based symbol repository that could allow users to contribute, browse, and share symbols. It is also possible to envision features in such a tool that would allow users to discuss and vote on symbols and symbol categories in much the same way as is done in the standardization process we have outlined here.

Our experiences designing and evaluating a new process for standardizing symbols makes it clear that, while the goal of having usable and useful map symbol standards is an important one, the way toward achieving that goal requires substantial effort on the part of mapmakers, even when the process is facilitated in an asynchronous, distributed manner. Even then, our process required manual moderation in order to flexibly tailor each round of activities. A long-term goal should be to identify parts of our process and other processes

that can be blended into existing mapping tools to make the act of standardizing symbols transparent to the end-user, while still resulting in high-quality, refined symbology.

References

Akella M K, 2009, "First responders and crisis map symbols: clarifying communication" *Cartographic and Geographic Information Science* **36** 19–28

ANSI, 2006, "ANSI INCITS-415 2006 Homeland Security Mapping Standard—Point symbology for emergency management", American National Standards Institute, Washington, DC

Bertin J, 1983 *Semiology of Graphics: Diagrams, Networks, Maps* (University of Wisconsin Press, Madison, WI)

Board C, 1967, "Maps as models", in *Models in Geography* Eds R J Chorley, P Haggett (Methuen, London, UK) pp 671–725

Brewer C A, 2005 *Designing Better Maps: A Guide for GIS Users* (ESRI Press, Redlands, CA)

Cai G, Yu B, 2009, "Spatial annotation technology for public deliberation" *Transactions in GIS* **13** 123–146

Chaparro B S, Hinkle V D, Riley S K, 2008, "The usability of computerized card sorting: a comparison of three applications by researchers and end users" *Journal of Usability Studies* **4** 31–48

Cooke N J, 1994, "Varieties of knowledge elicitation techniques" *International Journal of Human–Computer Studies* **41** 801–849

Dalkey N C, 1969, "The Delphi method: An experimental study on group opinion", RM-5888-PR, The Rand Corporation, Santa Monica, CA

Department of Defense, 2008, "Common warfighting symbology: MIL-STD-2525C", US Department of Defense, Washington, DC

Dodge M, Kitchin R, Perkins C, 2009 *Rethinking Maps: New Frontiers in Cartographic Theory* (Routledge, London)

Dymon U J, 2003, "An analysis of emergency map symbology" *International Journal of Emergency Management* **1** 227–237

Dymon U J, Mbobi E K, 2005, "Preparing an ANSI standard for emergency and hazard mapping symbology", paper presented at International Cartographic Conference, La Coruna, Spain; copy available from …………………………………………..

Funkhouser H G, 1937, "Historical development of the graphical representation of statistical data" *Osiris* **3** 269–404

GICHD, 2005, "Cartographic recommendations for humanitarian demining map symbols in the Information Management System for Mine Action (IMSMA)", Geneva International Centre for Humanitarian Demining, Geneva

Haklay M, 2010 *Interacting with Geospatial Technologies* (John Wiley, Chichester, Sussex)

Harvey F, Chrisman N, 1998, "Boundary objects and the social construction of GIS" *Environment and Planning A* **30** 1683–1694

Joly F, 1971, "Problemes de standardisation en cartographie thematique" *International Yearbook of Cartography* **11** 116–119

Kitchin R, Dodge M, 2007, "Rethinking maps" *Progress in Human Geography* **31** 331–344

Koláčny A, 1969, "Cartographic information - a fundamental concept and term in modern cartography" *The Cartographic Journal* **6** ….–….

Komkov A M, 1971, "The international language of geographical maps" *International Yearbook of Cartography* **11** 209–215

Kostelnick J C, Dobson J E, Egbert S L, Dunbar M D, 2008, "Cartographic symbols for humanitarian demining" *The Cartographic Journal* **45** 18–31

Linstone H, Turoff M, 1975 *The Delphi Method* (Addison-Wesley, Reading, MA)

MacEachren A M, 1995 *How Maps Work* (Guilford Press, New York)

Martin G, Black M, 2007, "Australasian All-Hazards Symbology Project", Spatial Vision Innovations Pty Ltd, Melbourne

Monmonier M, 1991, "Ethics and map design: six strategies for confronting the traditional one-map solution" *Cartographic Perspectives* **10** 3–8

Nikishov M I, Preobrazhensky A I, 1971, "The problems of the unification of the contents and conventional sign standardization on economic maps" *International Yearbook of Cartography* **11** 127–136

Petchenik B B, 1977, "Cognition in cartography" *Cartographica: The International Journal for Geographic Information and Geovisualization* **14** 117–128

Rado S, Dudar I, 1971, "Some problems of standardization of transportation map symbols in thematic mapping" *International Yearbook of Cartography* **11** 160–164

Ratajski L, 1971, "The methodological basis of the standardization of signs on economic maps" *International Yearbook of Cartography* **11** 160–164

Robinson A C, Chen J, Lengerich G, Meyer H, MacEachren A M, 2005, "Combining usability techniques to design geovisualization tools for epidemiology" *Cartography and Geographic Information Science* **32** ….–….

Robinson A C, Roth R E, MacEachren A M, 2011, "Understanding user needs for map symbol standards in emergency management" *Journal of Homeland Security and Emergency Management* **8** 1–16

Robinson A H, 1952 *The Look of Maps* (University of Wisconsin Press, Madison, WI)

Robinson A H, 1973, "An international standard symbolism for thematic maps: approaches and problems" *International Yearbook of Cartography* **13** 19–26

Robinson A H, Morrison J L, Muehrcke P C, Kimerling A J, Guptill S C, 1995 *Elements of Cartography* (John Wiley, New York)

Roth R E, Finch B F, Blanford J I, Klippel A, Robinson A C, MacEachren A M, 2010, "The card sorting method for map symbol design", paper presented at International Symposium on Automated Cartography (AutoCarto*)*, ACSM/CaGIS, Orlando, FL; copy available from …………………….............

Spatial Vision Ltd, 2007, "Australasian All-Hazards Symbology Project", Spatial Vision Innovations Pty Ltd, Melbourne

Suchan T A, Brewer C A, 2000, "Qualitative methods for research on mapmaking and map use" *The Professional Geographer* **52** 145–154

Tomaszewski B M, MacEachren A M, 2006, "A distributed spatiotemporal cognition approach to visualization in support of coordinated group activity", in *3rd International ISCRAM Conference* Eds B V d Walle, M Turoff (…………………. , Newark, NJ) pp 1–5

Van Elzakker C, Wealands K, 2006, "Use and users of multimedia cartography", in *Multimedia Cartography* Eds W Cartwright, M Peterson, G Gartner (Springer, Berlin) pp 487–504

Wood J R, Wood L E, 2008, "Card sorting: current practices and beyond" *Journal of Usability Studies* **4** 1-6

Free Classification of Canadian and American Emergency Management Map Symbol Standards

Raechel A. Bianchetti, Jan Oliver Wallgrün, Jinlong Yang, Justine Blanford, Anthony C. Robinson and Alexander Klippel

206 Walker Building, Department of Geography, GeoVISTA Center, The Pennsylvania State University, University Park, PA 16802, USA
Email: rabianchetti@psu.edu

Emergency management in transnational contexts can be a challenging endeavour. Cultural and language differences among multiple countries can hinder the exchange of information during dynamic emergency response. With increasing international threats and the explosion of near real-time data availability, the emergency response process has become mired in complex communication practices. Maps have the potential to provide an intuitive medium for communication and means for establishing situation awareness during emergency events. The development of map symbol standards is one method for improving communication efficiency. This paper evaluates how the design of two national emergency management map symbol sets (American ANSI and Canadian EMS) influences map-readers' conception of represented information.

Keywords: map symbology, cognition, free classification, emergency management

INTRODUCTION

A key to successful emergency management is the dissemination of timely, accurate information. Emergencies occurring at international borders pose unique challenges to the exchange of information as they cause damage across national boundaries and may require multi-cultural and/or multi-linguistic communication (Boin and Rhinard, 2008). For example, the Kashmir earthquake that occurred on 8 October 2005, caused the loss of more than 74,000 lives and devastated three countries – India, Pakistan and Afghanistan. Response and eventual recovery efforts during events such as this earthquake are guided by information sources generated within individual countries and shared between relief groups and government agencies. Cultural and language differences between countries can affect the ability of emergency collaborators to communicate effectively and to jointly respond to the crisis event. In some cases, it is possible to implement protocols for international reactions to emergency events (Voigt *et al.*, 2007). In other cases, such as emergency management map symbol standards, international agreement has yet to be attained.

A crisis has the potential to impact typical communication processes between agencies, and limit technological capabilities for disseminating new information (Kapucu, 2006). Contemporary crises have repeatedly highlighted critical issues arising from the lack of efficient communication

(Manoj and Baker, 2007). In these cases, reliance on materials generated during preparation for catastrophic events could limit the abilities of managers to synthesize information from various international stakeholders. For example, during the response to the Kashmir earthquake mentioned above, abundant national and international organisations migrated to the region to establish the rescue and response commission (Hicks and Pappas, 2006). In such a case, effective collaboration relies on multiple actors' abilities to understand a heterogeneous set of information.

GIS technology and careful map construction potentially improves the quality of emergency management responses to crisis events (Cova, 1999; Cutter, 2003). While GIS technologies provide managers with the ability to assess new information that is dynamically being shared from a number of sources, conscientious map design allows users to quickly assess and internalize new information about the emergency event. Collaboration among multiple actors in the face of emergencies is both dynamic and reciprocal in nature. Collaboration between actors in either same-place or different-place scenarios may be among pairs of individuals, small teams within or across agencies, or between/among agencies and organisations (MacEachren and Cai, 2006). The intricacies of both the emergency management process and event information make it imperative that communication be unambiguous for its users. As MacEachren and Brewer (2004) point out, 'visual representations have a particularly

important role to play as mediators of geocollaborative activities' (p. 2). Map communication supports collaboration among multiple emergency responders.

Alexander (2004) examined the effects of knowledge, ability, experience and training on the development of maps during emergency response situations. Emergency managers and graduate students were asked to generate maps in response to a textual description of an emergency event, and the map artefacts were then examined. Sixty-seven maps were analysed for six qualities – level of detail, position, and size of features, balance of style and realism, neatness, textual features, and the inclusion or exclusion of a tactical plan on the map. Using these criteria, the maps were classified into seven categories. The results of this analysis showed a wide range of perceptions of the emergency event, showing how individual knowledge and experience can influence the generation of geographic information in response to emergency events, and demonstrating the need for mediation to connect different perceptions.

Map symbol standardisation is one method for enhancing map communication (Akella, 2009). Standardisation includes a process by which a formal symbology is defined for an intended use case. In the absence of symbol standardisation, emergency officials are left to create and employ their own *ad hoc* symbologies (Robinson *et al.*, 2011), similar to the maps generated by participants in the Alexander (2004) study. While these *ad hoc* symbologies are representative of the creator's conception of emergency management events, they may not be representative of other user's conceptions. In international situations, where symbologies pass between users with different cultural backgrounds, these *ad hoc* symbologies can be risky. One particular example where culture plays an important role in user conception of map symbols is the representation of medical facilities (Dymon, 2003; Korpi and Ahonen-Rainio, 2010). Dymon (2003) compared the designs of 44 medical facilities to choose the most representative for the ANSI standard, while Korpi and Ahonen-Rainio (2010) note that symbols representing medical facilities include such diverse symbols as the star of life and the rod of Asclepius. To overcome the risks associated with the use of *ad hoc* symbologies, cartographic symbol standardisation can improve communication by providing users with a pre-determined set of unambiguous symbols.

Previous analyses of map symbol standards have focused on maps as artefacts of the overall design process. Kent and Vujakovic (2009) assessed the stylistic design of topographic maps across Europe using content analysis to describe the number and types of features on topographic maps. The authors concluded that stylistic characteristics of European maps link back to their geographical origins. They describe cartography as a natural visual language for communicating place, where national standards serve as dialects of a larger cartographic language. Another example of the use of content analysis, Ciolkosz-Styk (2011) assessed city maps across Europe by tracking the presence of different types of information on the maps. In both cases, authors assess the general cartographic styles used in a set of maps, but the question remains of how map users interpret these different cartographic styles.

Acknowledging the importance of maps for emergency communication, symbol standards for emergency management, such as the American ANSI INCITS 415-2006 (referred to as ANSI for the hereafter) and the Canadian Emergency Mapping Symbology (referred to as EMS for the hereafter), have been proposed to improve map-based communication. The ANSI and EMS symbol standards present a unique situation for studying the effects of cartographic style, in particular shading and shape, on map user understanding of map symbols. The focus of our present research is to evaluate these two standardization efforts for emergency management.
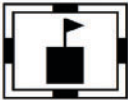
ANSI INCITS 415-2006

The ANSI symbol standard was created following a mandate by the Federal Geographic Data Committee's Homeland Security Working Group (FGDC HSWG) (Dymon and Mbobi, 2005). The creation of the ANSI standard involved surveying and comparing existing symbology and defining a set of symbols for review by emergency management and related officials. The final symbol set consists of 198 unique pictorial symbols categorized into four standard categories: 'Incidents,' 'Natural Events,' 'Operations' and 'Infrastructure'. We use the term 'standard category' for standard prescribed categories used to organize the ANSI and EMS symbols. In contrast, participant categories are referred to as categories for the remainder of the paper. To differentiate between the standard categories, frames around the actual symbols are used (Figure 1). The Incidents and Natural Events standard categories share the same diamond shaped frame, but utilize white (Incidents) or black (Natural Events) background to differentiate between the two categories. Both the Infrastructure and Operations standard categories utilize white backgrounds. The Infrastructure frame is rectangular, while the Operations frame is circular. Representative symbols from each of the four standard categories are presented in Figure 1 along with their category and the ANSI reference definition (http://www.fgdc.gov/HSWG/index.html).

Emergency Mapping Symbology

In 2010, the Emergency Mapping Symbology was introduced as a new standard for emergency management in Canada (Sondheim *et al.*, 2010). Designed to support multi-agency collaboration, the EMS symbol set was developed through the cooperation of government and non-government agencies from Canada and the USA, including the FGDC. The standard was designed by evaluating the ANSI symbols to determine how well the ANSI standard met the mapping needs for Canadian emergency management; the ANSI standard was then extended to meet the Canadian needs, and then stylistically modified.

The EMS symbology consists of 305 pictorial symbols classified into five standard categories: 'Event/Incident (General Case)', 'Event/Incident (Water, Weather)', 'Infrastructure', 'Operation' and 'Aggregate, Other'. Representative symbols from each of the four EMS standard categories that were represented in the classification activity and which are common to the ANSI categories are presented in Figure 2.

Symbol	ANSI Domain	ANSI Definition
	Incident	Wild Fire (Fire Incident) - An uncontrolled fire in a wooded area.
	Natural Event	Tsunami (Hydro-Meteorologic) - A great sea wave produced by an earthquake or volcanic eruption, characterized by high speed of propagation, long wavelength, long period, and low observable amplitude on the open ocean. (Source: Dictionary of Geological Terms, 3rd Ed.)**
	Infrastructure	Schools (Educational Facilities) - A facility for the primary and s econdary education of children. (Source: Adapted from Merriam-Webster Online Dictionary definition)
	Operation	Emergency Staging Areas (Emergency Operation) - A designated place where emergency management forces, equipment, and supplies are assembled prior to engagement in operations.

* Additional defintion given: An ocean wave produced by a submarine earthquake, landslide, or volcanic eruption. These waves may reach enormous dimensions and have sufficient energy to travel across entire oceans. Tsunamis have no connection with tides as inferred by the common use of the term tidal wave.*

Figure 1.   Representative symbols from each of the four ANSI standard categories, labels and definitions. Category membership is visually distinguished by different frames

In contrast to ANSI, the EMS symbol set uses a gradient colour fill to convey the standard categories, a function fulfilled by the frame in the ANSI symbol set. To differentiate weather-related events and other incidents, the colour blue is introduced into the symbol design. Another feature of the EMS not provided by ANSI is the optional black or white borders around the symbols, intended to increase figure–ground separation.

METHODS

Free classification

We used a free classification experimental design in this research, sometimes referred to as card sorting or unsupervised learning. Free classification is a knowledge elicitation method developed within psychology to analyse participant's conception of a set of entities through their categorisation (grouping) practices (Cooke, 1994). Participants in free classification categorize objects based on their own measure of similarity (Spencer, 2009) and the results for multiple participants are compared. This method is based on the assumption that although people may create

categories differently, enough similarities exist between people for their categories (and thus realities) to be comparable (Kelly, 1970; Goldstone, 1994). Free classification experiments have been successfully employed to evaluate point symbols and glyphs (Klippel *et al.*, 2009; Roth *et al.*, 2011).

Participants

Forty-nine undergraduate students participated in a free classification study for course credit. The participants were assigned to one of two conditions, ANSI or EMS. Twenty-four participants (11 females) freely classified the ANSI symbols, average age 20.5 years; 25 participants (13 females), average age 20.2 years, freely classified the EMS symbol set. The degree programmes that participants identified varied, but the majority of students were geography students (31 of 49 participants).

Material

In order to maintain consistency between the symbols used as stimuli, subsets of the ANSI and EMS symbol standards (EMS=127 and ANSI=127) were selected for the free classification study. These subsets of the two symbologies

Symbol	EMS Domain	EMS Definition
	Event/Incident (general case)	Forest Fire - An uncontrolled fire in a forested area.
	Event/Incident (water,weather)	A great sea wave produced by an earthquake, underwater landslide, volcanic eruption or other massive displacement of water, characterized by high speed of propagation, long wavelength, long period, and low observable amplitude on the open ocean.
	Infrastructure	School - A building or collection of buildings comprising a primary, middle or high school.
	Operation	Emergency Staging Areas (Emergency Operation) - A designated place where emergency management forces, equipment, and supplies are assembled prior to engagement in operations.

Figure 2. Representative symbols from each of the EMS categories, their labels and definitions. Category membership is visually distinguished by colour

were chosen so that the point symbol features would be consistent between the two symbol sets. While some features vary in their visual style, the concept they represent is the same. For example, the EMS symbol for the concept 'military' in the EMS standard is two crossed swords, while the 'military' symbol in ANSI is crossed guns. To eliminate the influence of the visual indicators of standard category on the free classification process, these indicators were removed from each of the symbol sets. The EMS symbols were converted to greyscale, while the ANSI symbol frames and background fill were removed. These alterations led to the creation of single pictorial symbols for classification.

Example symbols from both ANSI and EMS are provided in Figure 3.

Design and procedure

The research presented here takes a cognitive approach to assess user interpretation of map point symbols. The goal of this work was to evaluate the effects of map symbol style on map-reader conception of the represented information.

Two free classification experiments were carried out to assess whether the visual style differences between the EMS symbols (rounded, gradient filled symbols) and ANSI

Figure 3. A sample of ANSI and corresponding EMS symbols used in the free classification experiments

symbols (solid black) affect people's conception of the visually presented information. To this end, we utilized a GIS lab that seats 16 students in front of individual 24′ screens. Each workplace was separated from others by using view blocks. The experiment was administered through our grouping software experiment platform called CatScan (Klippel *et al.*, 2011).

Participants were randomly assigned to one of the two conditions (ANSI or EMS) and to their workplace. Participants provided consent and entered basic demographic information such as age, gender and field of study. Next, the participants were provided with a brief introduction to the classification task. The instructions explicitly stated that there is no right or wrong grouping when creating categories, and that it was up to the participants to define criteria to place symbols into same or different groups, that is, symbols they considered similar to each other. Participants were not told that the symbols represented emergency management concepts, only that they were viewing map symbols. Before the primary symbol classification task, they performed a practice task to group unrelated stimuli to ensure that participants had no problems with the general idea of a classification experiment.

Participants performed three main tasks during the experiment: creating groups of individual symbols, naming/labelling each of the groups created, and then describing their grouping rationale for every group they created. In the first part of the experiment, participants created as many groups as they considered appropriate and placed the symbols into the groups. All groups were created from scratch, that is, no number of groups was suggested in the initial screen participants saw (see top half of Figure 4). Participants were able to create groups, place icons into groups, move them between or out of groups, and delete groups.

Upon finishing, the classification task participants were shown the groups they created, one after another. In this stage, participants were asked to (1) provide a short label for the groups that they created; and (2) provide a detailed textual description of the rationale for their grouping choices. An example of the group labelling and description phase of the experiment is shown in Figure 5.

RESULTS

We analysed the results of the experiment using three methods. First, *t*-tests were used to assess whether there were differences in the number of categories that participants created while grouping the symbols and to assess whether any differences in grouping time were significant. Second, heat maps (Wilkinson and Friendly, 2009) ordered alphabetically or based on Ward's cluster analysis were used to compare the placement of symbols in the participants' categories (groups) with the standard categories (Infrastructure, Operations, Incidents and Natural Events). Finally, to explore the terms that participants used in describing the groups they created, we used word cloud visualisation.

Statistical analysis: task length and number of categories

Results of the *t*-test suggest that there is no difference in the time between the ANSI and EMS symbol sets that it took participants to complete the grouping task [EMS: $M=1051.48$, $SD=321.49$; ANSI: $M=977.20$, $SD=316.57$; $t(47)=0.81$, $P=0.42$]. Likewise, there was no significant difference in the number of groups used in the free classification of EMS versus ANSI symbol sets [EMS: $M=15.32$, $SD=7.70$; ANSI: $M=13.5$, $SD=6.87$; $t(47)=0.87$, $P=0.38$]. This suggests that the differences in the symbol style did not significantly affect the length of the classification process or the number of categories (groups) that participants used during classification.

Visualizing symbol relationships

A heat map is a tool for visualizing similarity measures generated during the classification exercise. Each pixel within the heat map represents how many times the corresponding two symbols were grouped together using colour and lightness: the darker the red colour of a pixel, the more often two symbols have been placed into the same category, while light yellow-to-white pixels indicate that two symbols are less commonly grouped together. Figures 6 (ANSI=24) and 7 (EMS=25) present the heat maps ordered by the standard category first and symbol label alphabetically second.
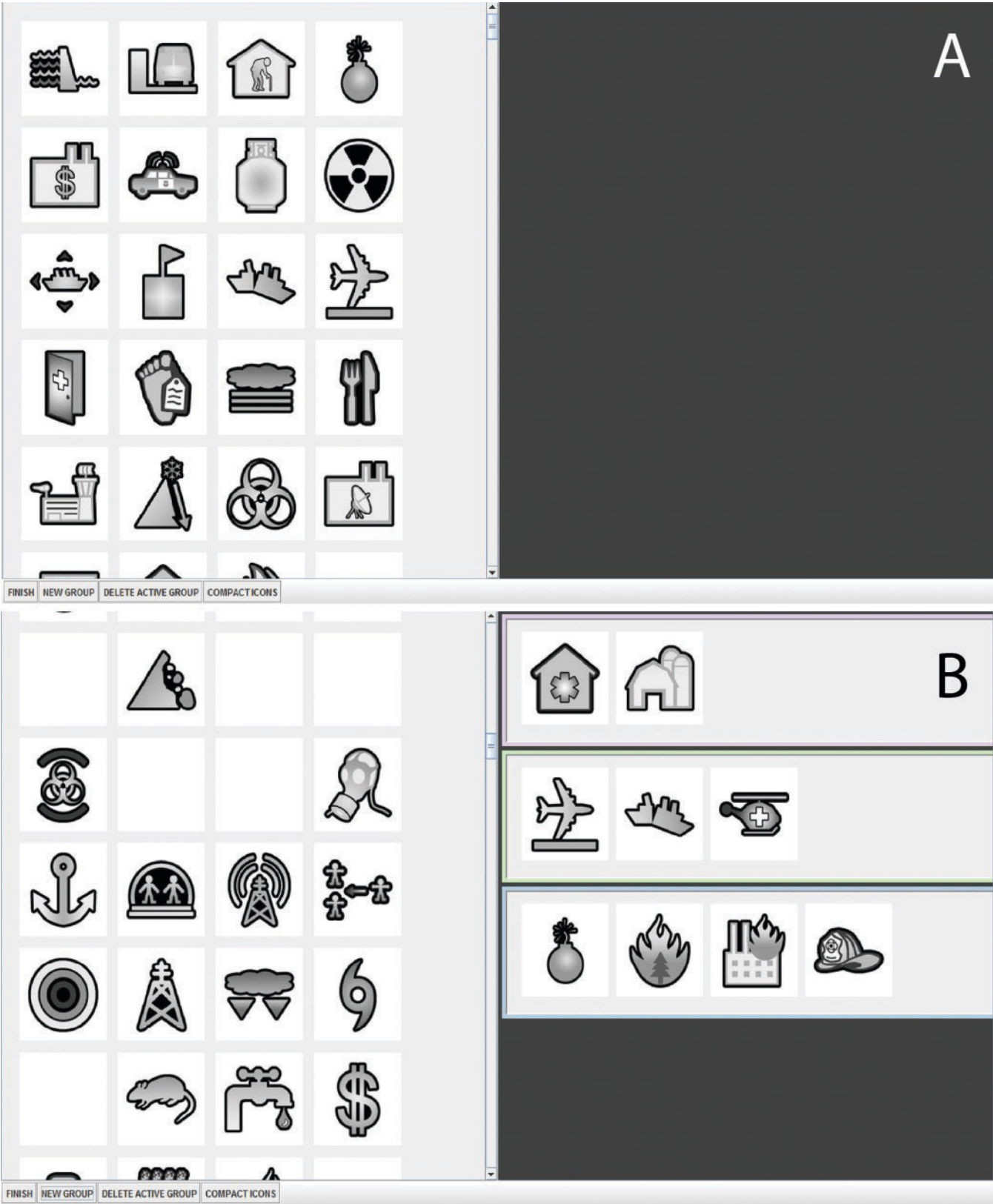
There is little evidence to suggest that participant's categorisation matched that of the four standard categories at a global level. The only exception to this seems to be the Incidents category, which is both smaller and more consistent in design than the other categories. Other highly fragmented clusters were brought out by reordering the heat map by the Ward's cluster analysis results (Figures 8 and 9).

To get a sense of how participants grouped the symbols, we examined the composition of the participants' categories (groups) by the standard categories. This allows us to determine where the two groups of participants agreed and disagreed in their category structures. Figures 8 and 9 present the ordered heat maps and pie charts showing the composition of the groups by the standard categories.

It is clear that the EMS set tended to group into more categories than the ANSI symbol set when comparing the composition of groups within the EMS and ANSI results (see also the higher average number of groups, 15.32 and 13.5, respectively). In addition, the number of categories that were comprised of a single standard category was also greater in the EMS symbol set (EMS=3 and ANSI=1). Several of the groups show similarities across the ANSI and EMS symbol sets. The first group (EMS-A and ANSI-A) is comprised of transportation features depicted in Figure 13 as well as Infrastructure symbols that depict the transportation hubs for each of the transportation types.

Both the ANSI and EMS categories have a homogenous category of Operations (EMS-D and ANSI-E). This set of symbols is represented by a variety of emergency operation equipment, but each of the symbols also includes a cross symbol that is commonly associated with emergency response. The use of the cross is used for symbols in both ANSI and EMS.

Figure 4. A mock-up of the symbol grouping portion of the experiment: the top part (A) shows the initial screen that participants saw; the bottom part (B) depicts an ongoing experiment showing that groups have been created
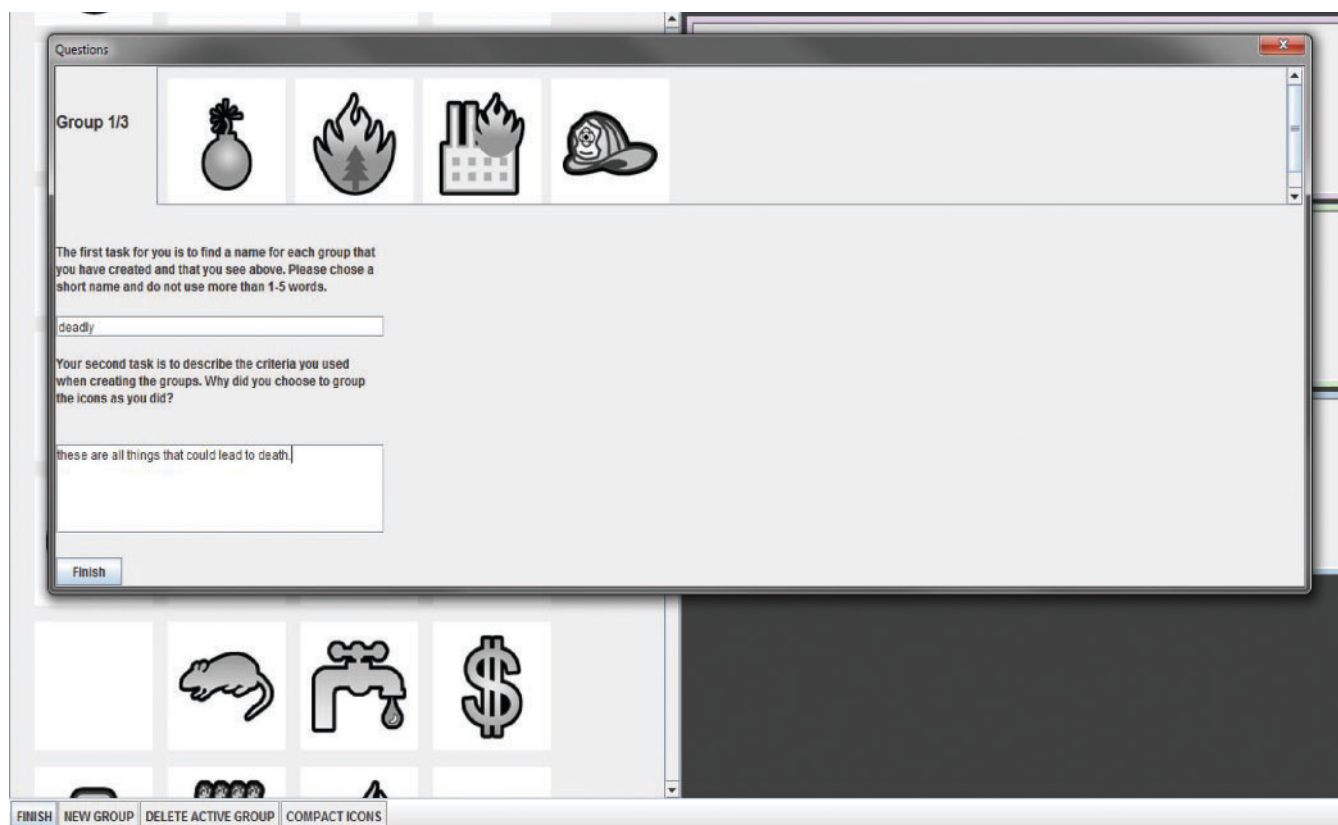
Figure 5. A mock-up of the participant entering group label and description information

One notable similarity across both symbol sets was the classification of several Natural Events symbols. In both the ANSI and EMS cases, the Natural Events symbols representing hail, rain and tsunami were grouped together (Figure 8, ANSI-D; Figure 9, EMS-C). This is particularly interesting because of the drastic iconicity differences between these symbols in ANSI and EMS. Hail and rain
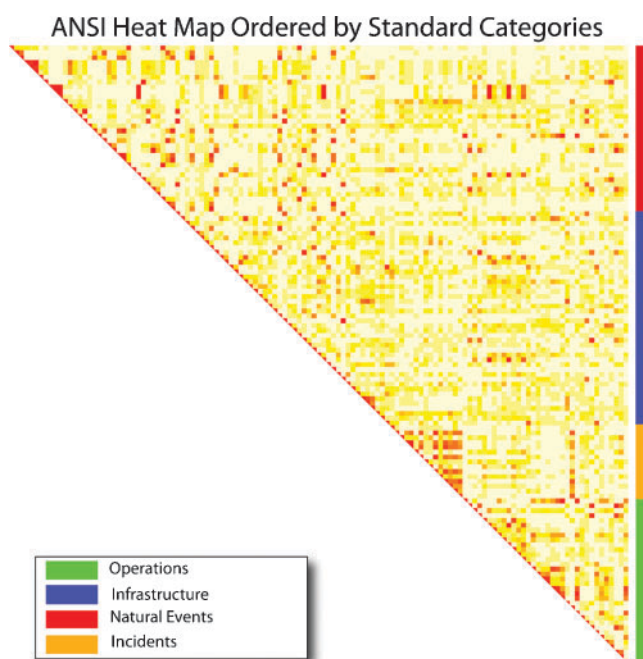
Figure 6. The ANSI heat map ($n=24$) ordered by standard category. Darker red colours indicate higher similarities
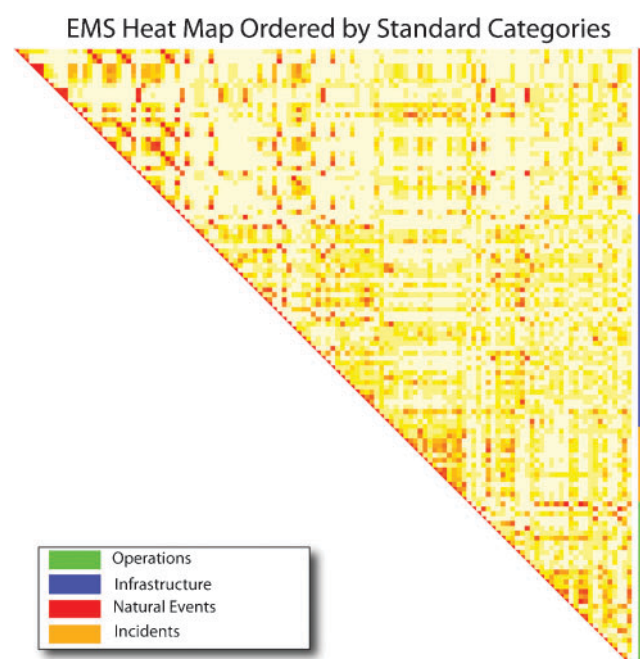
Figure 7. The EMS heat map ($n=25$) ordered by standard category. Darker red colours indicate higher similarities

Figure 8. ANSI heat map and the proportion of symbols representing each standard category (note pie chart colours match key)

Figure 9. EMS heat map and the proportion of symbols representing each standard category (note pie chart colours match key)

are both represented by abstract geometric features in the ANSI symbol set, but in the EMS symbol set, the Rain symbol is represented by an umbrella and rain-drops, while the Hail symbol is represented by a cloud and two triangles.

Word clouds for visualizing symbol category descriptions

A preliminary assessment of the labels and descriptions that participants used to identify their categories was conducted using Wordle (http://www.wordle.net). Wordle is an online tool developed to assess the frequency of word use from textual documents. Word clouds have been used previously to support user decisions about term relevance (Gottron, 2009) and for the visual exploration of terms. For this analysis, we removed a list of words that are not directly related to the semantics of the icons such as pronouns.
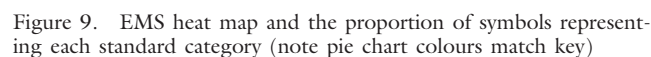
The labels and longer descriptions were used to generate four word clouds. Figure 10 shows two of the short label word clouds. Category labels frequently used across both ANSI and EMS include 'transportation', 'water', 'people' and 'fire'. A similar pattern is also visible in the word clouds of the long category descriptions in Figure 11. The frequencies of the term 'transportation' across both labels and categories are most likely due to the number of features within both symbol sets that depict modes of transportation (Figure 14). Similarly, the popularity of the term 'people' is likely due to the number features that incorporate a human in their design.

DISCUSSION

The work presented here addresses two aspects of symbol sharing, whether the classification tendencies of users match the classification structures imposed by the map standard, and the stylistic design aspects that affected participant

Figure 10. Word clouds generated using the category labels for the (A) ANSI and (B) EMS symbol sets
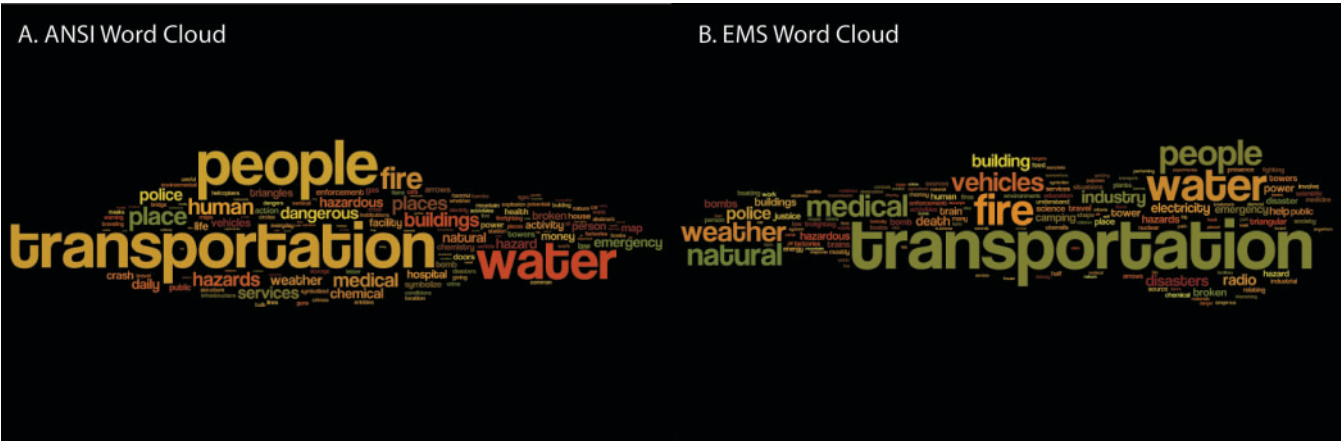
Figure 11.   Word clouds generated using the category description for the (A) ANSI and (B) EMS symbol sets

classification choices. Understanding the natural classification tendencies for these standards can inform the refinement of each standard and provide guidelines for the development of new standards. Additionally, understanding how stylistic differences between the two standards affect user classification can be used to improve symbol designs.

This research brought to light three different stylistic factors that affected user classification strategy: iconicity, repetitive use of a central pictograph and the use of minor symbols to convey differences between central figures. These three factors influenced the classification behaviour in different ways.

The use of a central pictograph for conveying unity across several symbol feature types tended to increase participants' likelihood of grouping the symbols into the same category. An example of this was the emergency operations category. Here several symbols representing aspects of emergency response were grouped together in both the ANSI and EMS cases due to the use of the cross typically used to represent emergency response. The use of a common symbol such as the cross for multiple symbologies could be used to improve map readers' sense of subcategories. These symbols are presented in Figure 12.

A second method that was successful in prompting participants to generate subsets of common symbols was the modification of a common central symbol. This is most evident with the symbols that represent transportation incidents (Figure 13). In the case of EMS, participants grouped the multiple modes of transportation into a single transportation category, while not including the ambulance or medical helicopter features, pictured in Figure 12, into the group, opting instead to include these features with other medical Operations features. The cross worked well as a group-inducing mechanism when it is meaningful.

A third aspect of the map symbol design that prompted users to group symbols was the level of iconicity. Map point symbols should be designed to prompt users to conceptualize their meaning accurately (Muehrcke, 1974); still, many of the point symbols from the ANSI and EMS symbol sets leave much to the imagination during interpretation. Illuminating how iconicity choices affect the classification of these point symbols would create a deeper understanding of another aspect of the design of EMS map symbology. Iconicity is particularly interesting because of the differences between the iconicity levels of the EMS and ANSI symbol

Figure 12.   Medical operations symbols consistently classified together for both ANSI and EMS based on the cross symbol

Figure 13.   EMS and ANSI symbols for transportation. Both symbol sets use a common pictorial symbol, the airplane in this case, to represent a mode of transportation, while using changes in the symbol or the addition of handles to represent changes in the incident type

Figure 14.   ANSI and EMS symbols representing weather features. The level of realism decreases down the list of EMS symbols, while it increases in the ANSI symbols

sets in the case of weather symbols (Figure 14). Despite the use of abstract icons, such as the point symbol for rain, the ANSI weather symbols were still consistently found within the same group. This same pattern was found in the Canadian group, though the level of iconicity used for the EMS symbols was much more realistic.

Participants used the stylistic design choices to their benefit when classifying point symbols. During emergency response, these same principles could possibly be used to interpret foreign map symbols. One example from these two symbol standards that directly applies to cross-border communication is the representation of transportation. Transportation features cross three of the four standard categories that were considered in this study. In the case of the Operations category, we have seen several different types of transportation being used for emergency response. The Infrastructure category uses vehicles for the representation of transportation hubs. Finally, the Incidents category utilizes the transportation mode in its representation of different types of transportation incidents. In all of these cases, modes of transportation common to the USA or Canadian cultures are used in the representation. In this case, the two cultures are similar enough that the transportation types remain the same. In each of the uses of the transportation type, the symbol is altered to reflect a more specific role within the standard category. For instance, the use of the cross on the helicopter defined it as an emergency helicopter. In symbol standards that are not so culturally linked, the use of common stylistic design choices could help map readers with little knowledge of the country's culture to draw conclusions about relationships between multiple point symbols. In contrast to the case presented here, it is probable that other cultures would have

other transportation modes, such as the use of collective taxis or animals.

CONCLUSIONS

During the Kashmir earthquake of 2005, multiple international organisations came together to provide relief to the nations effected by the disaster. Thompson (2011) points out some of the difficulties that arose during the response phase including a lack of a general geographic information depository and the lack of a common operating picture. Confusion brought on by responders' differing perceptions of the crisis led to poor decision-making and wasted time. The work presented here has the potential to improve these communications by addressing how the variation between symbol design between two geographic information sources affects map reader's understanding to the map symbols.

The research presented here is a first step in determining which stylistic choices influence map reader's understanding of map symbols. Here we have compared the influence of differences in visual style. The scope of this current research is limited to two symbol standards from English-speaking countries and neighbours, the USA and Canada. To further our understanding of how these stylistic difference affect map-user conception, it will be important to extend this study to symbol standards developed in countries with different cultural heritages and native languages.

The results of this study suggest that participants did not use categories equivalent to the ANSI or EMS standard categories under the free classification paradigm. This is most likely due to the heterogeneous nature of point symbols within each of the ANSI and EMS standard categories. Each standard category consists of a number of feature types that when analysed together without explicit indication of the standard classifications seem unrelated. Participants used many more categories (groups) during the classification, 13–15 categories on average. Participants created many smaller consistent categories that were influenced by three stylistic factors of the point symbol design, level of iconicity, the modification of a common central pictorial symbol, and the addition of a common marker to show relationships between disparate pictorial symbols. This suggests that the use of four general categories may be too broad for participants to differentiate between categories using the pictorial symbols alone, but that smaller subsets of symbols can be differentiated by the use of the pictorial symbols. To support cross-border communication, using stylistic choices that enhance these natural subgroups could potentially make communication more efficient.

Several questions remain about user interpretation of map symbol features. Future analysis of these symbol sets will look at how the categorisation of map symbol labels compares to the results presented here. Additionally, we would like to examine the meanings that map readers take from these point symbols across more significantly different cultural backgrounds. Classification could also be used to determine if there is any difference in participants' ability to classify symbols with the presence of either frames or colour, as prescribed by the ANSI and EMS standards.

BIOGRAPHICAL NOTES

Raechel Bianchetti is a third-year PhD candidate at the Pennsylvania State University, University Park, Pennsylvania. Her previous degrees focus on physical geography and the application of remote sensing for forestry and geologic applications. Her dissertation focuses on the cognitive aspects of aerial photograph interpretation with the goal of using geovisual analytics approaches to improve forest disturbance detection and attribution. Her interest in remote sensing is fuelled by a love of landscape photography and an appreciation for early photographic interpretation. Her other research interests include uncertainty representation, geology and land cover mapping.

ACKNOWLEDGEMENTS

REFERENCES

Akella, M. K. (2009). 'First responders and crisis map symbols: clarifying communication', **Cartography and Geographic Information Science**, 36, pp. 19–28.

Alexander, D. E. (2004). 'Cognitive mapping as an emergency management training exercise', **Journal of Contingencies and Crisis Management**, 12, pp. 150–159.

Boin, A. and Rhinard, M. (2008). 'Managing transboundary crises: what role for the european union?', **International Studies Review**, 10, pp. 1–26.

Ciolkosz-Styk, A. (2011). 'Comparison of city maps' content of Western Eastern, and Central Europe', **AUC Geographica**, 1, pp. 7–14.

Cooke, N. J. (1994). 'Varieties of knowledge elicitation techniques', **International Journal of Human–Computer Studies**, 41, pp. 801–849.

Cova, T. J. (1999). 'GIS in emergency management', **Geographical Information Systems**, 2, pp. 845–858.

Cutter, S. L. (2003). 'GIScience, disasters, and emergency management', **Transactions in GIS**, 7, pp. 439–446.

Dymon, U. and Mbobi, E. (2005). 'Preparing an ANSI Standard for Emergency and Hazard Mapping Symbology', in **International Cartographic Conference (ICC 2005)**, 1–6, A Coruna, Jul 9–16.

Dymon, U. J. (2003). 'An analysis of emergency map symbology', **International Journal of Emergency Management**, 1, pp. 227–237.

Goldstone, R. L. (1994). 'The role of similarity in categorization: providing a groundwork', **Cognition**, 52, pp. 125–157.

Gottron, T. (2009). 'Document word clouds: visualising web documents as tag clouds to aid users in relevance decisions', **Lecture Notes in Computer Science**, 5714, pp. 94–105.

Hicks, E. K. and Pappas, G. (2006). 'Coordinating disaster relief after the South Asia earthquake', **Society**, 43, 42–50.

Kapucu, N. (2006). 'Interagency communication networks during emergencies', **The American Review of Public Administration**, 36, pp. 207–225.

Kelly, G. A. (1970). 'A brief introduction to personal construct theory', in **Perspectives in Personal Construct Theory**, ed. by D. Bannister, pp. 3–20, Academic Press, London.

Kent, A. and Vujakovic, P. (2009). 'Stylistic diversity in European state 1 : 50 000 topographic maps', **The Cartographic Journal**, 46, pp. 179–213.

Klippel, A., Hardisty, F. and Li, R. (2011). 'Interpreting spatial patterns: an inquiry into formal and cognitive aspects of Tobler's First Law of Geography', **Annals of the Association of American Geographers**, 101, pp. 1011–1031.

Klippel, A., Hardisty, F. and Weaver, C. (2009). 'Star plots: how shape characteristics influence classification tasks', **Cartography and Geographic Information Science**, 36, pp. 149–163.

Korpi, J. and Ahonen-Rainio, P. (2010). 'Cultural constraints in the design of pictographic symbols', **The Cartographic Journal**, 47, pp. 351–359.

MacEachren, A. M. and Brewer, I. (2004). 'Developing a conceptual framework for visually-enabled geocollaboration', **International Journal of Geographical Information Science**, 18, pp. 1–34.

MacEachren, A. M. and Cai, G. (2006). 'Supporting group work in crisis management: visually mediated human–GIS–human dialogue', **Environment and Planning B Planning and Design**, 33, pp. 435.

Manoj, B. S. and Baker, A. H. (2007). 'Communication challenges in emergency response', **Communications of the ACM**, 50, pp. 51–53.

Muehrcke, P. (1974). 'Beyond abstract map symbols', **Journal of Geography**, 73, pp. 35–52.

Robinson, A. C., Roth, R. E. and MacEachren, A. M. (2011). 'Understanding User needs for map symbol standards in emergency management', **Journal of Homeland Security and Emergency Management**, 8, pp. 33.

Roth, R. E., Finch, B. G., Blanford, J. I., Klippel, A., Robinson, A. C. and MacEachren, A. M. (2011). 'The card sorting method for map symbol design', **Cartography and Geographic Information Science**, 28, pp. 89–99.

Sondheim, M., Leeming, G. and Charmley, D. (2010). **Emergency Mapping Symbology**, Refractions Research, Victoria, BC.

Spencer, D. (2009). **Card Sorting: Designing Usable Categories**, Rosenfeld Media, Brooklyn, NY.

Thompson, W. C. (2011). 'An assessment of geo-information use during the 2005 Kashmir earthquake response and recommendations to improve future use', **International Journal of Emergency Management**, 8, pp. 26–41.

Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K. and Mehl, H. (2007). 'Satellite image analysis for disaster and crisis-management support', **IEEE Transactions on Geoscience and Remote Sensing**, 45, pp. 1520–1528.

Wilkinson, L. and Friendly, M. (2009). 'The history of the cluster heat map', **The American Statistician**, 63, pp. 179–184.

Geo-Social Visual Analytics

Wei Luo, Alan M. MacEachren

GeoVISTA Center, Department of Geography, Pennsylvania State University, 302 Walker
Building, University Park, PA, 16802, USA

Abstract: Spatial analysis and social network analysis typically consider social processes in their own specific contexts, either geographical space or network space. Both approaches demonstrate strong conceptual overlaps. For example, actors who are close to each other tend to have greater similarity than those who are far apart; this phenomenon has different labels in geography (spatial autocorrelation) and in network science (homophily). In spite of those conceptual and observed overlaps, the integration of geography and social network context has not received the attention needed in order to develop a comprehensive understanding of their interaction or their impact on outcomes of interest. In order to address this gap, this paper discusses the integration of geographic with social network perspectives applied to understanding social processes in place from two levels: the theoretical level and the methodological level. At the theoretical level, this paper argues that the concepts of nearness and relationship in terms of a possible extension of the First Law of Geography is a matter of both geographical and social network distance, relationship and interaction. At the methodological level, the integration of geography and social network contexts are framed within a new interdisciplinary field: visual analytics, in which three major application-oriented subfields (data exploration, decision-making, and predictive analysis) are used to organize discussion. In each subfield, this paper presents a theoretical framework first, and then reviews what has been achieved regarding geo-social visual analytics in order to identify potential future research.

Keywords: Geography, Social Network, Visual Analytics, First Law of Geography, Data Exploration, Decision-Making, Predictive Analysis.

1 Introduction

Modern society has become an increasingly interconnected world of techno-social systems embedded with dynamic multi-scale networks (*e.g.*, the internet, transportation). The complex interactions within and among these networks always have geographical constraints, whereas they also change or reshape the traditional notion of geographical effects (*i.e.*, distance) [93]. To effectively understand the

interaction between space and techno-social networks, recent research in physics emphasizes the power of networks in which space becomes a background to visualize and understand network analysis results [152] whereas research in geography encourages the integration of spatial thinking into traditional social science through the concepts of space, place and time [66-67], but often treats networks in a simplistic way. This paper argues that space and social networks should be considered simultaneously, a perspective which has not received enough attention. We examine the interaction of geographical and social network contexts from the perspective of the new multidisciplinary research field of Visual Analytics, which is defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [153]. Visual analytics provides a framework for integration of computational analytical methods with visual interfaces to both the information of interest and the computational methods that enable human analysts to cope with large, complex, and heterogeneous data sources and complex questions that these data sources make it possible to address.

Understanding large and complex techno-social networks and their interaction with space at geographic scales requires advances in computational methods. However, computational methods alone have limits and biases because of the predefined structures they have, which greatly limit their analytical power. The process and results of any computational techniques have limited value without input from human analysts to select appropriate methods, to set parameters, to interpret results, to understand what to do next, and to draw conclusions [10]. Visualization of data and computational processing gives users an intuitive representation, greatly promoting application of human perceptual and cognitive information processing capabilities. A simple combination of visualization with computational analysis, however, is not sufficient. Thus, the goal for visual analytics is to integrate human and computational reasoning in more fundamental ways, bringing the experts' background knowledge, creativity, and intuition into the analysis process through an interactive visual environment, in order to combine the strengths of humans and computers to enable an insight gaining process [86].

Visual analytics provides a potential conceptual approach and set of tools to integrate geographic and social network contexts of human processes, but the application of visual analytics to this challenge remains relatively underdeveloped in the literature. Thus, the goal of this paper is to examine the interaction of both contexts from the perspective of visual analytics. To achieve this goal, we address four objectives: (1) to present a theoretical framework in which geography and social network contexts could be combined, (2) to explore the potential common areas with the goal of linking the theoretical framework to visual analytics in order to develop an integrated analytical method we are calling "geo-social visual analytics", (3) to review what has been achieved in relation to the integration, and (4) to identify potential future research for advancing their integration. From here on, we will refer to the interaction of

geographical and social relationships as geo-social relationships, and the interaction between both in terms of visual analytics as geo-social visual analytics.

2   Geo-Social Relationships at a Conceptual Level

Social processes take place within particular contexts. Spatial analysis in social science is used to identify geographical patterns that result from social processes and to understand how space affects such processes. Most spatial analysis is based on an explicit or implicit assumption of the First Law of Geography [157]: "Everything is related to everything else, but near things are more related than distant things". Social network analysis is used to understand how relationships among actors (*i.e.*, individuals, groups, or other social collectives) within a network affect or are affected by social processes [162]. Social network analysis has an assumption, complementary to the First Law cited above, that actors with similar relations may have similar attributes/behaviors. Spatial analysis and social network analysis consider social phenomena in their own specific contexts, either geographical space or network space, but both contexts should be considered together when they contribute simultaneously.

Both contexts demonstrate strong conceptual overlaps. Hess [76] proposes a geographically informed theoretical framework to understand the behaviours of social actors in geography through integrating territorial embeddedness, network embeddedness and societal embeddedness (Figure 1). This framework is used to study economic actions from a critical human geography perspective, but this paper aims to extend it into a generic framework to study geo-social relationships. The concept of embeddedness has been prominently used by geographers to understand the behaviours of social actors in specific contexts [127]. Societal embeddedness refers to societal (*i.e.*, cultural, political, etc.) background from which actors come, in which actions of actors are influenced, and to which actors make contribution. Network embeddedness refers to the importance of relational aspects (*i.e.*, social relations, cultural relations) among social actors to shape the actors' behaviours and of actors' behaviours to change relations. Territorial embeddedness refers to the specific places in which the actors behave: how the places influence actors' behaviours and attributes; how the actors' behaviours change the territory. The overlap area in Figure 1 between the territorial embeddedness and the societal embeddedness fits the First Law of Geography [157]. The overlap area between the network embeddedness and the societal embeddedness fits the homophily principle in social network analysis theory: similarity breeds connection [118]. The overlap area between the territorial embeddedness and the network embeddedness fits a common phenomenon: how space constrains the development of networks and how networks reshape the space. The three overlap areas in Figure 1 indicate that they are not mutually exclusive, but interact with each other; the emphasis is on the interaction between geographic and social network context and the impact of the interaction on outcomes of interest. The

overlap area suggests a possible extension of the First Law of Geography: *Everything is related to everything else, but near things are more related than distant things. Nearness can be considered a matter of geographical and social network distance* [60].
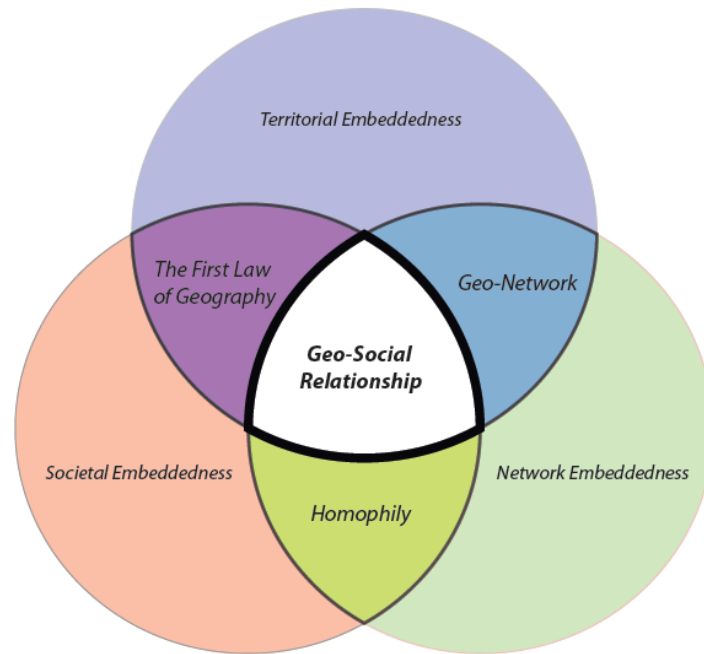

Figure 1: Proposed conceptual framework for geo-social relationships based on fundamental categories of embeddedness [76]

In addition to distance, the other two important concepts implied in the First Law of Geography are relationship and interaction: ''everything is related to everything else''. Flint [59] argues that "the nature of a place is the combination of both locations and their connections to the rest of the world". Prager [125] also argues that geographical locations would be unrelated without relationships and interactions, whereas such relationships and interactions would be meaningless without the context that geographical locations provide. Those arguments emphasize the concept of social construction of space in which social relationships and processes should be refocused, in addition to assuming that the geography of location defines the spatiotemporal process. Some of those social relationships may be constrained within the place, whereas others may stretch out to link geographical locations to wider relations and processes [116]. There is an increasing understanding of the importance of combining geographical space and network from different sub-disciplines within geography, such as political geography [101], economic geography [142], and geographical information science (GIS) [26]. Staeheli [146] argues that spaces become "social locations" embedded in "webs of cultural, social, economic, and political relationships".

Ashdown [15] even argues that the political power is now shifting from a dominance of western culture to a collective governance at the global space, because we have come into a new interlocked age (*i.e.,* of complex, interconnected networks) that causes our destinies to be shared with our enemies. We believe that connections to other locations matter most for the future of different locations in the new age, and that by linking individual location to the group of locations, the science of social networks can explain key aspects of how observed spatial-social patterns evolve. Therefore, to know the spaces, we must understand how spaces are connected geographically and socially. This contention may be increasingly true in today's digitally hyper-connected world.

We are living life in the network: we check our e-mail, make a phone call, take transportation, or update our status in Facebook [98]. Our ties to others affect our ideas and behaviours (*i.e.,* emotional, sexual, and health-related) [42], and the interaction of such individual-level behaviours develops macro social phenomena observed in a spatio-temporal framework. Network science [35] is regarded as the approach to study such phenomena in our interconnected age [167-168]. A growing literature in network science explores how everything is related to everything else [20, 83, 148]. For example, the famous small world experiment [120] formalizes the notion that each person only has six degrees of separation from anyone else on earth [89]. Sui [149] further argues that Tobler's First Law of Geography is a big idea for a small world, because it only takes a few steps to turn a large world into a small world. Based on all of the discussions, we argue that, in addition to considering geographical and social network distance as the necessary components of nearness in the First Law of Geography, relationships can also be considered a matter of geographical and social relationship and interactions.

In addition to the exploration of geographical and social relationships and interactions, the conceptual framework (Figure 1) allows such relationships to be put into certain societal contexts (*i.e.,* political, economic, cultural). On the one hand, the premise "near things are more related than distant things" in the extended First Law of Geography proposed in this paper implies that certain local factors and circumstances can make geo-socially close areas different from geo-socially distant areas. On the other hand, the premise "everything is related to everything else" indicates that there are factors that make contributions to patterns and connections among areas. The societal context provides the framework to explore different factors (*i.e.,* political, economic, cultural) behind observed geo-social patterns, and how such geo-social patterns interact with those factors to generate new geo-social patterns.

In work that complements the discussion above of geo-social integrations at the conceptual level, Adams, Faust and Lovasi [2] identify five conceptual strategies for the integration based on current geo-social relationship research: (1) spatial impacts on the development of social networks over varying spatial scales, such as offices [135,

171], communities [45, 58], and so on; (2) the impact of social network on the places people select to inhabit [163]; (3) use of peer network structures to determine neighbourhood boundaries [77]; (4) the interactive impacts between spatial and social relationships [104, 126, 140]; and (5) multiple context impacts on outcomes related to social, health, and other processes [48, 119, 150]. The five conceptually geo-social integrations also fit the conceptual framework (Figure 1). The first integration focuses more on spatial constraints; the second one emphasizes the network effects on residence selection; the third and fourth ones stress the interaction between geographical and network relationships; and the last one highlights the multiple context impacts on outcomes.
.

The proposed conceptual framework comes from a critical human geography perspective, so how can we apply it to the visual analytics domain? Previous research already set examples to integrate critical human geography perspectives into visualization through identifying areas of common ground at the conceptual level of two methods, such as feminist visualization [94] and grounded visualization [88]. Knigge and Cope [88] integrate grounded theory and visualization through summarizing six points. We borrow five of them and use them to integrate our proposed geo-social relationship conceptual framework and visual analytics at the theoretical level. First, the proposed conceptual framework and visual analytics are exploratory approaches which involve iterative explorations with matching mental models to construct knowledge. Second, both are iterative approaches which involve recursive processes of data collection, visualization, and analysis with critical thinking at each step. Third, both methods are enriched through connecting real-world phenomena and human experiences to broader processes. Fourth, both methods require multiple interpretations and representations, because they believe there is no single correct way to interpret and visualize data [144]; Fifth, both methods recognize the importance of situating knowledge construction into the historical, geographical, and cultural context.

3   Geo-Social Visual Analytics at a Methodological Level

Following the introduction to geo-social relationships at a conceptual level, this section discusses how to put the geo-social relationships into practice. From an application perspective, visual analytics methods can be classified into three groups focused on support for: data exploration, decision-making, and predictive analysis. This classification is used here to organize the methodological level in terms of geo-social visual analytics. Visual analytics aims to amplify the human reasoning process, so it must build on an understanding of that reasoning process [154]. For the above three categories in visual analytics, this paper presents reasoning frameworks for each group first, and then discusses corresponding geo-social visual analytics technologies in order to identify potential future research directions.

3.1 Data Exploration

3.1.1 Conceptual Framework

Data exploration is a primary task in visual analytics to make sense of overwhelming amounts of disparate, conflicting, and dynamic data in a novel manner. Here, we use the Feature-ID model for geovisualization (Figure 2) [107], an extension of an earlier pattern-matching model of cartographic visualization [109], as the reasoning framework for data exploration in terms of geo-social visual analytics. The pattern-matching model, in turn, draws upon a general scientific visualization perspective based on human cognition [63] to support understanding of human-display interaction in the context of map-based geovisualization.

The iterative process between human and computer interaction through "seeing", "interpreting", and "constructing-knowledge" shown in Figure 2 is an insight gaining process, which is a primary goal for visual analytics. Yi et al. [179] characterize insight gaining as a multi-step process: provide overview, adjust, detect pattern, and match mental model. A reasonable insight gaining process starts with seeing an overview of a domain area. Having a big picture may or may not lead to direct insight, but it is a good starting point for people to take actions to make additional inquires about the areas in which they are interested in gaining more knowledge. Adjust refers to a process through which people adjust the level of abstraction to explore a data subset of interest in order to make more sense of the data. Detect pattern refers to identifying interesting results that can include specific distributions, anomalies, clusters, and trends in the datasets. Match mental model (equivalent to "instantiate schema" in Figure 2) refers to reducing human cognitive load and amplifying recognition through providing a visual representation of data to decrease the gap between the data and user's mental model of it, as well as the gap between the visual representation and real-world knowledge. The whole process of visual interactive exploration is a mental model building process from knowledge development to critical breakthrough [39] that involves the iterative interaction among "seeing", "interpreting", and "constructing-knowledge" at each step. Therefore, geovisualization works as a larger cognitive system to support a human reasoning process. Geo-social visual analytics aims to understand the interaction between two contexts and the impact of multiple contexts on reasoning outcomes through integrating social network space into geographic space to provide users a broader perspective.
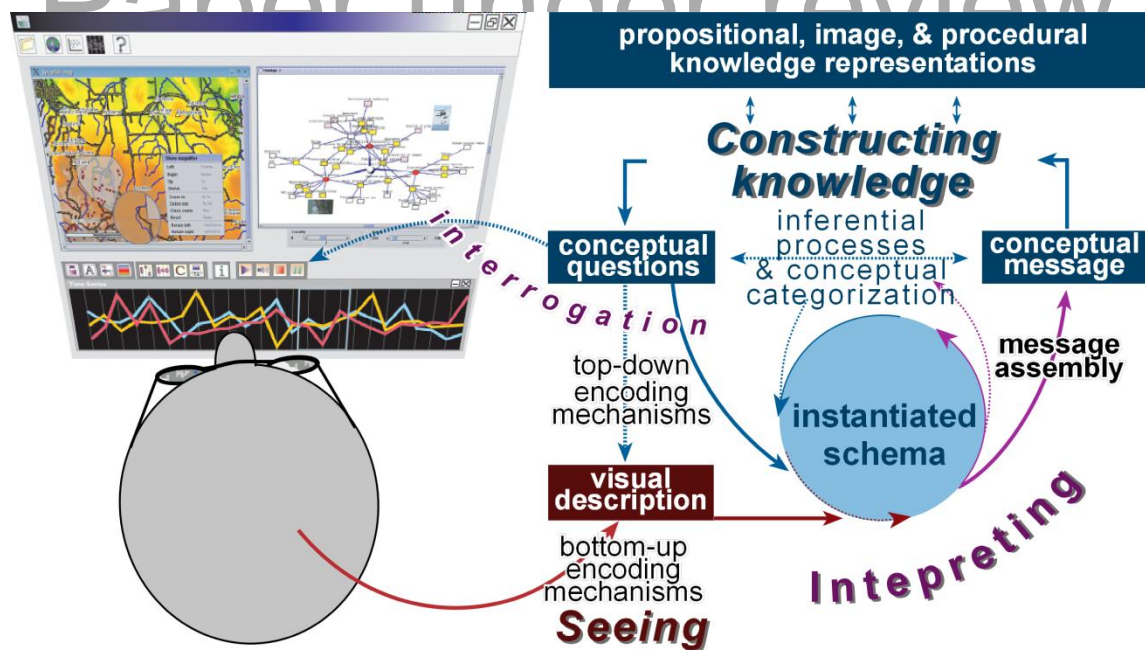
Figure 2: Feature-ID model of geovisualization; elaboration of ideas first presented in Figure 8.1 in [107].

3.1.2   Geo-Social Visual Analytics in Data Exploration

Network representations, especially network theories have not been fully considered in geographical information science [125], but they have great potential to offer insight into complex geographical phenomena in terms of geo-social interactions. The following section aims to fill this gap. Given complex relationships between geographical space and social space at different spatial scales, this paper characterizes the relationships into two groups: (1) geo-social relationships among geographical areas (*e.g.,* nation, state, county), and (2) geo-social relationships among individuals at discrete locations (*e.g.,* locations of mobile phone use, individual household locations, *etc.*).

3.1.2.1   Geo-Social Relationships among Geographical Areas

Geo-social relationships among geographical areas cover a wide range of topics: such as migration flows at the city scale [128], state scale [124, 158], or country scale [74]; transportation flows [12, 47, 73]; international trade among countries [56-57], sports competition among countries [4]; and so on. In an early example shown in Figure 3(a), Tobler [159] uses a network representation to describe the migration among different states in the U.S from 1965 to 1970. In work grounded in geovisualization and geovisual analytics, Guo [71] proposes an integrated interactive visualization framework that is used to effectively discover and visualize major flow patterns and multivariate relations from the county-to-county migration data in the U.S (Figure 3(b)). In complementary recent research, Wood et al. [173] propose an origins and

destinations (OD) map to preserve all origin and destination locations of the spatial layout through constructing a gridded two-level spatial treemap.

<Figure 3(a) is inserted here>
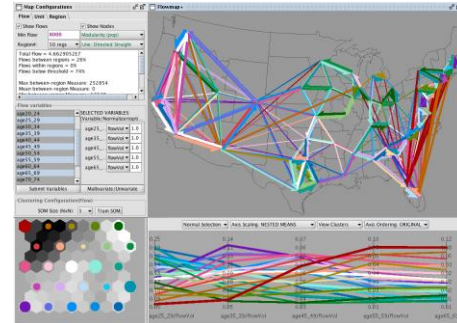
Figure 3(a)   Figure 3(b)

Figure 3(a): State to state migration from 1965 to1970. Reproduced from Figure 8 in [159]; permission to reproduce is pending the review outcome. Figure 3 (b): Flow mapping and multivariate visualization of large spatial interaction data. Figure courtesy of Diansheng Guo. For additional information on the research it is derived from, see [71].

The above work assumes that geographic location defines the spatial-social process with explicitly spatial representation and implicitly network representation, but the assumption only holds partially true for the modern interconnected world. Such representations reflect a situation in which the current integration of network analysis in GIS only focuses on a mathematical perspective that emphasizes graph theory and topology components of network [44]. Such representations do not allow users to explore the relationship between geographical space and social space, because they ignore network theory behind the network representation. Miller [121] also suggests that a Euclidian metric should not be the only function to measure closeness. Geographical proximity does not necessarily mean social closeness. Conversely, geographical long distance does not necessarily result in social isolation. From a cognitive perspective, explicitly spatial representation and implicitly network representation can mislead human intuition about social relationships among actors. Thus, it is necessary to involve explicit network representations to consider the importance of social position, social distance, and social space.

Andris [13] makes a list of benefits to involving an explicit network representation within a geographical environment: 1) Network community structure methods can identify clusters to understand the group of interconnected places as a unit rather than dense collocations; 2) Node measures (*i.e.*, degree, betweenness) can show the power of places; 3) Network system measures like degree distribution, closeness distribution, and clustering coefficients can indicate the role of any connected geographic region

over the whole system; 4) Multiple social flow layers can be added simultaneously like spatial overlay functions in a Geographical Information System (GIS) to better evaluate interaction between places; 5) Explicit network representation performs better to model the case in which spatial closeness does not correspond to stronger social flows between places.

Given the benefits with explicit network representations in a geographical environment, Luo et al. [106] introduce a spatial-social network visualization tool, the *GeoSocialApp*, that supports network, geographical, and attribute spaces in this way to allow the exploration of spatial-social networks among them (Figure 4). With explicit network spaces (in a dendrogram view and node-link view), users can have an intuitive understanding of social position, social distance, and social groups directly. For example, with the international trade network among 192 countries in 2005 as a case study, two groups identified through the dendrogram view show a core-periphery structure in the node-link view in which the red nodes are in the core and the yellow nodes are in the periphery. Since each node represents one country in the map view, the results in the map also show that the countries in the world have a hierarchical structure in which red nodes in the node-link view are economic core countries without highlight and yellow nodes are economic periphery countries with highlight. In complementary work, Thiemann [151] develops SPaTo Visual Explorer to allow the exploration of spatial-social networks with the interchange of spatial and network representation. Unlike geographical distance measures, a new shortest-path distance based on node centrality measures is implemented into SPaTo Visual Explorer [174]. This tool can easily identify the shortest social distance among different cities based on the worldwide air-transportation network.
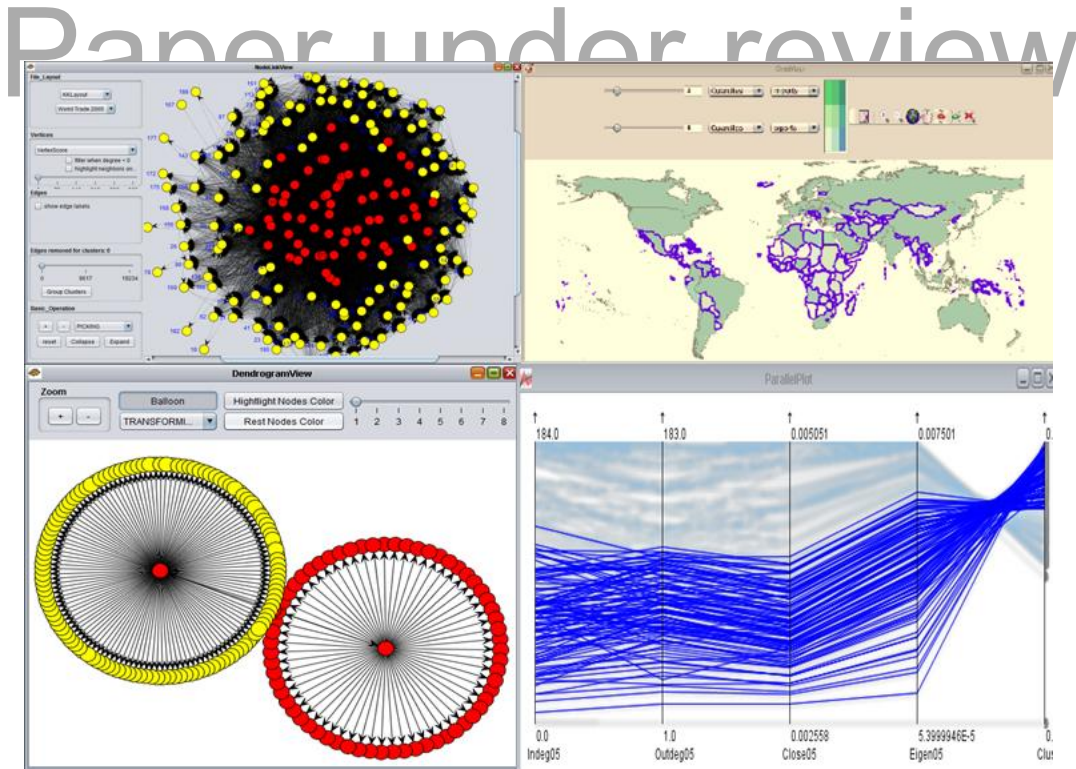
Figure 4: The analysis of international trade data among 192 countries in 2005 with GeoSocialApp: Node-Link (Upper Left), Dendrogram (Lower Left), Bivariate Choropleth Map (Upper Right) and Parallel Coordinate Plot (Lower Right). For a preliminary view of this tool, see Luo, et al [106]. Each node in the node-link view and the dendrogram view corresponds to one country in the bivariate choropleth map; the Parallel Coordinate Plot view is used to explore network attributes (i.e., indegree, outdegree) for countries. Red dots in the node-link and dendrogram view are periphery countries and these are highlight on the map and Parallel Coordinate Plot with blue outlines.

3.1.2.2   Geo-Social Relationships among Individual Locations

Geo-social relationships among individual locations include road networks [17, 46], commuting behaviours [95], location-based social networks [50], and social media networks [178]. In the geovisualization domain, Kwan and Lee [96] use GPS data collected in Lexington, Kentucky to illustrate the 3D space-time paths of women without children under 16 years of age, and find their trips mainly use highways and major arterials (Figure 5). In the geovisual analytics domain, Shen and Ma [141] create MobiVis which allows visual analytics of social and spatial information in a human interaction network over time, and they illustrate how easily this tool supports comparison of behaviour patterns of individual and group behaviours with the MIT Reality Mining Dataset [51]. Those studies, and most similar research studies, illustrate that geovisualization and geovisual analytics can reveal distinctive patterns of spatial

and social behaviours of different human interaction groups in a straightforward way [40, 82, 99, 130, 143].

<Figure 5 is inserted here>

Figure 5: Space-time paths based on GPS data collected in Lexington, Kentucky. Reproduced from Figure 9 in [96]; permission to reproduce is pending the review outcome.

Complementary to the visualization advances outlined above, computational methods in terms of spatial-social human interactions focus on developing quantitative representations of human movements. For example, with mobile phone data, Gonzalez et al. [65] and Rhee et al. [131] find that human trajectories are characterized by a regular, time independent characteristic length scale and are more attracted to more popular places, like home or work. With the circulation of bank notes in the United States, Brockmann et al. [33] find that the travelling distances of human mobility decay as a power law, and that the distribution of the time people stay in one small, spatially confined region follows algebraically long tails. Chaintreau et al. [38] and Karagiannis et al. [84] observe that inter-contact time between mobile devices shows an approximate power law in the range of 10 minutes to 1 day. Overall, all of the above studies suggest the existence of scale-free characteristics observed in most networks in which a small number of nodes have a high degree distribution and a large number of nodes have a small degree distribution [6, 19] in spatial and temporal dimension. In other words, the number of people in terms of spatial distance or inter-contact time has a scale-free distribution: a small number of people travel a long distance or have a long inter-contact time with others, whereas a large number of people travel a short distance or have a short inter-contact time with others.

Cho, Myers and Leskovec [41] further identify the impact of spatial and social factors on human movement: short-range travel is spatially and temporally periodic with little impact by the social ties, which have a strong impact on long-distance travel. Balcan et al. [18] develop a unified model to study the multiscale nature of human mobility and its relationship with epidemic spread, including airline traffic network and short-range commuting interactions. Crandall et al. [43] even develop models to quantify how likely it is that two people know each other, if they have a very close geographic distance at approximately the same time. These results open new directions for new perspectives on not only link prediction but also network dynamics with spatial, social, and even multiscale considerations. The topic of predictive analysis will be addressed directly in section 3.3 below.

Both current geo-social visual analytics and pure computational methods have advantages and disadvantages for studying human interaction related to movement in

geographic space. Geo-social visual analytics applied to the study of human movement can facilitate answering qualitative research questions in terms of "who", "where", "when", and "what". But, thus far, geo-social visual analytics has not integrated computational advances that represent the interaction of "who", "where", "when", and "what" through a mathematical description; this further limits the prediction power of existing methods. Computational approaches can explore the mathematical relationships among "who", "where", "when", and "what", but it is a challenge to give users a qualitative understanding of human movement without appropriate visual representation and interactive techniques, which limits the possibility to involve humans in the analytical process in a straightforward way. Therefore, more powerful and insightful understanding is possible by integrating visual methods with computational tools that can scale analysis to larger data volumes and formally evaluate validity of insights. It is necessary to bring advantages from visual methods and computational approaches to complement each other, which also matches the goal of visual analytics to integrate the strength from human and computer sides to support the analytical process.

3.2   Decision-Making

In addition to enabling an efficient insight gain from a complex dataset, another major application in visual analytics is to use the insight to support a decision-making process. To design a visual analytic tool to effectively support human decision-making related to the interactions among geographic and social contexts, it is important to understand how people process information and how people make decisions in real situations.

3.2.1   Conceptual Framework

Decision-making is a process to reduce uncertainty and doubt, enabling individuals to take a reasonable course of action facing complex decision problems, often in time pressure situations [75]. The process of decision-making consists of three steps: analyse the situation, find out relevant alternatives, and select an alternative by certain criteria [90]. Here, we draw upon two theoretical perspectives to frame the discussion of geo-social visual analytics for decision-making: Situation Awareness and Spatial Multicriteria Decision Analysis.

The conceptual framework from Situation Awareness (SA) can represent the decision-making process from a cognitive perspective and also integrate data exploration, decision-making and predictive analysis in the context of visual analytics. SA can be defined as "the human user's internal conceptualization of a situation" [91]. Endsley [54] defines three levels of SA: the first level is the perception of elements in the current situation, the second level is the comprehension of the current situation, and the third level is the projection of future status (Figure 6). A reasonable decision-making process

should be based on an understanding of the current situations from the first two steps of SA, and also a prediction for future situations. The first two steps also match Figure 2 in terms of a mental model building process to comprehend the current situation.

Spatial multicriteria decision analysis aims to integrate GIS and multicriteria decision making (MCDM), and both of them can provide different techniques and methodologies to transform geographical data and the decision-maker's preferences to obtain information and knowledge to support decision making [112]. More details regarding GIS-based MCDM (GIS-MCDM) can be found in Malczewski [113]. Spatial decision analysis is an inherently multicriteria decision process, involving economic, social, environmental, and political dimensions [87]. The territorial embeddedness and societal embeddedness in Figure 1 can represent the essence of multicriteria decisions in spatial decision analysis. In the section below, we propose adding another dimension into spatial multicriteria decision analysis, the social network.
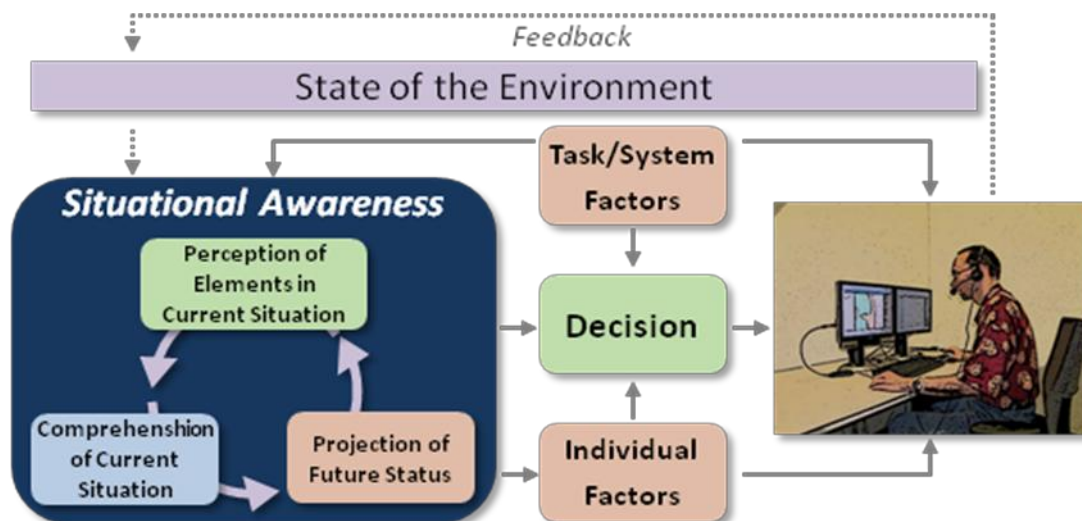
Figure 6: The process of decision-making for situation awareness. Adapted from [90].

3.2.2 Geo-Social Visual Analytics in Decision-Making

Spatial data analysis and social network analysis have their independent advantages to support decision-making. We argue here that their integration can be more powerful in support of decision-making than the sum of the parts. Limited research has been carried out thus far that achieves such integration. This section proposes two categories of geo-social integrations in terms of decision-making: the first one discusses an integrative approach toward spatial and social network factors to support decision-making process; the second one discusses how social network structures do impact GIS-based MCDM.

3.2.2.1   An Integrative Approach of Spatial and Social Network Factors

Spatial data analysis is used to detect and visualize spatial patterns (*i.e.*, disease, crime), and relate these patterns to salient explanatory covariates (*i.e.*, economic and demographic factors), and then these insights are used to give decision-making support for polices [16]. However, many social phenomena are complex systems that mainly grow from the bottom-up, while traditional spatial data analysis focuses on top-down methods that cannot deal with the question of how the phenomena being analysed evolves over space and time [22, 78].

Network science, a bottom up approach, provides the potential to link individual behaviours and interactions among individuals to the size, scale, and shape of social phenomena observed in a spatio-temporal framework [24]. The two most typical network phenomena: small-world networks (characterized by high local clustering and short average node-to-node distance) [169] and scale-free networks (in which, as noted above, a small number of nodes have a high degree distribution and a large number of nodes have a small degree distribution) [6], have shown a strong relationship with space and time. For example, the famous small world experiments to study the average path length for social networks of people observed these relationships at two geographical levels: U.S. [120] and world [167]. Additionally, small-world and scale-free properties have been demonstrated to exist in many spatial-social networks (i.e., World Wide Web graph, power grid graph, and road networks) [5, 19, 69, 103, 169, 176]. Even many traditional spatial phenomena exhibit scale-free characteristics such as city and company growth [23, 147]. Finally, as discussed in the data exploration section, the existence of scale-free characteristics has been extended from social relationships into spatial and temporal dimensions through analysing mobile phone and GPS data.

Geo-social visual analytic tools have the potential to directly enable decision-making that incorporates understanding of both geographic and social factors in an integrated way. One prototype of how such tools might work, TwitterHitter, was introduced by White and Roth [170]. The objective of TwitterHitter is to harvest information from Twitter.com to support the functions of crime analysis; these functions include decisions related to ongoing investigations as well as those related to deployment of personnel. TwitterHitter provides functions to plot a linked map-timeline view of the recent spatiotemporal activities of suspects on Twitter, and also can generate a directed network graph of the suspect's known associates (*i.e.*, Twitter friends) (Figure 7). Some other spatial data analysis methods can also be used, such as geographically weighted regression [61], to understand the etiology of the criminal activity, with the collected tweets or their attributes as potential explanatory variables in the analysis. In complementary work focused on decisions related to disease outbreaks, Guo [72] proposes a geo-social visual analytic approach to analyze large spatial human interaction data to support effective pandemic control measures. The approach includes two linked views: a reorderable matrix and a map view (Figure 8) to make it

easy to enable pattern interpretation in a geographical context and social context simultaneously. The geo-social interaction patterns provide valuable insight toward identifying critical locations and regions to suggest hypothetical control strategies for a pandemic outbreak based on synthetic population data.

<Figure 7 is inserted here>

Figure 7: Individual linked map-timeline and social network analysis views in TwitterHitter. Reproduced from Figures 3 and 4 in [170]; permission to reproduce is pending the review outcome.

<Map view is inserted here>
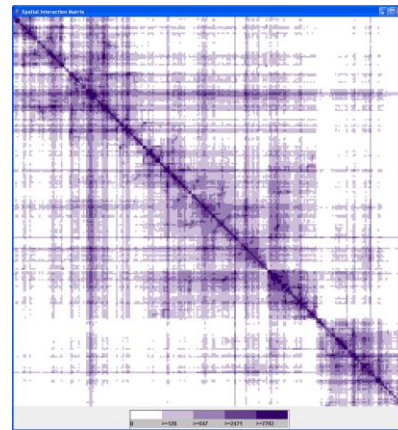
Figure 8: A map view (left) and a re-ordered matrix view to represent spatial human interaction data [72]. . The map view is reproduced from Figure 6 in [72]; permission to reproduce is pending the review outcome. The re-ordered matrix figure is courtesy of Diansheng Guo.

The prototype tools developed by White and Roth [170] and by Guo [72] illustrate that spatial data analysis and network analysis can support decision-making from different perspectives. Spatial data analysis is good at supporting decision-making in terms of what, where, when, and why questions, whereas network analysis supports who and how questions. Both analysis methods can handle what-if questions to assist decision-makers in evaluating different decision scenarios. Thus, combining both can support decision-making in terms of what, who, where, when, why, how, and what if questions. For example, spatial analysis in crime analysis demonstrates that explanatory factors relevant to spatial clusters of crime include, but may not be limited by, alcohol outlet densities [68], single person households [36], and depression [134]. Social network analysis in crime analysis has many important implications for crime investigations, such as targeting criminal leaders [175] and fighting organized crimes proactively [117] Decision-making in terms of disease control should not only require the observation of corresponding spatial patterns and driving factors behind these patterns [165], but also needs the information on how diseases are transmitted from person to person [27].

Similarly, decision-making related to evacuation (*e.g.,* in response to a hurricane threat) requires both spatial analysis related to location of people, evacuation routes, etc. as well as an understanding of how social connections impact individual evacuation decisions and behaviour [7].

3.2.2.2   The Impact of Social Network Structures on Group Decision-Making

In group/participatory settings, GIS-MCDM involves a series of activities, including defining problems, selecting evaluation criteria by group members, determining individual and collective preferences in terms of evaluation criteria and/or alternatives, sensitivity analysis with evaluation criteria and alternatives, exploring alternative combinations of individual preferences into group judgements, supporting group interaction to refine individual and group preferences, and having a final ordering of alternatives to make a compromise alternative available [102, 114]. Different stakeholders can be involved in the process of GIS-MCDM to face a variety of decision-making problems, such as environmental planning, transportation, urban planning, and so on. Social relationships among those stakeholders have significant impact on their behaviours, which further has implications for their decision abilities [29]. However, GIS-MCDM has not taken the impact of social relationships on actors' decision making into account.

The potential importance of social relationships for GIS-MCDM is illustrated in work by Bodin and Crona [29]. They review the role of social networks in terms of different relational patterns on governance process and outcomes: 1) high network density can facilitate collective action, reduce conflicts, and enhance knowledge development; 2) low degree of cohesiveness (*i.e.,* clearly distinguishable subgroups) has negative effects on collaborative processes among subgroups [70]; 3) bonding ties among subgroups is beneficial for conflict resolution and collective action; 4) high degree of network centralization is positively correlated with collective action [139]. Furthermore, this paper shows that none of the above network characteristics has a monotonically increasing positive effect on collective actions and conflict reduction, and that increasing one characteristic may cause the reduction of another. Therefore, how to maximize the positive effects of the individual and mix level of different network characteristics presents a key research challenge in terms of group decision-making.

The integration of visualization techniques into GIS-MCDM has received an increasing attention [11, 81], but those studies focus on individual decision makers rather than groups [114]. Consequently, the collaborative tasks in GIS-MCDM with visualization/visual analytics have not been explored, not to mention considering the impact of social network on decision-making. Here, we highlight a geo-social visual analytics tool developed to analyse public decision-making processes, and discuss the possibility to extend this tool into MCDM domains considering the impact of social network structures.

Paper under review

Aguirre and Nyerges [3] introduce a novel geo-social visual analytics method that they label "grapevine" (Figure 9) that is directed to analysis of the very complex geo-social information generated within applications of web-based public participation systems for participatory learning and decision-making. The authors applied the grapevine tool to analysis of data collected during a month-long, online and asynchronous citizen advisory activity focused on planning for transportation in Puget Sound. The analysis enabled by the tool allowed Aguirre and Nyerges to partially confirm a hypothesis about analytic–deliberative decision-making, "that decisions are better when they come from a combination of analysis and deliberation rather than from analysis alone." But, it also allowed them to identify key challenges in supporting deliberative processes that attempt to engage a wide cross-section of the public in deliberation that includes technical information and complex problems.

<Figure 9 is inserted here>

Figure 9: The static display of the grapevine. Reproduced from Figure 4 in [3] ; permission to reproduce is pending the review outcome.

The Grapevine tool is intended to help researchers understand the complex geo-social activities making up technology-enabled public decision-making. There are three underlying network structures in this tool: the main stem of the grapevine connects one node to another that represents users' posts; participants vote for each other's posts; participants reply to each other's posts. Aguirre and Nyerges discuss the potential to use social network analysis to understand the frequency of interactions and roles of people from a theoretical perspective, but how to use social network analysis in the real case study with the Grapevine tool has not been explored. Therefore, the Grapevine tool can be extended from three perspectives to integrate social network into GIS-MCDM in group/participatory settings for the future work. First of all, the Grapevine tool focuses on individual decision makers rather than groups. Secondly, the impact of structural social networks on decision-making reviewed by Bodin and Crona [29] can be considered. Thirdly, although the Grapevine tool is not a GIS-MCDM, it can generate individual and collective alternatives for MCDM to allow decision-makers to choose and negotiate to support collaborative tasks.

This section argues two categories of geo-social integrations in terms of decision-making, which is an iterative process that requires the support of a visual-interactive environment, especially in time-pressure situations [153]. One common problem with current geo-social visual analytic tools in terms of decision-making is that they are good at helping analysts understand the current situations from the first two steps of

Please do not cite

SA, but lack the power to make decisions relevant to the current SA or predictions for future situations.

3.3   Predictive Analysis

As discussed in the decision-making section, the SA model also provides the theoretical framework for predictive analysis in the context of visual analytics. Predictive analysis is not independent from the first two steps of the SA model: it requires the understanding of the past and current situations through data exploration. An internal conceptualization of a situation, aided by predictive models and human reasoning, is the key to predictive analysis.

Developing mathematical models to support predictive analysis starts with the understanding of patterns found in real-world data. For example, based on the common property: scale-free characteristics observed in many large networks (*i.e.,* actor collaboration graph, World Wide Web graph), Barabási and Albert [19] build a preferential attachment model to explain the development of scale-free networks in which networks tend to continue to grow with new vertices, and new vertices have a preferential attachment to vertices that are already well connected. The preferential attachment model has been used to make predictions of network growth with scale-free characteristics, but this model does not consider the impact of geographical constraints on the network growth. As discussed in the data exploration section, scale-free characteristics have been extended into spatial and temporal dimensions in terms of human mobility. Lee et al. [100] develop a new mobility model called SLAW (Self-similar Least Action Walk) that can capture all human mobility features reviewed in the data exploration section, including the Lévy flight travel patterns [33], spatial heterogeneously bounded mobility [65], power-law inter-contact times (ICTs) [38, 84], and fractal waypoints [131]. However, none of those mathematical models in terms of geo-social relationships have been implemented into geo-social visual analytics to empower prediction.

Recent studies have shown that mathematical models have a better prediction performance considering more context information (*i.e.,* spatial, network, societal) in Figure 1 rather than just one context. For example, Andris, Halverson and Hardisty [14] develop a new model considering physical and social space for predicting future migration, and the model outperforms a gravity model considering physical space alone with U.S. Migration flows among major cities. In complementary work, Takhteyev et al. [150] find that pre-existing ties (*i.e.,* frequency of air travel) between places and people is the best predictor of Twitter ties compared to three other spatial and social factors including geographic distance, national boundaries, and language. A related study shows that using place-based attributes (*i.e.,* social, economic and ecological context) can successfully predict community membership more than 70% of the time in a large social network of cell phone towers [37].

Geo-social visual analytics in terms of predictive analysis implements mathematical models in a visual-interactive environment to allow users to select appropriate methods, to set parameters, to interpret results, to understand what to do next, and draw conclusions based on different scenarios. For example, Brigantic et al. [32] introduce a visual analytic tool (PanViz) with metapopulation-based epidemic models to rapidly assess alternative mitigation strategies in terms of pandemic influenza to give decision makers support. The PanViz system was subsequently extended and deployed in the Indiana State Department of Health Planning (Figure 10) to support analysis of potential epidemic control strategies [111]. While PanViz demonstrates the potential of geovisual analytics, it does not explicitly include capabilities to incorporate social network information into the analysis. Bisset and Marathe [28] have developed a similar tool that does include such capabilities: EPISIMS with individual-based epidemic models to simulate the dynamics of millions of individuals, traffic of entire cities, and disease spread, respectively. In related research, Broeck et al [34] present a visual analytic tool "GLEaMviz" available to the public that allows the user to set a variety of parameters to simulate the human-to-human infectious disease spread across the world.
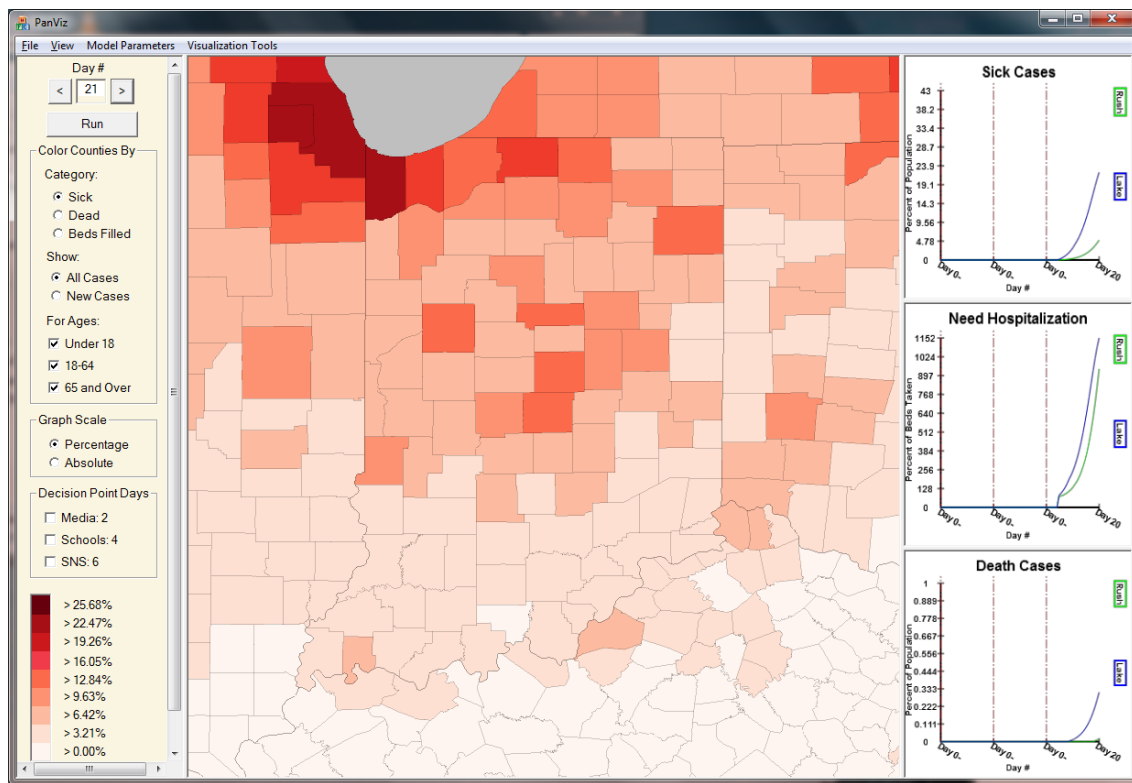
Figure 10: PanViz Applied at a State Level. Figure Courtesy of Ross Maciejewski. For Additional Information on the Research It Is Derived From, See [111].

One big challenge in terms of predictive analysis is that geo-social systems are highly sensitive to social adaptive behaviours [164]. In crisis situations (*i.e.*, pandemics, natural disasters), the geo-social systems behave abnormally which is still a challenge to predict. Vespignani [164] proposes three challenges for future work to predict social adaptive behaviours in the period of crisis: collecting data on information spread and social reactions in the period of crisis; developing models to quantify the effect of risk perception and awareness phenomena of individuals; and deploying monitoring infrastructures capable of informing computational models in real time. The three suggestions also need geo-social visual analytics to support methods to address all three challenges in an integrative system.

In research directly focused on predictive analytics for crisis events, Bengtsson et al. [25] focus on displacement after the 2010 Haiti Earthquake. Specifically, they track population with phone call data from six weeks before the disaster to five months after. Their estimates in terms of the number, the timing, and geographical distribution of population movements correspond well with a retrospective survey. A follow-up study identifies that the destinations in which people stayed had significantly social bonds and the time that the displaced population stays outside the city follows a skewed, fat-tailed distribution [105]. Bengtsson and colleagues have taken the first step to demonstrate that the prediction of population movements with the use of phone data in disaster response is possible, but substantial future work is still required before common usage. For example, cell phone data availability during natural disasters is still limited [25]; data coverage varies over space, time, and different groups of people [64]; data privacy is always a big concern [97].

Above all, the process of predictive analysis is an iterative process that needs human interpretation, control, experiences, and imagination, especially in crisis periods. As discussed in the decision-making section, current geo-social visual analytic tools in terms of decision-making are good at enabling understanding of the current situations from the first two steps of SA, but lack the power to make predictions for future situations. Meanwhile, there is a gap between the insight gained in the process of data exploration and predictive analysis. For example, the current predictive epidemic visual analytic tools implement epidemic models to simulate disease transmission and design corresponding control strategies without flexible approaches analysing human spatial-social interactive clusters like Figure 8, as knowledge input to improve control strategy design.

4   Discussion

Geo-social data do not make any sense when being abstracted from their appropriate contexts [1]. Geo-social contexts do not only demonstrate conceptual and observed overlaps, but also shed light on data from different perspectives. Downs and DeSouza

[49] argue that spatial thinking serves three purposes: 1) a descriptive function, 2) an analytic function, and 3) an inferential function. We follow Downs and DeSouza's lead and propose extending consideration of spatial thinking into spatial-social thinking. Social thinking describes how human interactions impact their ideas, emotions, and behaviours [172]. A social network approach provides a means to describe human interactions, analyse and infer how such interactions impact social thinking process. Therefore, the three purposes that spatial thinking serves can be applied to social thinking to develop integrative spatial-social thinking for people to acquire knowledge. The three purposes also match the primary goal of visual analytics: interactive visual interfaces should support human analytical reasoning in an efficient and effective way. Supporting spatial and social thinking to address the three purposes outlined is a prototypical example of meeting this goal.

Based on the above discussion, geo-social visual analytics has multiple research questions that need to be addressed in relation to data exploration, decision-making, and predictive analysis. In addition, data exploration, decision-making, and predictive analysis have inner connections: knowledge development in the data exploration process can support decision-making and predictive analysis, which sometimes leads decision-making and predicative analysis to develop a new knowledge construction process. Therefore, the overarching goal of visual analytics is to involve human ability with the whole complex analytical process rather than each step separately [8].

5   Challenges

Our analysis of the literature provides several potential research directions for further investigation of geo-social visual analytics. We conclude this review paper by highlighting nine core challenges that will require interdisciplinary efforts to meet. Each challenge is outlined below.

5.1   *Developing a framework to collect geo-social relationship data and assess their fitness for different applications while also considering the potential negative consequences for human privacy of collecting these data.* Geo-social analytics, especially during natural disasters, provides additional motivation for future collection of network and spatial data. Most popular geo-social network data collection methods include surveys [129], crawling social media sites [138], collecting data from mobile devices such as cell phones [52], and leveraging wireless sensor technology [85]. Details about each method and their pros and cons can be found in Salathé et al [137]. A key problem here is that there is no theory/framework to assess whether the collected data are suitable to study different applications. For example, Salathé and Jones [136] study disease transmission at individual-level through building social interaction networks.

Nodes represent individuals and edges consider both friendships in Facebook and physical proximity in real world (*i.e.*, the same dorm, the same class). However, the demographics of social media users is a biased sample of the whole population (in relation to age, gender, race, etc) and such networks are still at a rather coarse resolution for the study of disease transmission. As more complete data sets become available, individual-level network data with spatial and temporal information may make it possible to predict human behaviours better, but collection of individual level data raises a range of privacy concerns [31].

5.2 *Understanding the interaction of geography and network space, as well as their respective and coupled impacts on outcomes of interest.* Recent research has found that spatial proximity and network relationships matter on outcomes of interest [150, 163]. For example, Onnela et al. [123] study the spatial/geographical constraints on social interactions, and find that space shapes social groups in the way that small social groups are geographically tight, but become clustered with group size exceeding beyond about 30 members. These results suggest that spreading processes face different network and spatial constraints, and further research is required to investigate how geographical and social relationships exactly operate in different geo-social systems.

5.3 *Developing theory, methods, and tools to consider spatial and network factors simultaneously.* Most geo-social analytical approaches use independent traditional spatial analysis and social network analysis methods simultaneously to explore the same datasets [53]. It is necessary to develop new theory and methods that integrate spatial and social factors together. For example, Radil et al. [127] propose a new method that borrows the concept of social position to explore an actor's position in a spatial contiguity matrix simultaneously with his or her position in social networks. The proposed method can identify statistically significant violence patterns which cannot be captured by the classical spatial autocorrelation method: Global Moran's I [122, 156]. Radil and his colleagues' work is only one attempt to explore new geo-social theory and methods. As discussed in this paper, spatial analysis and social network analysis exhibit strong conceptual and observed overlaps, so much more future work is needed.

5.4 *How does spatial decision-making promote the transition of daily life habits, and further impact local communities and networks?* As reviewed in the decision-making section, GIS-MCDM includes broad application domains (e.g., transportation, urban planning). Decision-making in terms of those domains involves interests among different stakeholders, but it also has a corresponding influence on the practices of everyday life. The influence will further impact local community interaction and network structures. For example, people who used to

live in *hutong* (one mega-block high-rise neighbourhood) in Beijing report a substantial disruption of the high quality and frequency of local interaction they had in *hutong* compared to that after they were relocated to apartment complexes on the city periphery when their neighbourhood underwent urban renewal [133]. Tita et al. [155] point to a similar impact of urban redevelopment of on social networks in their argument that the clear north-south geographical division in the gang rivalry networks in a section of Los Angeles is due to a landscape feature: the San Bernadino Freeway. Although spatial decision making has a significant role in shifting local community network structures, research in GIS-MCDM has not taken such shift into account.

5.5   *How do social media change the traditional decision-making approaches?* The emergence of social media (*i.e.,* Facebook, Twitter, LinkedIn) changes the world via collective power through on-line social network. One person can communicate with hundreds or even more people about products, news, cultures, and any information. The communication occurs in a smaller world, because Kwak et al. [92] find that the average path length of a Twitter network is 4.12 compared to "six degrees of separation"[120] in the real world. The impact of people-to-people communication has greatly changed the traditional sense of decision-making, because social media based conversations help people to be accountable and occur outside of the direct control of decision-makers [115]. For example, social media is playing an increasing role in the most recent anti-government protests, including the Arab Spring, Occupy Wall Street, and the London Riots [161]. MacEachren et al. [110] leverage Twitter into a web-enabled geovisual analytics application and discuss how social media can offer strategies for disaster and emergency management. While social media have a transformative impact on the traditional decision-making approaches, how responding organizations can successfully leverage these technologies is just beginning to be considered [177]. Before good use of social media in decision-making can be achieved, there are many unexplored research questions. For example, how does information diffuse via social media? How do social media change human behaviours in normal and crisis situations? How do social media transform individual voice into collective power to be accountable?

5.6   *Developing new geo-social visual analytics methods to incorporate data exploration, decision-making, and predictive analysis as a whole.* As discussed above, most geo-social visual analytics methods only support one step, so visual analytics cannot effectively transform knowledge through visual exploration into complex analytical strategies directly. One possible solution is to improve inter-disciplinary cooperation through understanding the human analytical reasoning of real decision-makers to design visual analytic tools accordingly [9], such as has been attempted for maritime anomaly detection [132], bridge management

system analysis [166], and other application domains. In addition, visual analytics have not synergistically integrated computational methods to maximize human conceptual, perceptual and reasoning capabilities in the whole scientific and problem-solving process. One possible theoretical framework to link the whole complex analytical process can be found in Gahegan [62], situating human reasoning, concretized representation, conceptual structures, visualization representation, and mathematical models into the whole science process. To support a full range of applications of geo-social visual analytics, however, the approach must be generalized beyond the context of scientific research, which was the target of Gahegan's model.

5.7   *Integrating distinct applications of cognitive science to support geo-social visual analytics.* Cognitive science provides theoretical frameworks for the design of geovisualization tools [108], it provides a conceptual approach (e.g., distributed cognition) to understand human reasoning as enabled by visual tools, and it also offers fundamental theories and approaches to understand and model human behaviors in network science. For example, the Organizational Risk Analyzer (ORA) uses both network theory and social psychology to model human behaviors, and ORA has been used to analyze 1500 videos made by insurgents in Iraq and effectively reduce sniper activity by 70% [30]. Therefore, cognitive science should not only be used to design visualization tools, but to support geo-social analytic models as well.

5.8   *Understanding the dynamics of geo-social relationships and processes.* Geo-social relationships are not static but dynamic, so it is important to understand the change of the relationships over time and compare the dynamic change to the static understanding [2] as well as to investigate the linked geographic and social process that drive the change. Hess [76] also discusses the need to involve a temporal concept into the proposed categories of embeddedness (Figure 1) through taking into account developments over time and changes in the spatial configuration of networks at different scales. We argue that three models can be developed to represent temporal geo-social relationships among geographical or individual units. The first model has fixed nodes with unchanging geographical relationships and varying social relationships. The model is based on the geo-social phenomena in which geographical relationships are fixed because of their relative geographical locations (*e.g.*, cities, states, and nations), and social relationships change over time. Those geo-social phenomena include the dynamic trade network at a country-to-country scale [55, 180], human migration in the U.S. at a county-to-county scale [71], and technology adoption (i.e., Twitter) at a city-to-city scale [160]. The second model has fixed nodes with changing geographical relationships and relatively fixed social relationships (i.e., human movement, mobile-based social media). The last model can support more

complex geo-social dynamic behaviours in which networks can expand and recede [22]. Recent developments that apply the concept of "rendezvous" (bringing sensors close to one another in space or time [79] to shed light on human mobility characteristics [80]) provides data and methods to support the last two models.

5.9 *A new concept of scale?* Scale is a fundamental concept in geography. Scale-free distributions in social networks also exhibit self-similarity on all length scales [145]. As discussed above, many spatial-social phenomena show self-similarity observed across different spatial scales. Barabási [21] claims that although there are diverse dynamical processes on networks (*i.e.,* the spread of viruses and ideas on physical space network and the flow of information over cyberspace network), it is possible that these dynamical processes share some common characteristics. Based on this argument and the relationships between geography and networks discussed in this paper, we propose that it is important to ask what are the spatial factors behind those common characteristics; are those spatial factors scale-dependent? Through understanding geo-social relationships, is it possible to reconceptualize the concept of spatial scale as constructed by relational spatial units (*i.e.*, human, company, city, country)? Does the cyber world bring a new life to the concept of scale? What are the common and different characteristics between the scale in the physical world and the cyber world? How does information spread in a cyberspace network interact with information spread in the physical space network?

To sum up, geo-social visual analytics is based on the conceptual extension of the First Law of Geography: Everything is related to everything else, but near things are more related than distant things [157]; Nearness and relationship can be considered a matter of geographical and social network distance, relationship and interaction. At the methodological level, geo-social visual analytics should facilitate answering research and application questions in terms of "who", "where", "when", "what", "why", "how", and "what if". Therefore, geo-social visual analytics should rely on human reasoning frameworks to effectively integrate cognitive science, network visualization, geovisualization, spatial data mining, network data mining, spatial analysis, social network analysis, and new geo-social analytical approaches to support the whole scientific and problem-solving process.

6 Acknowledgements

7 References

[1] ABBOTT, A. Of time and space: The contemporary relevance of the Chicago School. *Social Forces 75*, 4 (1997), 1149-1182. doi:10.1093/sf/75.4.1149.

[2] ADAMS, J., FAUST, K. and LOVASI, G. S. Capturing context: Integrating spatial and social network analyses. *Social networks 34*, 1 (2012), 1-5. doi:10.1016/j.socnet.2011.10.007.

[3] AGUIRRE, R. and NYERGES, T. Geovisual evaluation of public participation in decision making: The grapevine. *Journal of Visual Languages & Computing 22*, 4 (2011), 305-321. doi:10.1016/j.jvlc.2010.12.004.

[4] AHMED, A., FU, X., HONG, S. H., NGUYEN, Q. H. and XU, K. Visual Analysis of History of World Cup: A Dynamic Network with Dynamic Hierarchy and Geographic Clustering. In *Visual Information Communication*, M. L. Huang, Q. V. Nguyen and K. Zhang, Eds. Springer, 2010, pp. 25-39. doi: 10.1007/978-1-4419-0312-9_2.

[5] ALBERT, R., ALBERT, I. and NAKARADO, G. L. Structural vulnerability of the North American power grid. *Physical Review E 69*, 2 (2004), 025103. doi:10.1103/PhysRevE.69.025103.

[6] ALBERT, R., JEONG, H. and BARABASI, A. Diameter of the world wide web. *Nature 401*, 6749 (1999), 130-131. doi:10.1038/43601.

[7] ALSNIH, R. and STOPHER, P. R. Review of procedures associated with devising emergency evacuation plans. *Transportation Research Record: Journal of the Transportation Research Board 1865*, 1 (2004), 89-97. doi:10.3141/1865-13.

[8] ANDRIENKO, G., ANDRIENKO, N., JANKOWSKI, P., KEIM, D., KRAAK, M., MACEACHREN, A. and WROBEL, S. Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science 21*, 8 (2007), 839-858. doi:10.1080/13658810701349011.

[9] ANDRIENKO, G., ANDRIENKO, N., KEIM, D., MACEACHREN, A. M. and WROBEL, S. Challenging Problems of Geospatial Visual Analytics (editorial introduction). *Journal of Visual Languages & Computing 22*, 4 (2011), 251-256. doi:10.1016/j.jvlc.2011.04.001.

[10] ANDRIENKO, G., ANDRIENKO, N., KOPANAKIS, I., LIGTENBERG, A. and WROBEL, S. Visual analytics methods for movement data. In *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery* F. Giannoni and D. Pedreski, Eds. Springer Berlin Heidelberg. New York, 2008, ch. 13, pp. 375-410. doi: 10.1007/978-3-540-75177-9_14.

[11] ANDRIENKO, N. and ANDRIENKO, G. Informed spatial decisions through coordinated views. *Information Visualization 2*, 4 (2003), 270-285. doi:10.1057/palgrave.ivs.9500058.

[12] ANDRIENKO, N. and ANDRIENKO, G. Spatial generalisation and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics 17*, 2 (2010), 205-219. doi:10.1109/TVCG.2010.44.

[13] ANDRIS, C. *Metrics and methods for social distance*. Massachusetts Institute of Technology, 2011.

[14] ANDRIS, C., HALVERSON, S. and HARDISTY, F. Predicting migration system dynamics with conditional and posterior probabilities. In *Spatial Data Mining and*

Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on (Fuzhou, China, 2011), IEEE, pp. 192-197. doi: 10.1109/ICSDM.2011.5969030.

[15] ASHDOWN, P. *The global power shift*. TED, http://www.ted.com/playlists/73/the_global_power_shift.html, 2012.

[16] BAILEY, T. C. and GATRELL, A. C. *Interactive spatial data analysis*. Longman Scientific & Technical Essex, 1995.

[17] BAK, P., OMER, I. and SCHRECK, T. Visual analytics of urban environments using high-resolution geographic data. In *Geospatial Thinking Lecture Notes in Geoinformation and Cartography* M. Painho, M. Y. Santos and H. Pundt, Eds. Springer Berlin Heidelberg, 2010, pp. 25-42. doi: 10.1007/978-3-642-12326-9_2.

[18] BALCAN, D., COLIZZA, V., GON ALVES, B., HU, H., RAMASCO, J. J. and VESPIGNANI, A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences 106*, 51 (2009), 21484-21491. doi:10.1073/pnas.0906910106.

[19] BARAB SI, A. L. and ALBERT, R. Emergence of scaling in random networks. *Science 286*, 5439 (1999), 509-512.

[20] BARABASI, A. L. *Linked: How everything is connected to everything else and what it means*. Perseus, Cambridge, MA, 2002.

[21] BARABASI, A. L. Scale-free networks: a decade and beyond. *Science 325*, 5939 (2009), 412-413. doi:10.1126/science.1173299.

[22] BATTY, M. Network geography: Relations, interactions, scaling and spatial processes in GIS. *Re-presenting GIS* (2005), 149-170.

[23] BATTY, M. Rank clocks. *Nature 444*, 7119 (2006), 592-596. doi:10.1038/nature05302.

[24] BATTY, M. The size, scale, and shape of cities. *Science 319*, 5864 (2008), 769-771. doi:10.1126/science.1151419.

[25] BENGTSSON, L., LU, X., THORSON, A., GARFIELD, R. and VON SCHREEB, J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Medicine 8*, 8 (2011), e1001083. doi:10.1371/journal.pmed.1001083.

[26] BERA, R. and CLARAMUNT, C. Topology-based proximities in spatial systems. *Journal of Geographical Systems 5*, 4 (2003), 353-379. doi:10.1007/s10109-003-0115-y.

[27] BIAN, L. and LIEBNER, D. A network model for dispersion of communicable diseases. *Transactions in GIS 11*, 2 (2007), 155-173. doi:10.1111/j.1467-9671.2007.01039.x.

[28] BISSET, K. and MARATHE, M. A cyber environment to support pandemic planning and response. *DOE SciDAC Review Magazine* (2009).

[29] BODIN, Ö. and CRONA, B. I. The role of social networks in natural resource governance: What relational patterns make a difference? *Global Environmental Change 19*, 3 (2009), 366-374. doi:10.1016/j.gloenvcha.2009.05.002.

[30] BOHANNON, J. Counterterrorism's new tool:'metanetwork'analysis. *Science 325*, 5939 (2009), 409-411. doi:10.1126/science.325_409.

[31] BOHANNON, J. Investigating networks: the dark side. *Science 325*, 5939 (2009), 410-411. doi:10.1126/science.325_410.

[32] BRIGANTIC, R., EBERT, D., CORLEY, C., MACIEJEWSKI, R., MULLER, G. and TAYLOR, A. Development of a Quick Look Pandemic Influenza Modeling and Visualization Tool. In *Proceedings of the 7th International ISCRAM Conference* (Seattle, USA, 2010).

[33] BROCKMANN, D., HUFNAGEL, L. and GEISEL, T. The scaling laws of human travel. *Nature 439*, 7075 (2006), 462-465. doi:10.1038/nature04292.

[34] BROECK, W. V., GIOANNINI, C., GON ALVES, B., QUAGGIOTTO, M., COLIZZA, V. and VESPIGNANI, A. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infectious Diseases 11*, 1 (2011), 37. doi:10.1186/1471-2334-11-37.

[35] BUCHANAN, M. *Nexus: small worlds and the groundbreaking theory of networks*. W.W. Norton, New York, 2003.

[36] CAHILL, M. and MULLIGAN, G. Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review 25*, 2 (2007), 174-193. doi:10.1177/0894439307298925.

[37] CAUGHLIN, T. T., RUKTANONCHAI, N., ACEVEDO, M. A., LOPIANO, K. K., PROSPER, O., EAGLE, N. and TATEM, A. J. Place-Based Attributes Predict Community Membership in a Mobile Phone Communication Network. *PLoS ONE 8*, 2, e56057. doi:10.1371/journal.pone.0056057.

[38] CHAINTREAU, A., HUI, P., CROWCROFT, J., DIOT, C., GASS, R. and SCOTT, J. Impact of human mobility on opportunistic forwarding algorithms. In *IEEE Transactions on Mobile Computing* (Barcelona, Spain, 2007), IEEE Computer Society, pp. 606-620.

[39] CHANG, R., ZIEMKIEWICZ, C., GREEN, T. and RIBARSKY, W. Defining insight for visual analytics. *IEEE Computer Graphics and Applications 29*, 2 (2009), 14-17. doi:10.1109/MCG.2009.22.

[40] CHEN, J., SHAW, S. L., YU, H., LU, F., CHAI, Y. and JIA, Q. Exploratory data analysis of activity diary data: a space-time GIS approach. *Journal of Transport Geography 19*, 3 (2011), 394-404. doi:10.1016/j.jtrangeo.2010.11.002.

[41] CHO, E., MYERS, S. A. and LESKOVEC, J. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2011), ACM, pp. 1082-1090. doi: 10.1145/2020408.2020579.

[42] CHRISTAKIS, N. A. and FOWLER, J. H. *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown and Company, 2009.

[43] CRANDALL, D. J., BACKSTROM, L., COSLEY, D., SURI, S., HUTTENLOCHER, D. and KLEINBERG, J. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences 107*, 52 (2010), 22436-22441. doi:10.1073/pnas.1006155107.

[44] CURTIN, K. M. Network analysis in geographic information science: Review, assessment, and projections. *Cartography and Geographic Information Science 34*, 2 (2007), 103-111. doi:10.1559/152304007781002163.

[45] DARAGANOVA, G., PATTISON, P., KOSKINEN, J., MITCHELL, B., BILL, A., WATTS, M. and BAUM, S. Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. *Social networks 34*, 1 (2012), 6-17. doi:10.1016/j.socnet.2010.12.001.

[46] DEMŠAR, U., ŠPATENKOV , O. and VIRRANTAUS, K. Identifying Critical Locations in a Spatial Network with Graph Theory. *Transactions in GIS 12*, 1 (2008), 61-82. doi:10.1111/j.1467-9671.2008.01086.x.

[47] DEMŠAR, U. and VIRRANTAUS, K. Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science 24*, 10 (2010), 1527-1542. doi:10.1080/13658816.2010.511223.

[48] DOREIAN, P. and CONTI, N. Social context, spatial structure and social network structure. *Social networks 34*, 1 (2012), 32-46. doi:10.1016/j.socnet.2010.09.002.

[49] DOWNS, R. and DESOUZA, A. *Learning to think spatially: GIS as a support system in the K-12 curriculum*. National Research Council and National Academies Press, Washington, DC, 2006.

[50] DOYTSHER, Y., GALON, B. and KANZA, Y. Storing routes in socio-spatial networks and supporting social-based route recommendation. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks* (Chicago, Illinois, 2011), ACM, pp. 49-56. doi: 10.1145/2063212.2063219.

[51] EAGLE, N. and PENTLAND, A. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing 10*, 4 (2006), 255-268. doi:10.1007/s00779-005-0046-3.

[52] EAGLE, N., PENTLAND, A. S. and LAZER, D. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences 106*, 36 (2009), 15274-15278. doi:10.1073/pnas.0900282106.

[53] EMCH, M., ROOT, E. D., GIEBULTOWICZ, S., ALI, M., PEREZ-HEYDRICH, C. and YUNUS, M. Integration of Spatial and Social Network Analysis in Disease Transmission Studies. *Annals of the Association of American Geographers 102*, 5 (2012), 1004-1015. doi:10.1080/00045608.2012.671129.

[54] ENDSLEY, M. R. Theoretical underpinnings of situation awareness: A critical review. In *Situation awareness analysis and measurement*, M. R. Endsley and D. J. Garland, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2000, pp. 3-32.

[55] FAGIOLO, G. The international-trade network: gravity equations and topological properties. *Journal of Economic Interaction and Coordination 5*, 1 (2010), 1-25. doi:10.1007/s11403-010-0061-y.

[56] FAGIOLO, G., REYES, J. and SCHIAVO, S. On the topological properties of the world trade web: A weighted network analysis. *Physica A: Statistical Mechanics and its Applications 387*, 15 (2008), 3868-3873. doi:10.1016/j.physa.2008.01.050.

[57] FAGIOLO, G., REYES, J. and SCHIAVO, S. World-trade web: Topological properties, dynamics, and evolution. *Physical Review E 79*, 3 (2009), 0361151-03611519. doi:10.1103/PhysRevE.79.036115.

[58] FESTINGER, L., SCHACHTER, S. and KURT, W. *Social Pressures in Informal Groups*. Stanford University Press, Stanford, CA, 1950.

[59] FLINT, C. The theoretical and methodological utility of space and spatial statistics for historical studies: the Nazi Party in geographic context. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 35*, 1 (2002), 32-42. doi:10.1080/01615440209603142.

[60] FLINT, C., DIEHL, P., SCHEFFRAN, J., VASQUEZ, J. and CHI, S. Conceptualizing ConflictSpace: Toward a Geography of Relational Power and Embeddedness in the Analysis of Interstate Conflict. *Annals of the Association of American Geographers 99*, 5 (2009), 827-835. doi:10.1080/00045600903253312.

[61] FOTHERINGHAM, A., BRUNSDON, C. and CHARLTON, M. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons Inc, 2002.

[62] GAHEGAN, M. Beyond tools: visual support for the entire process of GIScience. In *Exploring Geovisualization*, J. Dykes, A. MacEachren and M. Kraak, Eds, 2005, ch. 4, pp. 83-99.

[63] GANTER, J. and MACEACHREN, A. M. *Cognition and the Design of Scientific Visualization Systems*. Department of Geography, Pennsylvania State University, Unpublished manuscript, 1989.

[64] GETHING, P. W. and TATEM, A. J. Can Mobile Phone Data Improve Emergency Response to Natural Disasters? *PLoS Medicine 8*, 8 (2011), e1001085. doi:10.1371/journal.pmed.1001085.

[65] GONZALEZ, M. C., HIDALGO, C. A. and BARAB SI, A. L. Understanding individual human mobility patterns. *Nature 453*, 7196 (2008), 779-782. doi:10.1038/nature06958.

[66] GOODCHILD, M., ANSELIN, L., APPELBAUM, R. P. and HARTHORN, B. H. Toward spatially integrated social science. *International Regional Science Review 23* (2000), 139-159. doi:10.1177/016001760002300201.

[67] GOODCHILD, M. and JANELLE, D. Toward critical spatial thinking in the social sciences and humanities. *GeoJournal 75*, 1 (2010), 3-13. doi:10.1007/s10708-010-9340-3.

[68] GORMAN, D. M., SPEER, P. W., GRUENEWALD, P. J. and LABOUVIE, E. W. Spatial dynamics of alcohol availability, neighborhood structure and violent crime. *Journal of Studies on Alcohol and Drugs 62*, 5 (2001), 628-636.

[69] GOVINDAN, R. and TANGMUNARUNKIT, H. Heuristics for Internet map discovery. In *INFOCOM 2000* (Tel Aviv, 2000), pp. 1371-1380 doi: 10.1109/INFCOM.2000.832534.

[70] GRANOVETTER, M. S. The strength of weak ties. *American Journal of Sociology 78*, 6 (1973), 1360-1380. doi:10.1086/225469.

[71] GUO, D. Flow mapping and multivariate visualization of large spatial interaction data. *Visualization and Computer Graphics, IEEE Transactions on 15*, 6 (2009), 1041-1048. doi:10.1109/TVCG.2009.143.

[72] GUO, D. Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science 21*, 8 (2007), 859-877. doi:10.1080/13658810701349037.

[73] GUO, D., LIU, S. and JIN, H. A graph-based approach to vehicle trajectory analysis. *Journal of Location Based Services 4*, 3 (2010), 183-199. doi:10.1080/17489725.2010.537449.

[74] GUO, D., WU, K., ZHANG, Z. and XIANG, W. WMS-Based Flow Mapping Services. In *2012 IEEE Eighth World Congress on Services* (Honolulu, HI, 2012), IEEE Eighth World Congress on Services, pp. 234-241. doi: 10.1109/SERVICES.2012.37.

[75] HARRIS, R. *Introduction to Decision Making*. http://www.virtualsalt.com/crebook5.htm, 1998.

[76] HESS, M. "Spatial" relationships? Towards a reconceptualization of embeddedness. *Progress in human geography 28*, 2 (2004), 165-186. doi:10.1191/0309132504ph479oa.

[77] HIPP, J. R., FARIS, R. W. and BOESSEN, A. Measuring 'neighborhood': Constructing network neighborhoods. *Social networks 34*, 1 (2012), 128-140. doi:10.1016/j.socnet.2011.05.002.

[78] HOLLAND, J. H. *Hidden order: How adaptation builds complexity*. Perseus Books, Cambridge, MA, 1996.

[79] HONICKY, R. Understanding and Using Rendezvous to Enhance Mobile Crowdsourcing Applications. *Computer 44*, 6 (2011), 22-28. doi:10.1109/MC.2011.129.

[80] HONICKY, R. E. *Towards a societal scale, mobile sensing system*. UC Berkeley, 2011.

[81] JANKOWSKI, P., ANDRIENKO, N. and ANDRIENKO, G. Map-centred exploratory approach to multiple criteria spatial decision making. *International Journal of Geographical Information Science 15*, 2 (2001), 101-127.

[82] JIA, T. and JIANG, B. Exploring Human Activity Patterns Using Taxicab Static Points. *ISPRS International Journal of Geo-Information 1*, 1 (2012), 89-107. doi:10.3390/ijgi1010089.

[83] JOHNSON, S. *Emergence: The connected lives of ants, brains, cities, and software*. Scribner, New York, 2012.

[84] KARAGIANNIS, T., LE BOUDEC, J. Y. and VOJNOVI , M. Power law and exponential decay of intercontact times between mobile devices. *IEEE Transactions on Mobile Computing 9*, 10 (2010), 1377-1390. doi:10.1109/TMC.2010.99.

[85] KAZANDJIEVA, M., LEE, J. W., SALATH , M., FELDMAN, M. W., JONES, J. H. and LEVIS, P. Experiences in Measuring a Human Contact Network for Epidemiology Research. In *ACM Workshop on Hot Topics in Embedded Networked Sensors (HotEmNets)* (New York, NY, USA, 2010), ACM. doi: 10.1145/1978642.1978651.

[86] KEIM, D., KOHLHAMMER, J., ELLIS, G. and MANSMANN, F. *Mastering the information age: solving problems with visual analytics*. Eurographics Association, Goslar, Germany, 2011.

[87] KIKER, G. A., BRIDGES, T. S., VARGHESE, A., SEAGER, T. P. and LINKOV, I. Application of multicriteria decision analysis in environmental decision making. *Integrated environmental assessment and management 1*, 2 (2009), 95-108. doi:10.1897/IEAM_2004a-015.1.

[88] KNIGGE, L. D. and COPE, M. Grounded visualization: integrating the analysis of qualitative and quantitative data through grounded theory and visualization. *Environment and Planning A 38*, 11 (2006), 2021-2037. doi:10.1068/a37327.

[89] KOCHEN, M. *The small world*. Ablex Norwood, NJ, 1989.

[90] KOHLHAMMER, J., MAY, T. and HOFFMANN, M. Visual Analytics for the Strategic Decision Making Process. In *GeoSpatial Visual Analytics*, R. D. Amicis, R. Stojanovic and G. Conti, Eds. Springer Netherlands, 2009, pp. 299-310. doi: 10.1007/978-90-481-2899-0_23.

[91] KOHLHAMMER, J. and ZELTZER, D. DCV: a decision-centered visualization system for time-critical applications. In *IEEE International Conference on Systems, Man and Cybernetics.* (2003), pp. 3905-3911 doi: 10.1109/ICSMC.2003.1244498.

[92] KWAK, H., LEE, C., PARK, H. and MOON, S. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* (New York, NY, USA, 2010), ACM pp. 591-600. doi: 10.1145/1772690.1772751.

[93] KWAN, M. Mobile Communications, Social Networks, and Urban Travel: Hypertext as a New Metaphor for Conceptualizing Spatial Interaction*. *The Professional Geographer 59*, 4 (2007), 434-446. doi:10.1111/j.1467-9272.2007.00633.x.

[94] KWAN, M. P. Feminist visualization: Re-envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers 92*, 4 (2002), 645-661. doi:10.1111/1467-8306.00309.

[95] KWAN, M. P. Gender, the Home-Work Link, and Space-Time Patterns of Nonemployment Activities. *Economic geography 75*, 4 (1999), 370-394. doi:10.2307/144477.

[96] KWAN, M. P. and LEE, J. Geovisualization of human activity patterns using 3D GIS: a time-geographic approach. In *Spatially integrated social science*, M.F.Goodchild and D.G.Janelle, Eds. Oxford University Press. New York, 2004, pp. 48-66.

[97] LANDWEHR, C., BONEH, D., MITCHELL, J. C., BELLOVIN, S. M., LANDAU, S. and LESK, M. Privacy and Cybersecurity: The Next 100 Years. *Proceedings of the IEEE 100*, 13 (2012), 1659-1673. doi:10.1109/JPROC.2012.2189794.

[98] LAZER, D., PENTLAND, A. S., ADAMIC, L., ARAL, S., BARABASI, A. L., BREWER, D., CHRISTAKIS, N., CONTRACTOR, N., FOWLER, J. and GUTMANN, M. Life in the network: the coming age of computational social science. *Science 323*, 5915 (2009), 721-723. doi:10.1126/science.1167742.

[99] LEE, J. Y. and KWAN, M. Visualisation Of Socio-Spatial Isolation Based On Human Activity Patterns And Social Networks In Space‐Time. *Tijdschrift voor economische en sociale geografie 102*, 4 (2011), 468-485. doi:10.1111/j.1467-9663.2010.00649.x.

[100] LEE, K., HONG, S., KIM, S. J., RHEE, I. and CHONG, S. Slaw: A new mobility model for human walks. In *INFOCOM 2009* (Rio de Janeiro, 2009), IEEE, pp. 855-863. doi: 10.1109/INFCOM.2009.5061995.

[101] LEITNER, H., SHEPPARD, E. and SZIARTO, K. The spatialities of contentious politics. *Transactions of the Institute of British Geographers 33*, 2 (2008), 157-172. doi:10.1111/j.1475-5661.2008.00293.x.

[102] LIMAYEM, M. and DESANCTIS, G. Providing decisional guidance for multicriteria decision making in groups. *Information Systems Research 11*, 4 (2000), 386-401. doi:10.1287/isre.11.4.386.11874.

[103] LIMTANAKOOL, N., SCHWANEN, T. and DIJST, M. Developments in the dutch urban system on the basis of flows. *Regional Studies 43*, 2 (2009), 179-196. doi:10.1080/00343400701808832.

[104] LOMI, A. and PALLOTTI, F. Relational collaboration among spatial multipoint competitors. *Social networks 34*, 1 (2012), 101-111. doi:10.1016/j.socnet.2010.10.005.

[105] LU, X., BENGTSSON, L. and HOLME, P. Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences 109*, 29 (2012), 11576-11581.

[106] LUO, W., MACEACHREN, A. M., YIN, P. and HARDISTY, F. Spatial-Social Network Visualization for Exploratory Data Analysis. In *SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN 2011) in conjunction with the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS) 2011* (Chicago, Illinois., 2011), ACM. doi: 10.1145/2063212.2063216.

[107] MACEACHREN, A. *How maps work: representation, visualization and design*. Guildford Press, New York, 1995.

[108] MACEACHREN, A. *How maps work: representation, visualization, and design*. Guilford Press, 2004.

[109] MACEACHREN, A. and GANTER, J. A pattern identification approach to cartographic visualization. *Cartographica: The International Journal for Geographic Information and Geovisualization 27*, 2 (1990), 64-81. doi:10.3138/M226-1337-2387-3007.

[110] MACEACHREN, A. M., ROBINSON, A. C., JAISWAL, A., PEZANOWSKI, S., SAVELYEV, A., BLANFORD, J. and MITRA, P. Geo-Twitter Analytics: Applications in Crisis Management. In *Proceedings, 25th International Cartographic Conference, Paris, France* (2011).

[111] MACIEJEWSKI, R., LIVENGOOD, P., RUDOLPH, S., COLLINS, T. F., EBERT, D. S., BRIGANTIC, R. T., CORLEY, C. D., MULLER, G. A. and SANDERS, S. W. A pandemic influenza modeling and visualization tool. *Journal of Visual Languages & Computing 22*, 4 (2011), 268-278. doi:10.1016/j.jvlc.2011.04.002.

[112] MALCZEWSKI, J. GIS - based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science 20*, 7 (2006), 703-726.

[113] MALCZEWSKI, J. *GIS and multicriteria decision analysis*. JOHN WILEY Wiley & SONS, INC., 1999.

[114] MALCZEWSKI, J. Multicriteria Decision Analysis for Collaborative GIS. In *Collaborative geographic information systems*, S. Balram and S. Dragićević, Eds. IDEA GROUP PUBLISHING, 2006. doi: 10.4018/978-1-59140-845-1.ch010.

[115] MANGOLD, W. G. and FAULDS, D. J. Social media: The new hybrid element of the promotion mix. *Business horizons 52*, 4 (2009), 357-365. doi:10.1016/j.bushor.2009.03.002.

[116] MASSEY, D. B. *Space, place, and gender*. University of Minnesota Press, Minneapolis, 1994.

[117] MCANDREW, D. The structural analysis of criminal networks. In *The social psychology of crime: Groups teams and networks*, D. V. Canter and L. J. Alison, Eds. Aldershot. Dartmouth, 1999.

[118] MCPHERSON, M., SMITH-LOVIN, L. and COOK, J. M. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001), 415-444. doi:10.1146/annurev.soc.27.1.415.

[119] MENNIS, J. and MASON, M. J. Social and geographic contexts of adolescent substance use: The moderating effects of age and gender. *Social networks 34*, 1 (2012), 150-157. doi:10.1016/j.socnet.2010.10.003.

[120] MILGRAM, S. The small world problem. *Psychology today 2*, 1 (1967), 60-67.

[121] MILLER, H. J. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers 94*, 2 (2004), 284-289. doi:10.1111/j.1467-8306.2004.09402005.x.

[122] MORAN, P. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10 (1948), 243-251.

[123] ONNELA, J. P., ARBESMAN, S., GONZ LEZ, M. C., BARAB SI, A. L. and CHRISTAKIS, N. A. Geographic constraints on social network groups. *PLoS ONE 6*, 4 (2011), e16939. doi:10.1371/journal.pone.0016939.

[124] PHAN, D., XIAO, L., YEH, R. and HANRAHAN, P. Flow map layout. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS)* (2005). IEEE Symposium on Information Visualization.

[125] PRAGER, S. D. Complex Networks for Representation and Analysis of Dynamic Geographies. In *Understanding dynamics of geographic domains*, K. Hornsby and M. Yuan, Eds. CRC, 2008, ch. 3, pp. 31-48.

[126] PRECIADO, P., SNIJDERS, T. A. B., BURK, W. J., STATTIN, H. and KERR, M. Does proximity matter? Distance dependence of adolescent friendships. *Social networks 34*, 1 (2012), 18-31. doi:10.1016/j.socnet.2011.01.002.

[127] RADIL, S., FLINT, C. and TITA, G. Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles. *Annals of the Association of American Geographers 100*, 2 (2010), 307-326. doi:10.1080/00045600903550428.

[128] RAE, A. From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems 33*, 3 (2009), 161-178. doi:10.1016/j.compenvurbsys.2009.01.007.

[129] READ, J., EAMES, K. and EDMUNDS, W. Dynamic social networks and the implications for the spread of infectious disease. *Journal of the Royal Society Interface 5*, 26 (2008), 1001-1007. doi:10.1098/rsif.2008.0013.

[130] REDA, K., TANTIPATHANANANDH, C., BERGER-WOLF, T., LEIGH, J. and JOHNSON, A. SocioScape–a Tool for Interactive Exploration of Spatio-Temporal Group Dynamics in Social Networks. In *Proceedings of the IEEE Information Visualization Conference (INFOVIS'09), Atlantic City, New Jersey* (2009).

[131] RHEE, I., LEE, K., HONG, S., KIM, S. and CHONG, S. *Demystifying the levy-walk nature of human walks*. 2008.

[132] RIVEIRO, M. *Visual analytics for maritime anomaly detection*. Örebro universitet, 2011.

[133] ROCK, M. *Splintering Beijing: Socio-spatial Fragmentation, Commodification and Gentrification in the Hutong Neighborhoods of 'old' Beijing* Pennsylvania State University, University Park, 2012.

[134] ROSS, C. E. Neighborhood disadvantage and adult depression. *Journal of Health and Social Behavior 41*, 2 (2000), 177-187. doi:10.2307/2676304.

[135] SAILER, K. and MCCULLOH, I. Social networks and spatial configuration—How office layouts drive social interaction. *Social networks 34*, 1 (2012), 47-58. doi:10.1016/j.socnet.2011.05.005.

[136] SALATH, M. and JONES, J. Dynamics and control of diseases in networks with community structure. *PLoS Computational Biology 6*, 4 (2010), e1000736. doi:10.1371/journal.pcbi.1000736.

[137] SALATH, M., KAZANDJIEVA, M., LEE, J. W., LEVIS, P., FELDMAN, M. W. and JONES, J. H. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences 107*, 51 (2010), 22020-22025. doi:10.1073/pnas.1009094108.

[138] SALATH, M. and KHANDELWAL, S. Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. *PLoS Computational Biology 7*, 10 (2011), e1002199. doi:10.1371/journal.pcbi.1002199.

[139] SANDSTR M, A. and CARLSSON, L. The performance of policy networks: the relation between network structure and network performance. *Policy Studies Journal 36*, 4 (2008), 497-524. doi:10.1111/j.1541-0072.2008.00281.x.

[140] SCHAEFER, D. R. Youth co-offending networks: An investigation of social and spatial effects. *Social networks 34*, 1 (2012), 141-149. doi:10.1016/j.socnet.2011.02.001.

[141] SHEN, Z. and MA, K. L. Mobivis: A visualization system for exploring mobile data. In *PacificVIS '08* (Kyoto, 2008), IEEE, pp. 175-182. doi: 10.1109/PACIFICVIS.2008.4475474.

[142] SHEPPARD, E. The Spaces and Times of Globalization: Place, Scale, Networks, and Positionality*. *Economic geography 78*, 3 (2002), 307-330. doi:10.1111/j.1944-8287.2002.tb00189.x.

[143] SLINGSBY, A., BEECHAM, R. and WOOD, J. Visual analysis of social networks in space and time. In *Nokia Data Challenge Workshop, Pervasive 2012* (Newcastle, UK, 2012).

[144] SLOCUM, T. A. *Thematic Cartography and Geovisualization* Prentice Hall, Upper Saddle River, NJ, 1999.

[145] SONG, C., HAVLIN, S. and MAKSE, H. A. Self-similarity of complex networks. *Nature 433*, 7024 (2005), 392-395. doi:10.1038/nature03248.

[146] STAEHELI, L. A. Place. In *A companion to political geography*, J. A. Agnew, K. Mitchell and G. Ó. Tuathail, Eds. Wiley-Blackwell. Malden, MA, 2003, pp. 158–170.

[147] STANLEY, M. H. R., AMARAL, L. A. N., BULDYREV, S. V., HAVLIN, S., LESCHHORN, H., MAASS, P., SALINGER, M. A. and STANLEY, H. E. Scaling behaviour in the growth of companies. *Nature 379*, 6568 (1996), 804-806. doi:10.1038/379804a0.

[148] STROGATZ, S. H. *Sync: The emerging science of spontaneous order*. Theia, New York, 2003.

[149] SUI, D. Z. Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers 94*, 2 (2004), 269-277. doi:10.1111/j.1467-8306.2004.09402003.x.

[150] TAKHTEYEV, Y., GRUZD, A. and WELLMAN, B. Geography of Twitter networks. *Social networks 34*, 1 (2012), 73-81. doi:10.1016/j.socnet.2011.05.006.

[151] THIEMANN, C. *SPaTo Visual Explorer*. RoCS, Northwestern University, http://www.spato.net/, 2011.

[152] THIEMANN, C., THEIS, F., GRADY, D., BRUNE, R. and BROCKMANN, D. The structure of borders in a small world. *PLoS ONE 5*, 11 (2010), e15422. doi:10.1371/journal.pone.0015422.

[153] THOMAS, J. and COOK, K. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society, 2005.

[154] THOMAS, J. and COOK, K. A visual analytics agenda. *IEEE Computer Graphics and Applications 26*, 1 (2006), 10-13. doi:10.1109/MCG.2006.5.

[155] TITA, G., RILEY, K. J., RIDGEWAY, G., GRAMMICH, C. A. and ABRAHAMSE, A. *Reducing gun violence: Results from an intervention in East Los Angeles*. Rand Corporation, 2011.

[156] TITA, G. E. and RADIL, S. M. Spatializing the social networks of gangs to explore patterns of violence. *Journal of Quantitative Criminology 27*, 4 (2011), 521-545. doi:10.1007/s10940-011-9136-8.

[157] TOBLER, W. A computer movie simulating urban growth in the Detroit region. *Economic geography 46* (1970), 234-240. doi:10.2307/143141.

[158] TOBLER, W. *Flow Mapper Tutorial*. http://www.csiss.org/clearinghouse/FlowMapper/FlowTutorial.pdf, 2007.

[159] TOBLER, W. R. Experiments in migration mapping by computer. *Cartography and Geographic Information Science 14*, 2 (1987), 155-163. doi:10.1559/152304087783875273.
[160] TOOLE, J. L., CHA, M. and GONZ LEZ, M. C. Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS ONE 7*, 1 (2012), e29528. doi:10.1371/journal.pone.0029528.
[161] TSOU, M.-H. and YANG, J.-A. Spatial Analysis of Social Media Content (Tweets) during the 2012 US Republican Presidential Primaries. In *GIScience* (Columbus, Ohio, 2012).
[162] VALENTE, T. *Social Networks and Health: Models, Methods, and Applications*. Oxford Univ Pr, 2010.
[163] VERDERY, A. M., ENTWISLE, B., FAUST, K. and RINDFUSS, R. R. Social and spatial networks: Kinship distance and dwelling unit proximity in rural Thailand. *Social networks 34*, 1 (2012), 112-127. doi:10.1016/j.socnet.2011.04.003.
[164] VESPIGNANI, A. Predicting the behavior of techno-social systems. *Science 325*, 5939 (2009), 425-428. doi:10.1126/science.1171990.
[165] WANG, J., CHRISTAKOS, G., HAN, W. and MENG, B. Data-driven exploration of 'spatial pattern-time process-driving forces' associations of SARS epidemic in Beijing, China. *Journal of Public Health 30*, 3 (2008), 234-244. doi:10.1093/pubmed/fdn023.
[166] WANG, X., DOU, W., CHEN, S. E., RIBARSKY, W. and CHANG, R. An Interactive Visual Analytics System for Bridge Management. *Computer Graphics Forum 29*, 3 (2010), 1033-1042. doi:10.1111/j.1467-8659.2009.01708.x.
[167] WATTS, D. J. *Six degrees: The science of a connected age*. WW Norton & Company, 2004.
[168] WATTS, D. J. *Small worlds: the dynamics of networks between order and randomness*. Princeton university press, Princeton, NJ, 2003.
[169] WATTS, D. J. and STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature 393*, 6684 (1998), 440-442. doi:10.1038/30918.
[170] WHITE, J. J. D. and ROTH, R. E. TwitterHitter: Geovisual analytics for harvesting insight from volunteered geographic information. In *Proceedings of GIScience, 2010* (Zurich, Switzerland: September 15. , 2010).
[171] WINEMAN, J. D., KABO, F. W. and DAVIS, G. F. Spatial and social networks in organizational innovation. *Environment and Behavior 41*, 3 (2009), 427-442. doi:10.1177/0013916508314854.
[172] WINNER, M. G. *Thinking about you thinking about me*. Michelle Garcia Winner San Jose, CA, 2002.
[173] WOOD, J., DYKES, J. and SLINGSBY, A. Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal 47*, 2 (2010), 117-129. doi:10.1179/000870410X12658023467367.
[174] WOOLLEY-MEZA, O., THIEMANN, C., GRADY, D., LEE, J., SEEBENS, H., BLASIUS, B. and BROCKMANN, D. Complexity in human transportation networks: a comparative analysis of worldwide air transportation and global cargo-ship movements. *The European Physical Journal B-Condensed Matter and Complex Systems 84*, 4 (2011), 1-12. doi:10.1140/epjb/e2011-20208-9.
[175] XU, J. and CHEN, H. Criminal network analysis and visualization. *Communications of the ACM 48*, 6 (2005), 100-107. doi:10.1145/1064830.1064834.

[176] XU, Z. and SUI, D. Z. Small-world characteristics on transportation networks: a perspective from network autocorrelation. *Journal of Geographical Systems 9*, 2 (2007), 189-205. doi:10.1007/s10109-007-0045-1.
[177] YATES, D. and PAQUETTE, S. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management 31*, 1 (2011), 6-13. doi:10.1016/j.ijinfomgt.2010.10.001.
[178] YAU, N. *Facebook worldwide friendships mapped*.
http://flowingdata.com/2010/12/13/facebook-worldwide-friendships-mapped/,
2010.
[179] YI, J. S., KANG, Y., STASKO, J. T. and JACKO, J. A. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV 08)* (New York, NY, USA, 2008), ACM, pp. 1-6. doi: 10.1145/1377966.1377971.
[180] ZHOU, M. and PARK, C. The cohesion effect of structural equivalence on global bilateral trade, 1948–2000. *International Sociology 27*, 4 (2012), 502-523. doi:10.1177/0268580912443577.

Information Visualization

Geovisual analytics to support crisis management: Information foraging for geo-historical context

Brian Tomaszewski and Alan M MacEachren

Article

Geovisual analytics to support crisis management: Information foraging for geo-historical context

Brian Tomaszewski[1] and Alan M MacEachren[2]

Abstract

Information foraging and sense-making with heterogeneous information are context-dependent activities. Thus visual analytics tools to support these activities must incorporate context. But, context is a difficult concept to define, model, and represent. Creating and representing context in support of visually-enabled reasoning about complex problems with complex information is a complementary but different challenge than that addressed in context-aware computing. In the latter, the goal is automated system adaptation to meet user application needs such as location-based services where information about the location, the user, and user goals filters what gets presented on a small mobile device. In contrast, for visual analytics-enabled information foraging and sense-making, the user generally takes an active role in foraging for the contextual information needed to support sense-making in relation to some multifaceted problem. In this paper, we address the challenges of constructing and representing context within visual interfaces that support analytic reasoning in crisis management and humanitarian relief. The challenges stem from the diverse forms of information that can provide context and difficulty in defining and operationalizing context itself. Here, we focus on document foraging to support construction of geographic and historical context for facilitating monitoring and sense-making. Specifically, we present the concept of geo-historical context and outline an empirical assessment of both the concept and its implementation in the Context Discovery Application (CDA), a web-based tool that supports document foraging and sense-making. We also discuss the CDA's transition into applied use for the United Nations to demonstrate the generality of underlying CDA concepts.

Introduction

Context is an important concept for understanding the world. A common question is "what is the context?" For crisis management, context includes where the crisis is occurring, what events have transpired, and who is involved. Although a ubiquitous concept, context is a difficult term to define and operationalize. Typically, it is thought of as a type of setting that gives meaning and describes the situation and circumstances of an entity.[1] Geography and history offer unique perspectives on context through study of the interconnectedness of phenomena, events, and places across multiple spatial and temporal scales through which situations are understood. The research we report here has two goals. Our first goal was to introduce a conceptual

[1]Department of Information Sciences & Technologies, Rochester Institute of Technology, Rochester, NY, USA
[2]Department of Geography and GeoVISTA Center, The Pennsylvania State University, University Park, PA, USA

Corresponding author:
Brian Tomaszewski, Department of Information Sciences & Technologies, Rochester Institute of Technology, 31 Lomb Memorial Drive, Rochester, NY 14623, USA.
Email: bmtski@rit.edu

framework for the fusion of geographical and historical epistemological perspectives into a form of context we define as geo-historical context (GHC). From a visual analytics perspective, our emphasis is not on context as input to automated filtering (as in context-aware computing) but as a framework for sense-making, with context actively assembled through analyst–system interaction. Our second goal was to develop, implement, and demonstrate application of methods for addressing the challenge of foraging for relevant information and using it to construct and represent GHC. Major contributions of the work presented here, as per our two research goals, are that we have (a) developed a conceptual and computational model that is instantiated to represent GHC to support sense-making, (b) implemented a visual analytic system to assemble context information, and (c) evaluated the system with crisis management domain experts.

Often, knowledge and awareness of past associations, concepts, and places are critical to how situations are understood.[2] Thus, foraging for and integrating information that can contextualize situations from geographical and historical perspectives depends upon recognizing links across information fragments derived over extended time spans, geographic scales, and conceptual meaning. GHC creates a basis for creating these linkages. It supports understanding the interconnectedness of phenomena, events, and place across multiple spatial and temporal scales and it enables situations to be reasoned about, often through visual representations such as maps. The complex nature of GHC requires formalization to impose structure on the seemingly limitless parameters that must be related in order to make practical use of GHC for crisis management and similarly complex domains.

Here, we outline the conceptual framework for a GHC model, describe its implementation in the Context Discovery Application (CDA), and present the results of usability and utility evaluation. United Nations (UN) staff participated in the utility evaluation, providing input on the potential of the GHC framework and its instantiation in the CDA to support work in humanitarian crisis management. More specifically, the GHC model and its implementation in the CDA were assessed for their potential to help analysts: (a) forage for, structure, and operate on heterogeneous information artifacts (documents, maps) that are (b) assembled, processed, interrelated, and interpreted in order to produce and represent GHC, which, in turn, (c) allows situations to be understood and reasoned about within sense- and decision-making activities.

The paper is organized as follows. First, we present the theoretical foundations for our approach to context. Next, we outline a model of GHC developed and implemented on this foundation. This is followed by a brief overview of the CDA prototype, focusing on how its functionality reflects the GHC model. We then report on a usability study to refine the CDA. Finally, we report on a utility study, which offers insight into the GHC framework from the perspective of international crisis management practitioners, and of the CDA as an implementation of the model. We end the paper with a discussion of how the CDA has been transitioned into applied use by the United Nations. We also offer some final conclusions on the overall research; these comments emphasize use of geovisual analytics tools for assembling context to support information foraging and sense-making tasks.

Theoretical foundations for GHC

Conceptualizing context

Context as an object of research is conceptualized in diverse ways across different domains; Bradley and Dunlop[3] provide a useful review and synthesis of perspectives from linguistics, computer science, and psychology. Starting with a standard dictionary definition of context as "the interrelated conditions in which something exists or occurs,"[4] we adopt Brezillon's[5] view that these conditions act as a filter or framework to support a human agent's reasoning in order to provide the correct meaning and interpretation for available information that is potentially relevant to a sense- or decision-making task at hand.

Following from this, we make a distinction between (a) contextual information, or information that creates, represents and provides context and thus implicitly provides a framework for problem solving, most often in the form of constraints and (b) contextualized information, or information that is the focus of attention and that has been given meaning through the framework provided by contextual information. For example, contextual information providing a framework for crisis management in a situation like the Haiti earthquake of 2010 might include a topographic map depicting terrain and infrastructure before the earthquake combined with information from news reports about building and infrastructure damage. This information framework helps to contextualize official information in situation reports about rescue team activities and distribution of relief supplies. At a later point in relief efforts, when an aftershock hits and situations change, the recently contextualized information about distribution of relief supplies becomes part of the overall framework of contextual information through which new situation reports are contextualized.

Context: static or dynamic (or both)?

Given the shift and duality that can occur between information being contextual and contextualized, it is important to distinguish between (a) context as a static set of information categories that can be used like a cookie-cutter for constraining situational factors and (b) context as ephemeral and evolving, with parameters and properties that change dynamically.

Static, pre-described categories or classes of information that can be used to represent context are often used in artificial intelligence (AI)-based efforts to model context. For example, the Context-Web Ontology Language (C-OWL), an extension to descriptive capabilities of the Web Ontology Language (OWL), is designed to formally capture static concept contexts within a single ontology in order to support machine-based matching of concepts with other ontologies.[6] Gahegan and Pike,[7] in work focusing on capturing, modeling, and representing how concepts are socially constructed in scientific processes, emphasize the static component of context, as a fixed set of concept properties that essentially serve as basic concept metadata (i.e., who created a concept and when, and how the concept was created). Information categories that create a static context have also been used for schema matching and query matching in heterogeneous geospatial database integration[8] and in semantic similarity matching procedures.[9] The CYC project's knowledge base and common-sense reasoning engine incorporates the notion of static context using predescribed categories. CYC includes 12 "mostly-independent dimensions along which contexts vary (Absolute Time, Type of Time, Absolute Place, Type of Place, Culture, Sophistication/Security, Granularity, Epistemology, Argument-Preference, Topic, Justification, and Anthropacity)"[10] (p. 4).

A dynamic, shifting view of context is more prevalent in ubiquitous computing and distributed/situated cognition research where the focus is on human/machine/artifact interactions. In these domains, context and content are not separable entities; instead, context arises and is produced by activity.[11] More specifically, context can be interpreted as emerging from both activity and combinations of tools, settings, goals, and artifacts imbued with history.[2] A dynamic perspective of context creates challenges to modeling context or formally representing context information (whether visually, in a database, or an ontology) in that it is difficult, if not impossible, to imagine all possible contextual states, information needed to convey those states, and appropriate action within a given state.[12]

Whether a conceptualization of context is based on a static set of pre-determined descriptive categories used for processing and integrating information, or as a dynamic state formulated from a complex series of interactions of artifacts, social interactions, environmental conditions, or their combination, contexts provide a mechanism for reasoning (both human and computer) with situational factors. Human reasoning, in particular, is critical to crisis management sense- and decision-making activities. A particular emphasis in research reported here is on understanding how the notion of context functions as a human reasoning framework and mechanism for such activities.

Theories of contextual reasoning

Formalization of context into logical theories for use as reasoning mechanisms has been an active area of inquiry in AI and knowledge representation/reasoning since the 1980s. A motivation behind formalizing context has been as input to models of human, context-based reasoning for understanding situational factors within automated, machine-based reasoning systems.

A concept addressed in these efforts is generality, related to the range of contexts across which assertions are true.[13] Stated simply, situations are unique, but unique systems are impractical and any knowledge representation or reasoning system that applies to all situations will be too general to be useful. The generality problem makes it difficult (if not impossible or desirable) to conceive of a universal knowledge representation and reasoning language based on a homogeneous world.[14] An approach to deal with the generality problem is the use of contexts to localize knowledge and then to find "compatibility" between localized contexts. Compatibility here is treated as the relations that can be defined between contexts that enable reasoning across contexts. Localized and compatible perspectives on context as a reasoning mechanism are the core ideas underlying local model semantics and multi-context systems, which are discussed in the following section and used as a theoretical principle underlying the GHC model presented in this work.

Local model semantics and multi-context systems

A multi-context system begins with the premise that context, as a formal structure for reasoning, is based on "local" facts derived from a global knowledge base and used for reasoning about a given goal.[15] Giunchiglia and Bouquet[15] argue that a local context of reasoning is based on a cognitive context, or an individual's cognitive representation of the world, as opposed to a pragmatic context, or the external structure of the world. For example, in a conversation between two people, the pragmatic context might be composed of the speakers themselves, the time the

conversation is taking place, and the location of the conversation. Reasoning with information that depends on the pragmatic context is utilized only as much as that information is represented or relevant within a given state of the cognitive context.[15]

The utility of situating reasoning in a cognitive context is that it accounts for different and/or conflicting perspectives within an agent's cognitive view of the world.[15] Perspectives taken by different agents often differ in level of detail and interpretation will depend on what is implicitly assumed.[16] As a very simple example, the statement "The report is due on April 25th" could also be true if expressed as "The report is due today" with an implicit assumption that today is April 25th.

Despite potentially differing local reasoning contexts to describe a given domain, compatibility and overlap can and do exist. From the local model semantics perspective, compatibility between local reasoning contexts refers to mutually influential relationships between local reasoning contexts where similar perspectives can describe the same piece of the world, but with different details.[16] For example, two people looking at a globe may both see the Atlantic Ocean, but one person can see only North America (from his/her viewpoint), and the other can see only Europe. Thus, compatibility emerges from the fact that their reasoning is related (they are both seeing the ocean), but distinct as they are looking at different land masses.

Ghidini and Giunchiglia [17] (p. 229) encapsulate these ideas in two basic principles for local model semantics:

- Principle 1 (of Locality). Reasoning uses only part of what is potentially available (e.g., what is known, the available inference procedures). The part being used while reasoning is called the context (of reasoning);
- Principle 2 (of Compatibility). There is compatibility among the kinds of reasoning performed in different contexts.

These local model semantics concepts have clear application to geographic problems such as a disaster relief, where context might differ on the basis of place, time, or concept/theme. Two local contexts could share a place but differ by theme (e.g., using a hydrological versus a transportation perspective) or time frame (based on the long history of a resident versus a short duration from an external emergency response manager brought in to help). Alternatively, local contexts might represent adjacent places that share only a border and common regional perspective. "Local" also can be defined at different geographic scales, such as the perspective of the county emergency manager whose local context is a single county and that of the state emergency manager whose local context is the entire state. Furthermore, the compatibility relations among the different kinds of local contexts will be different. To summarize, local reasoning contexts that are derived from subsets of global knowledge and then paired into compatibility relationships with other local reasoning contexts are the essence of local model semantics and multi-context systems.

Context modeling challenges

Information foraging and sense-making must integrate two perspectives on context. The first is static, descriptive context categories used for integrating and relating heterogeneous information. The second is the distributed/situated cognition perspective, where context emerges from a complex mixture of tools, artifacts, beliefs, and intentions imbued with history and existing within a social context of use where context is dynamic (subsumes static categories of context). Both static and dynamic contexts provide a mechanism for reasoning with situational factors. A context of reasoning to support information foraging and sense-making is a localized, cognitive view of the world based on a subset of facts that retains unique characteristics of a perspective. Compatibility needs to exist and be formalized for cross perspectives of local reasoning.

From this theoretical framework, two core challenges for developing a conceptual model and visual representation of GHC information can be defined. The first is that the sheer limitlessness of geography, the past, and other situational factors makes complete computational representation of geo-historical (or any other form of context[1]) unachievable. To address this challenge, it is necessary, for particular situations, to establish information categories that define a static context to underpin a model of geo-historical context that can implicitly intervene in a task, provide constraints, and explicitly contextualize information when needed. The second challenge is that although geographical or historical information can be used to formulate a static context (as discussed in the first challenge), these and other information elements are not context unto themselves, but rather become parts of a broader dynamic geographic–social context that the information categories defining a static context must adjust to as situations evolve. Conceptual models and visual representation of and interfaces to GHC information, therefore, must find a balance in meeting these two challenges.

The following is a brief humanitarian crisis example used to illustrate these challenges. Information categories that define a static context such as areas that
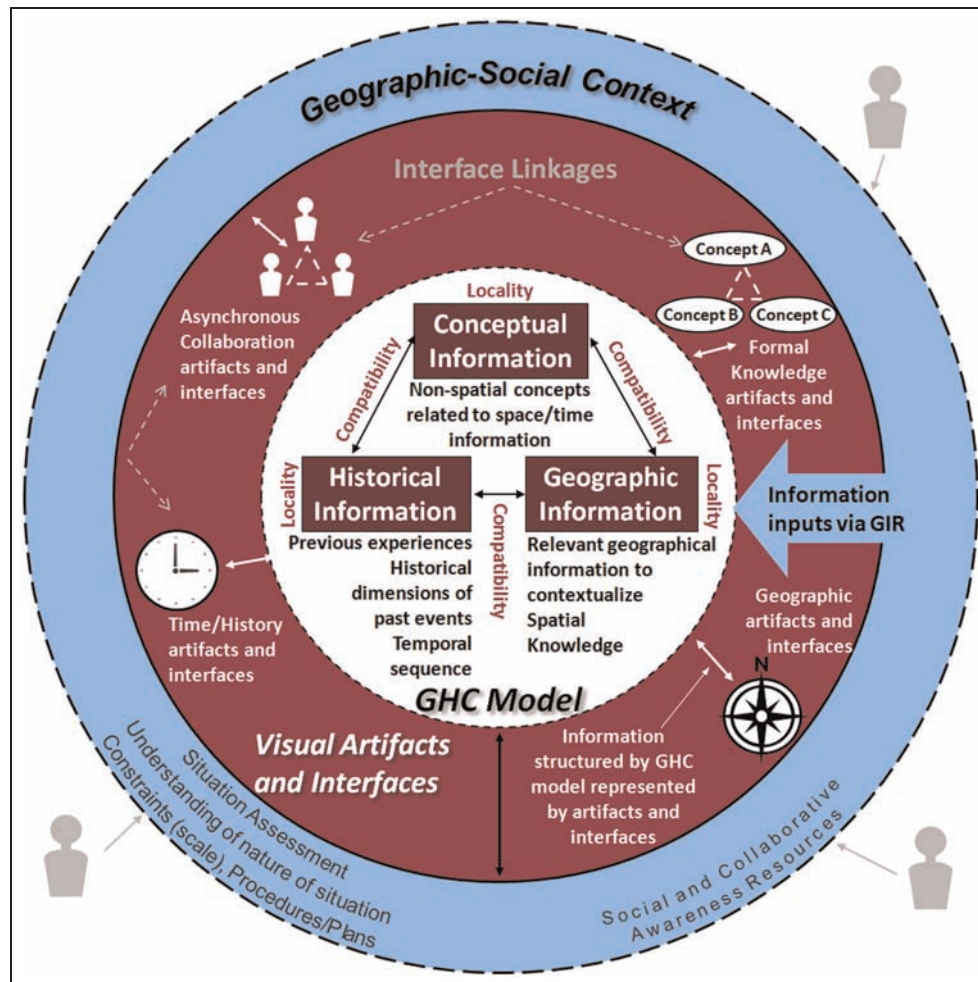
Figure 1. The overall process by which GHC is produced.

have been affected by famine over time and ethnic and tribal information about groups affected by the famine must be able to operate and adjust within dynamic social contexts such as the activities of an international aid organization. The organization will use static geo-historical contexts based on relevant information categories to derive geo-historical meaning to support their tasks and goals. An example goal might be to prioritize aid delivery based on locations of famine victims. As an analyst's work toward such a goal progresses, information categories that define a static context will need to adjust or expand as their relevancy to the situation evolves. For example, after addressing the initial needs of the famine victims by providing food relief within the static context of areas and people affected by the famine, the organization will then need to focus on myriad tasks that must be contextualized based on information categories such as social and political aspects of the affected area to ensure that relief efforts are lasting and sustainable.

The GHC model

Figure 1 is a high-level model representing the process of GHC production within a geographic–social context.

The GHC model serves two purposes. First, it provides a formal structure for the theoretical and conceptual components of GHC discussed in the "Theoretical foundations for GHC" section; these include (a) events, places, and concepts that represent a static form of context (as per the discussion in the "Context: static or dynamic (or both)?") as well as (b) relationships and constraints among these components such as scale, and spatial and temporal topological relationships—a set of relationships deemed robust and appropriate enough for the initial development of the GHC and its evaluation. The GHC model is specifically structured using three sub-models—geographical, historical, and conceptual—which represent windows into locality of context, along with

compatibility relations among the components, as per the ideas discussed in the "Context: static or dynamic (or both)?" section. In addition to being based on the ideas of local model semantics, this structure is also theoretically motivated by other data models that organize spatiotemporal phenomena into space/time/concept (or object) sub-models based on the intuition that these categories correspond with the way people think (cf. refs 18 and 19). Special characteristics of the GHC model when compared with other spatiotemporal models include the emphasis that the GHC model makes on (a) modeling context in particular, (b) modeling-derived knowledge and information rather than raw observational data, and (c) modeling reasoning contexts using local model semantics. For further discussion of the GHC model's geographical, historical, and concept sub-models, see ref 20.

Second, the model can be used as a conceptual template for structuring and representing specific information instances retrieved, compiled, developed, and ultimately used as part of foraging and sense-making processes. Context information can then be applied to fulfill a task or achieve a goal requiring geographical, historical, and thematic interpretations of situational factors. The GHC model is represented formally through an OWL-DL computational ontological structure (one of several sublanguages of the OWL), which is effective at representing, capturing, and describing aspects of real-world contextual information in computer readable formats.[21]

The GHC model has a simple structure. The three main sub-models (geographical, historical, and conceptual) begin as sub-classes of the supreme OWL Thing class. The specific structures of each respective sub-model were defined using existing, established ontological definitions wherever possible. In particular, the GHC ontology uses two existing ontologies as a starting point for dealing with space and time. The Geonames Ontology (GO) is used for representing discrete, coordinate-based geographic entities as per the conceptual structure of the geographic sub-model of the overall GHC model (http://www.geonames.org/ontology/). The OWL Time Ontology (TO)[22] is used for representing discrete instances of historical events in linear time as per the conceptual structure of the historical GHC sub-model.

Context Discovery Application (CDA)

The GHC conceptual framework has been instantiated in the CDA, a prototype geovisual analytics environment focused on document foraging and sense-making. In this section, we provide an overview of CDA functionality and outline a focused usability

assessment designed to "proof" a version of CDA prior to conducting the utility study presented in the next section.

CDA functionality

The following is a brief discussion of what a user might expect when working with a version of the CDA seeded with information relevant to humanitarian decision-making to ground specific functionality in a usage context (and present the version used in subsequently reported user studies). Thus, the version described has constraints over the "full", unconstrained CDA.[23] When started via loading the CDA web-client, the CDA will (a) visually render a small domain ontology in a graph display, (b) load a list of predetermined humanitarian project names into a dropdown list for selection by the user, (c) load a predetermined list of relevant specific date or time span event references into a timeline interface, (d) load a list of country names into a dropdown list for selection by the user to use with open text queries and, (e) display a base map of the world. Each of these items was compiled and incorporated into the CDA by the developer of the CDA using standard information technology tools such as ontology authoring, XML, and database systems for over a period of 1 month of part-time effort utilizing relevant information sources such as ReliefWeb (http://reliefweb.int/). To make the CDA a production tool, methods to enable end-user creation of this base information would be needed. But, the goal of the system reported was to act as a proof of concept for the GHC model introduced and to assess potential of the model to act as a framework for tools to enable real-world work, not to create a production tool. Thus, providing test users with initial base information loaded into the system seemed reasonable. All other functions are driven by actions taken by the user. A user will formulate a query by either entering an open keyword search much like a Google search or utilizing a query string automatically generated from the pre-determined list of humanitarian project titles. Using either approach, the user will then submit the query as the basis for processing by other CDA functionalities. In particular, the CDA supports ontology-enhanced queries to information sources such as Google News. Small domain ontologies (e.g., based on entities and relations extracted from an existing reference document) are used to support query expansion that makes it possible for an analyst to retrieve multiple, relevant documents without precise keyword matches in documents, an approach similar to that used in ref 24. The CDA uses the Named Entity Extraction methods of the open-source General Architecture for Text Engineering (GATE)

Figure 2. Representative example of the CDA.

environment to extract entities such as people, places, organizations, and things from the documents.[25] The CDA uses algorithms developed by the authors to disambiguate and geocode places extracted. Technical details of the CDA were described previously in ref. 23; here we sketch core features of the version adapted for usability and utility assessments to provide background for discussion of these assessments.

The CDA visually represents and allows users to explore implicit geographical information extracted from RSS feed-based sources. The information is represented in a set of linked views that included a two-dimensional (2D) map, Google Earth™ view, concept graph, timeline, named-entity extraction window, and web browser that enables the analyst to organize information, identify links, and use understanding derived to iteratively pose additional queries as understanding of situations and relationships is developed. The concept graph, timeline, and 2D map views are linked using geographic coordinates and temporal references. For example, a concept represented in the concept graph will have a latitude and longitude coordinate associated with the concept when the concept was created prior to analysis sessions. When the concept is clicked on, the 2D map will pan to the latitude and longitude coordinate so the user can see the geography associated with the coordinate. The 2D map and

Google Earth™ views are linked using a map synchronization algorithm discussed in ref. 23.

Figure 2 shows one view in which the user has selected a news story about humanitarian action in southern Sudan and how the user can browse through people, places, and organizations related to the story in the entity view (lower left). In this scenario, an analyst is cognitively gaining a contextual understanding of locations related to humanitarian events in the Sudan after doing a search. Locations involved in a complex crisis situation may have multiple conceptual and/or temporal dimensions that may be relevant for producing GHC. By "dimension", we mean some aspect of the GHC that can potentially be analyzed. For example, a conceptual dimension such as food insecurity and a temporal dimension such as the history of food security in a region. Conceptual and/or temporal dimensions can thus provide cognitive filtering capabilities that determine which locations are relevant based on relationships among concepts, events, and locations.

As a next step, the user has picked the timeline and concept map panel (Figure 3). A concept map depicting humanitarian relief activities is shown. Concepts that are extracted from a Sudan RSS feed and also found within the pre-determined small domain ontology discussed previously in this section are highlighted
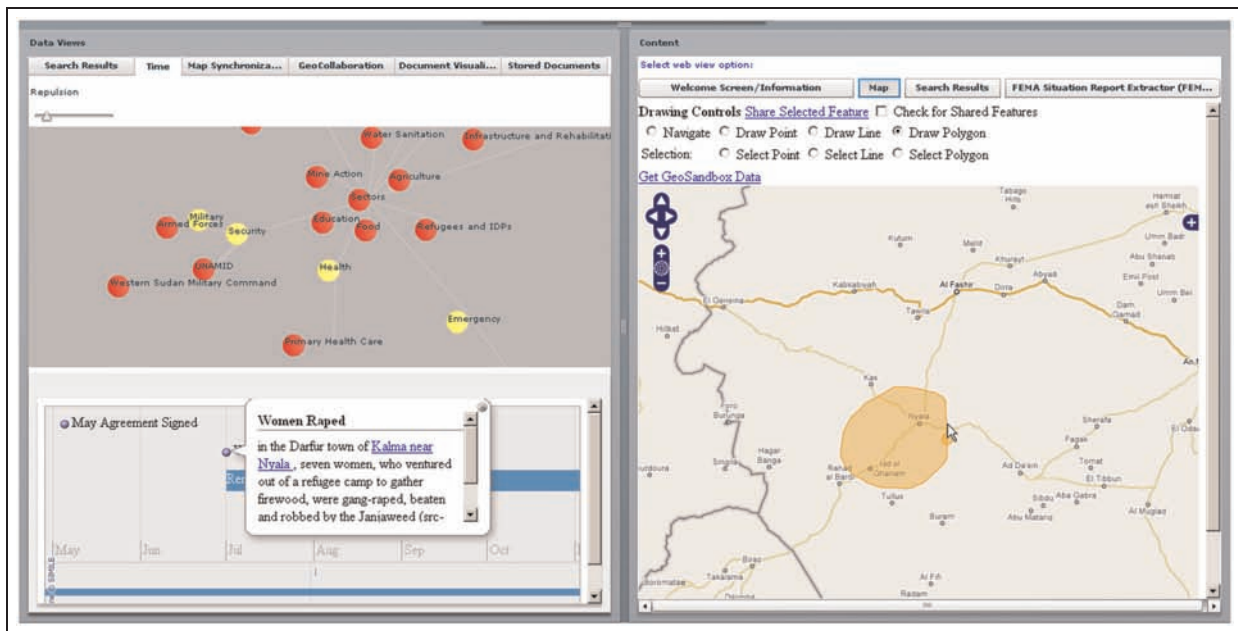
Figure 3. Concept map (top-left), timeline (bottom-left).

in yellow to provide a high-level overview of the conceptual content of the document. The concept map may prompt the user to recognize unanticipated concepts as contextually relevant.[26]

The scenario depicted illustrates: (a) selection of a concept, UNAMID, by clicking (Figure 3, upper left); (b) identification of a UNAMID security event from 2005 to which the concept is linked, resulting in autoscrolling of the timeline to highlight this event, or specific time reference, about atrocities (Figure 3, lower left); (c) mapping the event by clicking on a link to automatically re-center the map to the event (Figure 3, right), and (d) drawing an annotation on the map; see ref. 23 for a detailed account of this scenario and see ref. 27 for a video that illustrates how the dynamic changes help users to make sense of the interconnections among information fragments interactions. These interactions generate a set of explicit, linked artifacts such as map views, timeline references, highlighted concepts within the concept graph and external annotations that enable the analyst to construct GHC both in the analyst's mind as the linked artifacts enable analyst reasoning and through the linked artifacts themselves for interpreting events. The overall process is iterative, and the user can conduct further searches to find out more about topics learned during the initial search.

As an analyst works, he or she can simultaneously view geographic components of information extracted from text documents on the 2D map view built into the CDA and on an independent 3D Google Earth™

view depicting origins of news stories as points plus lines that connect to places mentioned (Figure 4).

The maps are dynamically linked and the 2D map includes a GeoSandbox feature enabling the analyst to save information to a central database; information that can be saved includes "snippet text" or text that surrounds a place reference, the URL of the source document, and latitude and longitude coordinates extracted from GE as deemed contextually relevant. Here, the user saves information from a news story mentioning Omdurman (see snippet in GE view, left) to the GeoSandbox within the CDA (markers indicate other saved information). The GeoSandbox is related to the idea by Wright et al.[28] of an analysts' Sandbox; their Sandbox offers a more general purpose concept-focused rather than geographic-focused organizational "space" for evidence fragments to support intelligence analysis. The information saved to the GeoSandbox repository can later be recalled using a database query with the results rendered in the CDA's map interface.

In addition to the above features, the CDA contains support for remote collaboration. One objective is to help analysts share analytic insights related to production of GHC and subsequent sense-making activities as per the ideas of situated cognition and dynamic context discussed in the "Context: static or dynamic (or both)?" section. This objective is addressed through the idea of a *geomessage* (Figure 5).

A geomessage uses a standard email message as a metaphor to support a spatiotemporally-enabled message for asynchronous geocollaboration. Conceptually,
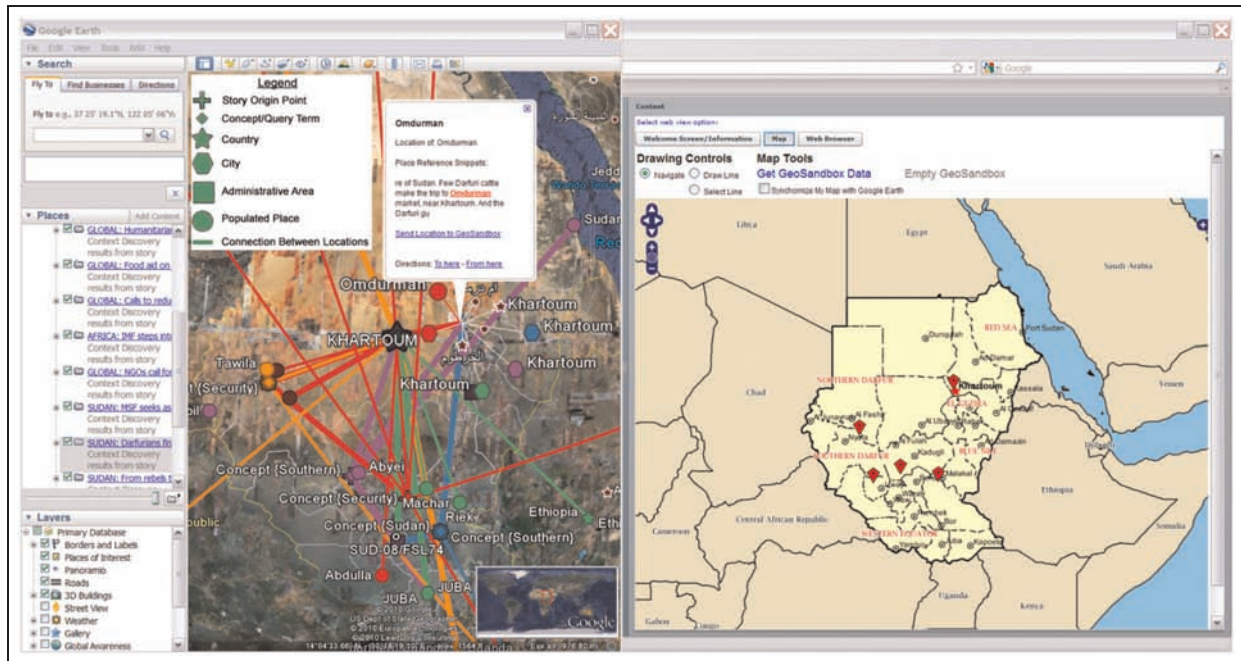
Figure 4. The GeoSandbox view linked to Google Earth™.

the body of a geomessage is a map extent linked to a text note like an email message.

Geomessages include an attachment mechanism designed explicitly to support geospatial artifacts that can be added to the user's map. In its current form, the geomessage supports attachments of map annotations and shared web map service (WMS) layers. WMS layers allow users to collaboratively create maps by sharing any individual and external geospatial resources from Open Geospatial Consortium (OGC)-compliant servers. Map layers from a given server can be loaded into the CDA by parsing the XML-based results from a GetCapabilities request made to the server (Figure 5).

Users can then select available layers and attach them to the geomessage. The benefits of this approach are that it (a) allows users to maintain a private view of their own data but share components as needed, (b) provides users with the flexibility to use multiple spatial data providers, (c) overcomes technical challenges to data sharing, such as when a novice user with geospatial technologies needs a layer for decision-making from a geospatial expert, and (d) supports user dialog through annotation tools that link formal geographic data to user knowledge, enabling holistic geographic awareness about a situation.[29]

CDA usability assessment and refinement

A targeted usability assessment of the CDA has been carried out. The primary goal was to develop a limited but robust version of the CDA through which to assess utility of the GHC conceptual framework (see "Evaluating the utility of the GHC framework"), and not to do a comprehensive user study. We summarize the methods and outcomes here, briefly.

Participants included two postdoctoral and three graduate students from the Penn State GeoVISTA Center; all were experienced with geographic information systems (GISs) and geovisual analytic technologies, enabling them to provide expert review from a tool design perspective. Five additional participants were from UN groups within the Office for the Coordination of Humanitarian Affairs (OCHA); all but one of the latter had some experience with GIS (and the one who did not was familiar with using Google Earth™); all were experts in humanitarian relief activities. The usability assessment had three components: a task analysis, a survey, and a focus group with the first two done individually by each participant and the last as a group.

Tasks. Each participant carried out tasks with representative features of the CDA such as extracting text from news articles and viewing the results in a map. Before the session, participants received instructions on how to use basic CDA features and on concepts such as the Consolidated Appeals Process (CAP). During the session, participants were asked to "talk aloud" while they worked and the work was observed and recorded. "Talk aloud" is slightly different than
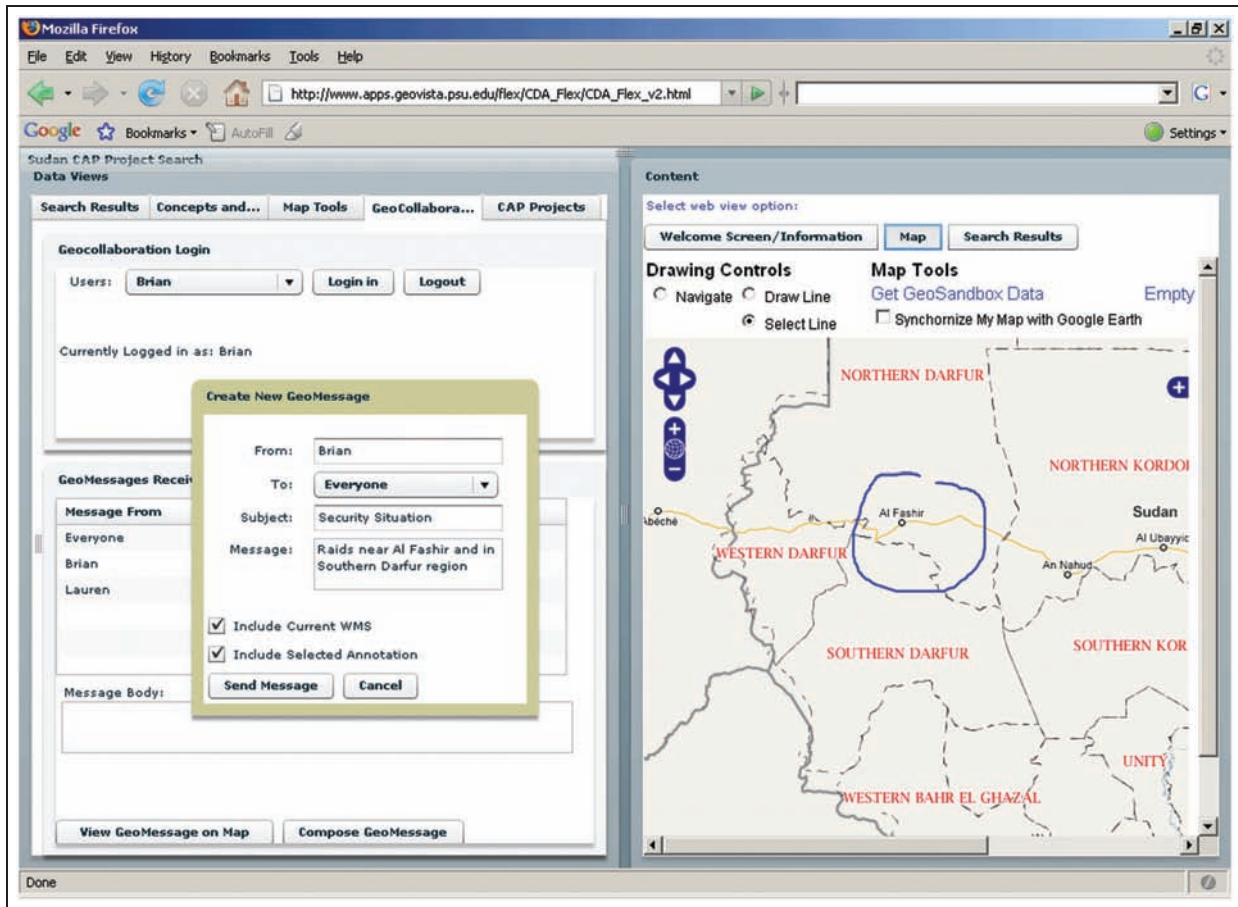
Figure 5. Example geomessage with map and annotation.

"think aloud" as in the talk aloud protocol, subjects describe but do not explain their actions.[30] The following six tasks were completed: (1) select (from a list) a CAP project to run a query against the Relief Web Sudan RSS feed, (2) review documents found in the search results interface (Figure 2, left), (3) use the space/time/concept interface (Figure 3) to review concepts, locations, and events that were potentially contextually relevant, (4) review search results in Google Earth, (5) use the GeoSandbox to save and select contextually relevant places, and (6) send a Geomessage. Tasks were conducted in 30-minute sessions.

Focus group. Given time constraints of UN personnel, only a focus group with Penn State participants was conducted. The goal was to leverage expertise of these individuals in design of geographic information technologies and visual interfaces focused on geographic analysis.

Surveys. Ten confidential survey forms were returned. The survey included Likert scale rankings (assessing understandability of the interface and of the displays,

consistency of design, acceptability of response time, and overall usability) and three questions allowing written response (focused on best features, worst features, and suggestions for improvement). The Likert ratings were summarized using descriptive statistics (mean, mediam, mode). Table 1 presents survey form response data from both the Penn State and UN usability studies. Figure 6 presents the information graphically using box plots, which provide information on the distribution of responses.

As can be seen in Table 1 and Figure 6, the range of responses to each question varied greatly. Most response questions averaged slightly above four points on the seven-point scale, indicating general, overall positive reactions to the usability of the CDA. Participants were most positive about system response time; in addition, the assessment of the consistency of interface component appearance was largely positive. However, results were not all positive; for the first question the central tendency was below the "average rating" and for the fourth and fifth central tendency was an "average rating." Responses to the first

Table 1. Usability study confidential evaluation form response data. The scale of measurement is 1 to 7.

ID	Question	Mean	Median	Mode
Q1	The software interface was easy to understand	3.6	3.5	2
Q2	The software responded to actions in expected ways	4.2	4.5	6
Q3	The visual appearance of interface components is consistent	4.4	5	5
Q4	The interface controls work in a consistent way	4.1	4.5	5
Q5	Visual displays of information were easy to understand	4	4	4
Q6	System response times were acceptable	5	5.5	6
Q7	Please rate the overall usability of the software	4.2	4.5	3

Color key
Negative response ▬
Neutral response
Positive response ▬
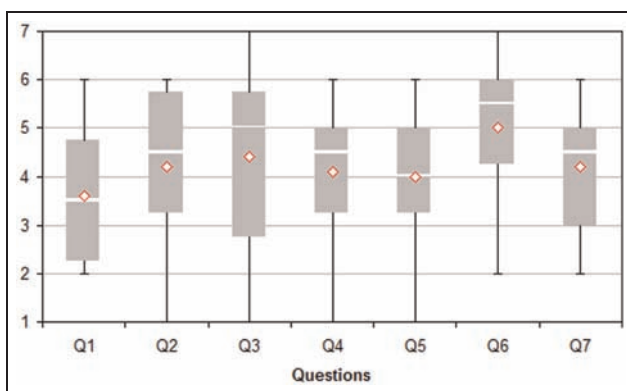
Figure 6. Usability study confidential evaluation form response question data box plots. Red diamond shapes in the center of each box are the average rating for the box.

question ("The software interface was easy to understand") indicate that a majority of users found the interface less easy to understand than average, and thus changes to the interface and/or improved training were deemed necessary. Responses to the fourth question add to this, with a rating for consistency of interface controls of average. Making the software interface and the visual displays easier to understand and controls more consistent were primary tasks undertaken between the usability and utility studies, as discussed in "Evaluating the utility of the GHC framework".

Data from talk aloud and observations during task completion, the focus group results, and the written responses to surveys were coded based on Nielsen's usability heuristics categories.[31] Table 2 presents illustrative quotes from the usability study pertaining to each usability heuristic. The full set of qualitative data from the task analysis, survey form written responses, and focus group discussions of the usability study, as classified into the 10 usability heuristics is provided in ref. 20.

As can be seen in Table 2, the usability study identified system issues related to all 10 usability heuristics.

Of the usability issues that were identified, the one category that had the most comments was error prevention. In particular, geocoding errors were deemed by study participants to be the main issue that needed to be remedied. The geocoding errors made some participants feel that the software could not be trusted in terms of the analysis it was providing. Several participants also commented that the software was difficult to understand in terms of control layout, flow between using tools for the various tasks, and presentation of the search results. Many comments were also made about the need to add a legend to the CDA's Google Earth™ output. Other minor specific design and functionality recommendations were made, such as adding better navigation for the concept view and redesign of the geocollaboration interface.

The combined analysis resulted in four primary changes to the CDA. The first change was to improve the geocoding algorithm's ability to identify geographic terms such as cities and countries and resolving identified terms to correct geographic coordinates. The second change was simplification of the overall user interface; from an interface with five main frames that each included many sub-controls to an interface with only two main frames and fewer controls. The third change was a modification of the search results interface, shifting from a tree view of retrieved documents with text fragments visible for the selected document to a view modeled on a Google Search results display at the top matched with a tagged document view below (as shown in Figure 2). The fourth change involved addition of a legend to the Google Earth display.

Evaluating the utility of the GHC framework

Utility of the GHC theoretical framework was evaluated through a study conducted at the New York offices of the UN. The emphasis was on evaluating the

Table 2. Usability issues from usability study.

Nielsen (2005) usability heuristic	Issue identified from usability	Illustrative quotes
Visibility of system status	Users asked for better system feedback	"That's why it is even more important to have those visual cues I think as to give people a sense of what is happening, so it has to be that when you do that search it comes back with something right away"
Match between system and the real world	Information not appearing in natural order	"Flow and visual organization of features is needed to better guide the users"
User control and freedom	User felt out of control when using the CDA	"I don't feel confident using the tools, that I know what's going to happen and what's linked to what … I don't feel in control much"
Consistency and standards	Search results deemed difficult to interpret Software was considered "inconsistent" (google earth CAP discussion)	"Too many clicks to review documents, needs to be more concise like Google" "Consistency is important, sometimes I click and a get a project note, sometimes I click on something that looks the same, like the outlines is thicker, and I get an explosion"
Error prevention	Users requested fixing geocoding errors. Search bug should be a top priority to fix because it will help to get a better idea of what will be shown	"A lot of room for error in analysis"
Recognition rather than recall	Users asked for a legend in the Google earth representation	"It would be ideal to add a legend graphic to Google Earth"
Flexibility and efficiency of use	Participants found concept map hard to navigate	"I would like to be able change the scale of this (the concept map), I have too many blinders, hard to navigate to get an overview, suppress the instances to get an overview"
Esthetic and minimalist design	Participants commented that the interface needed to be "more user friendly, more simple and attractive"	
Help users recognize, diagnose, and recover from errors	No explicit comments were made on this category, but when bugs/errors did occur, no feedback was provided to the user. Users only knew there was a problem from the developer (the author) being present and pointing out the problem. This issue was most prevalent in the geocollaboration tools	
Help and documentation	Participants found it hard to understand software—without explanation, it [the software] is difficult to understand	"I felt a bit lost … there were a whole bunch of tools that each one of which is individually interesting, but it was not clear how they connect and what I'm supposed to do"

GHC conceptual framework and the general strategy for instantiating it in the CDA, not to do a comprehensive evaluation of the CDA as production software. Thus, a limited version of the CDA was used that contained features relevant to humanitarian relief. The particular focus for the evaluation was on the problem of making humanitarian funding decisions. The scenario required contextualizing humanitarian crisis situations via analysis of implicit geographic information derived from open-source documents. More specifically, while using CDA, study participants conducted tasks focused on contextualizing and

reasoning with open-source information about humanitarian disaster relief projects in the Sudan.

Four UN staff members participated in the study, each from groups within the UN's OCHA, including the Field Information Services bureau and ReliefWeb. Having access to UN staff provided a key opportunity for examining a use of the GHC conceptual framework for application to real problems and not hypothetical scenarios. However, since UN OCHA personnel are constantly dealing with ever present world-wide disasters, the evaluation was necessarily much less controlled with fewer

participants than the typical laboratory study with student or similar participants.

The first part of the utility evaluation (task analysis) consisted of 1-hour sessions at the ReliefWeb office. The second and final part of the utility evaluation was a focus group evaluation. During the task analysis sessions, participants were observed conducting tasks and assistance was provided to the participants when requested.

To provide a comparable experience during task completion so that a follow-up focus group would be productive, the CDA was loaded with pre-compiled datasets for the task analysis sessions. Specifically, datasets from the ReliefWeb Sudan RSS feed and IRIN Sudan RSS feed were used. Each contained (a) documents that were processed by the CDA automated reasoning document classification and (b) a KML - Keyhole Markup Language representation of the RSS feed derived using the CDA.

Specific tasks were developed to assess the GHC model; these tasks incorporated geographic, thematic/conceptual, and historical components and the interrelations among them. Tasks were motivated by the following prototypical scenario that was also presented to participants at the beginning of each individual session.

Scenario: OCHA financial decision makers want an executive summary report on the evolving context of a select CAP project in the Sudan. They want to know how food security at local, regional, and international scales is playing out in the Sudan and how this may or may not relate to the efforts of the CAP project.

Two categories of cross-cutting tasks were developed to address this scenario. These tasks are targeted at establishing a preliminary context of a CAP project, thus to constrain/shape/contextualize a CAP project from geographic, historical, and thematic dimensions. One task sub-category includes tasks to assemble context information needed to answer a question (e.g., selecting the map in order to review basic CAP project information). The second sub-category includes tasks targeted at determining the contextual relevancy of a given piece of information (e.g., viewing a document's geographical footprint in Google Earth™).

Within these broad categories, distinct tasks were developed to utilize information from each of the GHC sub-models (geographic, historical, and conceptual/thematic) in different forms as per the ideas of locality (different view on the same world), and compatibility (interconnections existing with varying degrees of detail). The following are examples of GHC sub-model components related to sub-tasks that were developed to assess the GHC model in terms of locality:

- Geographic sub-model: View the document's geographical dimension in Google Earth (as seen in Figure 4).
- Concept sub-model: Review thematic information of potential relevance to a selected document that seems potentially useful (as partly seen in Figure 2 and also using the concept and timeline view of the CDA as in Figure 3).
- Historical sub-model: Review historical information of potential relevance to a selected document using the CDA concept and timeline view (see Figure 3).

The following are examples of tasks developed to assess the GHC model in terms of compatibility:

- Compile/synthesize your findings (a finding is something you think is potentially relevant that you would record/add to your report).
- Repeat (previous steps) as needed to support your analysis. For example, review several documents from different sources.

The intent of these tasks, in relation to evaluating compatibility in the GHC model, was to see how GHC sub-model information could be combined in order to represent the context of a humanitarian project. Each participant worked for approximately 1 hour and was asked to talk aloud during task completion. Participants used industry standard information technology tools such as Microsoft Word and Microsoft Powerpoint to compile and synthesize their findings. These tools were used as development of a specialized analysis synthesis tool was outside the scope of the CDA and the participants were comfortable with using these tools.

In addition to the specific detailed tasks outlined above, the participants were asked to create an executive summary report that outlined geographical, historical, and thematic dimensions of the CAP project's context. Figure 7 shows a map and the related excerpt from a report generated by one participant.

The report fragment and map shows multiple geographic dimensions of humanitarian projects in the Sudan in order to contextualize the project. On the map, colors represent individual news stories and lines represent locations computationally extracted from RSS feed to reveal potential geographic relationships between news stories. The ideas of locality and compatibility are reflected in this graphic in terms of how the study participant reports countries (locality) that
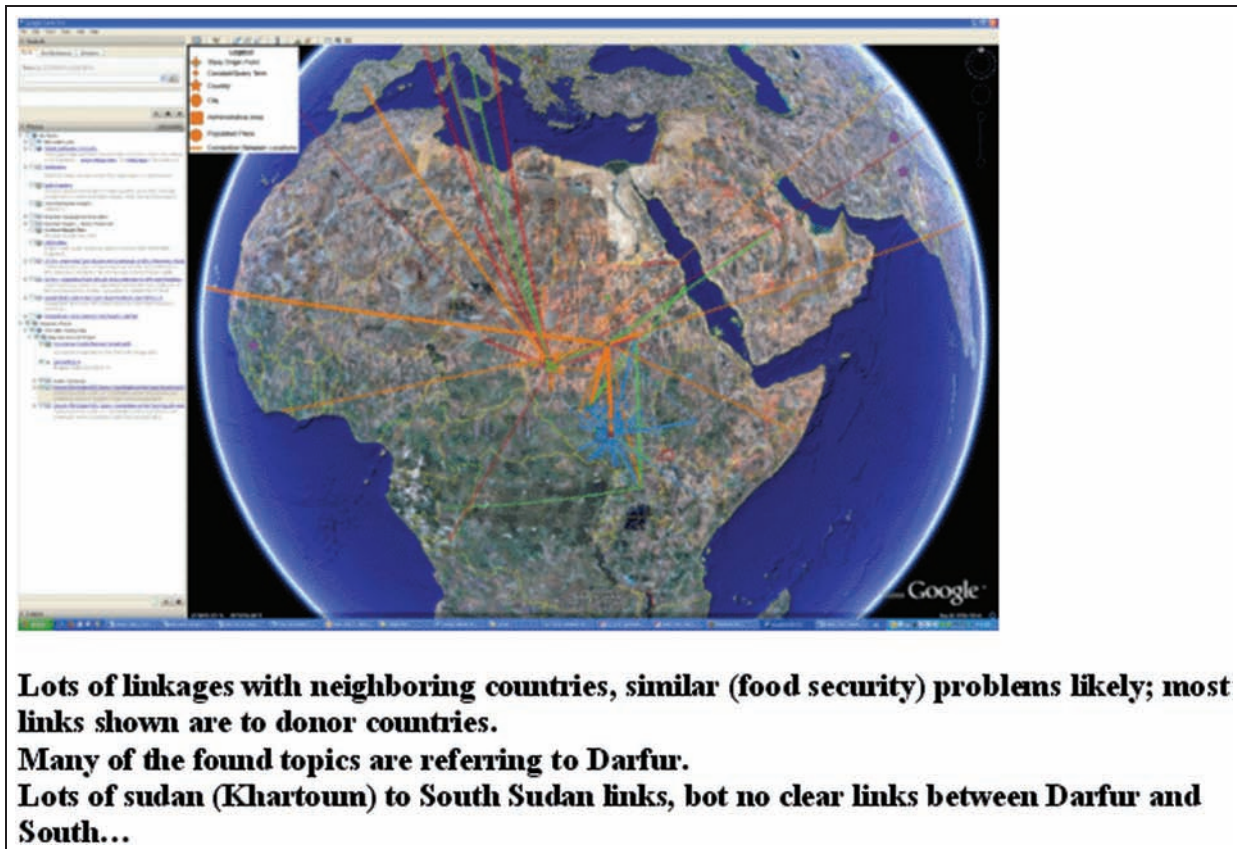
Figure 7. Example report section from one participant.

share similar but not identical food security problems (compatibility).

Following completion of the tasks, a focus group was used to prompt discussion targeted at key aspects of the GHC framework, its implementation in CDA, its applicability to the kind of task presented, and the general utility of the tools and the overall framework for supporting humanitarian relief activities.

Data from the utility assessment included notes taken while observing participants, transcription of the talk-aloud during portions of the task analysis sessions that focused on carrying out the task, and full transcriptions of the focus group session. Comments made in the transcriptions were analyzed using an approach outlined in Krueger[32] which has been utilized by cartographic/geography researchers such as Kessler.[33] In this approach, comments made during discussions are analyzed in terms of (a) frequency (or the number of times an individual or the group raises a specific comment), (b) extensiveness (which can be indicative of objection or support to a given topic), (c) intensity (potentially revealing feelings/emotions connected to a topic), and (d) what was not said (indicating that participants did not mention anything about a given issue).

Using an approach outlined by Kessler,[33] comments derived from the transcriptions were then matched to the specific research questions outlined above. Space permits only highlight from results.

In addition to assessing the viability of the tools developed (see ref. 34 for a discussion of evaluating the virtual globe aspects of the CDA), a sub-goal of the overall study was to examine whether the participants felt it was useful to consider context as a concept with geographic, historical, and conceptual components. The following are excerpts from the focus group transcriptions where the participants answered questions about utility of the GHC theoretical framework to support answering strategic-level questions.

As a prelude to examining GHC specifically, participants were first asked what the term "context" meant. P1 and P2 responded that context was a series of relationships between information components. This view matches closely with the theoretical perspective grounded in local model semantics outlined above. P1 even noted an idea discussed above that context acts like a framework: "it's also a very complex relationship let's say, because you're looking at some info and how it connects or relates to a ton of other

information may be around that subject, so that's, what I, I would consider a context of something." P4 made a comment that context was "the environment and situation", demonstrating how the terms context and situation can often be intermixed. As a follow-up to the basic question about a general definition of context, a question was then posed to the group about the notion of "analyzing a context" as per the language of the humanitarian case study the participants were conducting. P1 and P2 said that analyzing the context was in essence looking at the background to a specific situation. P4, however, restated that analyzing the context was in fact analyzing the situation.

Finally, probes directed discussion to whether it is useful to consider context to have geographical, historical, and conceptual components. All participants agreed that it is useful to consider context (and context for humanitarian projects in particular) to have geographical, historical, and thematic components. In agreeing with this, P4 related the notions of geographical, historical, and thematic components to humanitarian profile maps developed by ReliefWeb, as illustrated in the following quote: "yes because otherwise we couldn't, if we don't know the situation of one country … provide funding so that's why we also decided to put together a profile map. We can provide the situation of country, therefore they will learn of the situation and then they can provide some funding, so the situation, … the awareness of the situation is very important. It's key." Beyond the geographic dimensions mentioned explicitly, historical and thematic dimensions of context are implied in P4's quote implicitly. Humanitarian profile maps contain numerous references to thematic and historical information that is essential for understanding the situation. Also of note in the previous quote from P4 is the sentiment that the GHC helps provide a context in which decisions about funding can be made (although P4 didn't use the term context specifically). P1 and P2 also echoed this sentiment, and P3 agreed.

Participants were also asked to discuss how successful the CDA was at visually representing the geographic–social context of asynchronous group work that is characteristic of humanitarian relief efforts. During the questioning and subsequent discussion, P4 had stated that the collaborative aspects of the CDA were very useful – "the collaborative aspect and if it working since as it didn't work, that we would be able to collaborate, so you have the map, and you can draw top on it, and for example, and say 'oh look at that, I received the latest refugee information', and you can send and people can see it in a collaborative manner, and this I found it very well." As can be seen in P4's comment, a software bug with the geocollaboration system that was encountered did not greatly impact

P4's opinion about the potential of geocollaboration tools.

Further discussion about geocollaboration was prompted by asking "were you able to learn anything from your collaborators?" P4 indicated that since there was a bug, nothing could be learned from others, but that the geocollaboration tools would be useful as this is the era of Facebook. P2 reinforced this point by stating that CDA's geocollaboration tools are like a Myspace page.

Reports produced by participants were also reviewed. With the exception of Report 3, each of the four reports provided some indication of geographical, historical, and/or conceptual dimensions of the CAP projects contexts. The level of detail and means to represent and describe the GHC dimensions varied based on the form of information used by the report author. For example, participants 1 and 2 relied on Google Earth images with minimal accompanying text to describe contextual information. The author of Report 3 used numerous screen captures of web pages and the named entity view to describe geographical, historical, and thematic/conceptual dimensions, but did not include any map, concept, or timeline views. With the exception of Report 3, all participants used the geomessage tool to share findings. What was interesting to note was that two participants used the geomessage tool to develop messages with expressive annotations.

Three observations can be made from the focus group results. First, in spite of earlier usability assessments with UN personnel (who also participated in the utility study) and subsequent CDA revision, several usability issues were identified that were sufficiently serious to impede use for the tasks posed; in addition, one of the four participants needed to spend a substantial proportion of the hour-long session re-learning CDA functionality before starting on the tasks. This suggests that simpler interfaces and perhaps training in tool use will be necessary to put tools like these into practice. More positively, although participants could not agree on an exact definition of what context was, they all generally agreed that it was useful for context to be composed of geographic, historical, and thematic components for humanitarian information work. Finally, all participants, in different ways, saw the future value of specific components of the CDA (with an improved interface) for use in humanitarian information management.

Follow-up: the UN-OOSA RIVAF project

Since the CDA evaluation study, the CDA has been repurposed for use in disease dynamics analysis[35] and was also re-purposed by the United Nations Office for

Outer Space Affairs (UN-OOSA) for use in assessing the effects of the 2007–2009 global economic crisis (GEC) on the vulnerability of impoverished people around the world. The CDA is well suited to support such an analysis given the complex, heterogeneous, abstract nature of large volumes of data related to poverty indicators such as multi-scale economic markets, social network support, health and well-being and livelihoods. Specially, CDA visual analytic concepts and software tools were expanded upon and used, in part, on a research project titled "A Visual Analytics Approach to Understanding Poverty Assessment through Disaster Impacts in Africa", led by UN-OOSA and funded by the Rapid Impact and Vulnerability Analysis Fund of the UN Global Pulse initiative.[36,37] The objectives of the UN-OOSA RIVAF project are:

1. Understand GEC effects on relationships between livelihood, poverty, and vulnerability to natural disasters.
2. Understand how natural disaster impacts are potential indicators of GEC impacts on the poor and vulnerable.

To meet these objectives, a next generation of the CDA was developed via an online visual analytic environment called the "Visual Analytic Globe." This names change reflects the emphasis the new tool will make on incorporating space-based information as the per the mission of UN-OOSA Space-based Information for Disaster Management and Emergency Response Programme (UN-SPIDER)[38] into the overall visual analytic process via virtual globes.

Modifications to the CDA for development of the Visual Analytic Globe were motivated by (a) the results of the CDA usability and utility studies presented in this paper and (b) development of more robust geovisual analytic tools to support the tasks of the UN-OOSA RIVAF project. Specifically, the modifications include:

- Integrated, web-based virtual globe representations (vs. using an external standalone virtual globe) (Figure 8). Furthermore, each visual interface element (such as the space/time/concepts view discussed in the "CDA functionality" section but not shown in Figure 9) is now contained on a tab that can be easily repositioned, thus giving the analysts flexibility to configure the environment to support analytic tasks.
- Integrated quantitative data representations (Figure 9).

As can be seen in Figures 8 and 9, parts of the technology implementation have shifted from open-source to commercial tools to fit UN work practices. This provides some additional evidence that the core framework is fairly robust and extensible. While the information graphics tools included are relatively standard ones (in this case implemented using Esri[39] and Google[40] mapping tools), including them in the integrated environment meets clear needs of UN personnel to understand the effects of the global economic crisis on vulnerability using reliable, industry-standard tools vs. research prototypes.

Furthermore, core analytic capabilities of the CDA were expanded in the Visual Globe based on a user requirement that published documents be incorporated into the analytic processes of CDA use. This user requirement led to the a design choice to include new information retrieval tools that allow for the integration of archival documents such as government reports and studies by non-governmental organizations (NGOs). For example, numerous reports and studies on the effects of the global economic crisis such as socio-economic impacts were published in 2008–09. Often, these reports and studies contain valuable situational and contextual information of how people across varying geographic scales are coping with the effects of the GEC. Although valuable assets to support sense-making tasks, from an information retrieval and analytic perspective, the underlying knowledge in these assets are difficult to integrate with other data assets such as those that the CDA incorporates (as discussed in "CDA functionality") as these assets are often very large (25 + pages) documents that are time-consuming to analyze and make sense of. To overcome the challenges associated with incorporating archival documents into the CDA, a workflow was designed that allows the contents of archival documents to be structured to take advantage of the CDA's existing ontology-based querying tools and visual interfaces (see ref. 35 for discussion of these tools). In particular, the contents of each document were extracted from their source (often a PDF) and added to an XML template for storing the extracted contents in order to load them into a searchable document index.

Figure 10 graphically summarizes the workflow that was established to structure the contents of archival documents so the documents could take advantage of existing CDA tools and how documents processed in the workflow are incorporated into CDA visual interfaces.

Development of these new information retrieval tools to support annotation of archival documents to support subsequent access to the documents also provides evidence of how the core analytic framework of the CDA was capable of incorporating new types of data inputs to enable sense-making tasks. Ultimately,
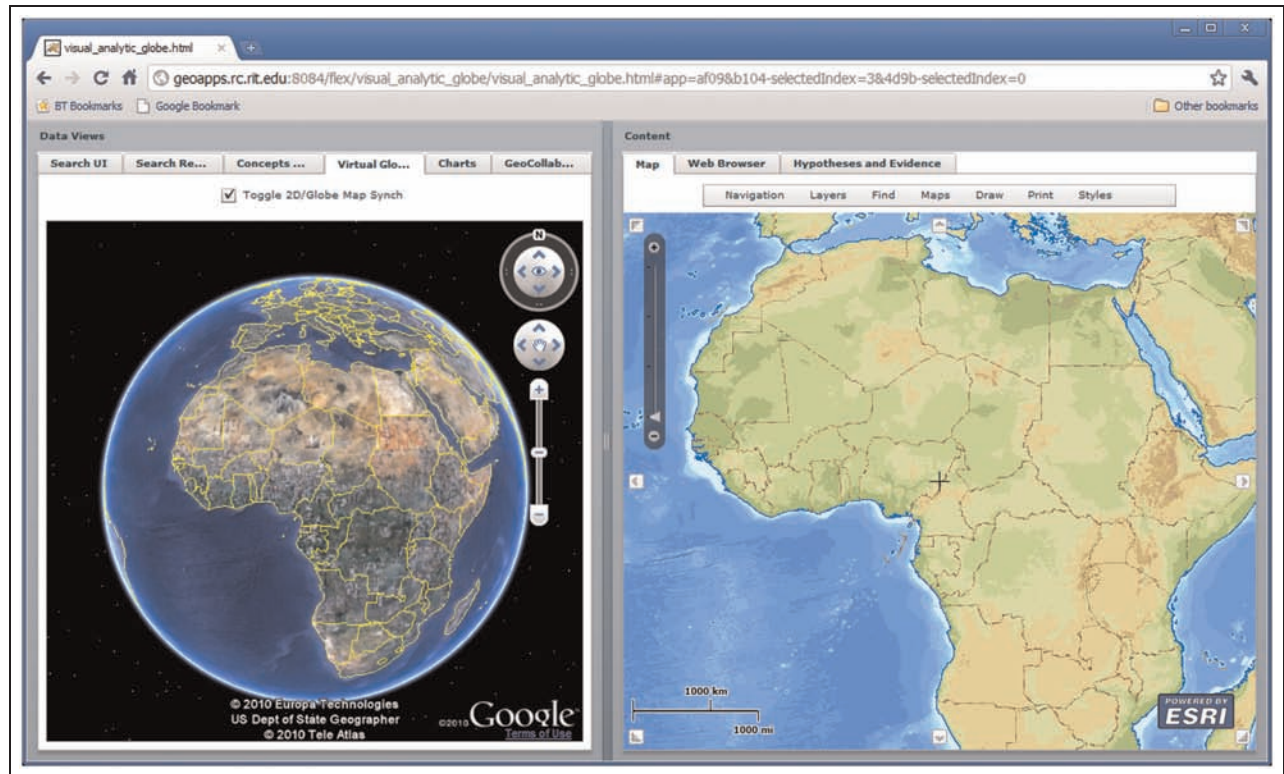
Figure 8. Integrated Virtual Globe and 2D map views. These views can be synchronized with one another. For example, panning and zooming the Virtual Globe shown on the left will make the 2D map show the right pan and zoom to the same display extent as the Virtual Globe. In this example, an analyst is beginning to conduct an analysis of Africa and is synchronizing the Virtual Globe and 2D map views so as to begin looking at specific areas within Africa, such as the cotton-growing area of Burkina Faso.

the CDA analytic framework can be expanded to include other types of data inputs, such as social media, in order to provide analysts with any relevant data artifacts that can be reasoned with and provide situational context using the CDA's visual interfaces.

Preliminary analysis

The following section briefly outlines select results from an analysis of Burkina Faso that was conducted for the RIVAF project based, in part, on use of the CDA analytic framework (as implemented in the Visual Globe tool). The example provided is meant to highlight the sense-making support and contextualization capabilities of the CDA and not to provide a full discussion of RIVAF project results, as such a discussion is beyond the scope of this paper. Subsections below summarize the analytic results then detail how the CDA analytic framework supported generation of these results.

Analytic results. Identifying the specific effects of the GEC on vulnerable communities in Burkina Faso was

a challenging, if not impossible, task. In the case of Burkina Faso, the challenge stemmed mainly from the complex web of vulnerabilities created via systemic poverty, economic disadvantages from being a land-locked country, limited government expenditures to fund activities such as public investment due to low state revenue and high reliance on external aid, and economic vulnerabilities created via trade fluctuation and a limited range of export commodities such as gold, livestock, and cotton.[41] According to the International Monetary Fund (IMF), although cotton contributes only 5–8% of gross domestic product (GDP), it accounts for 50–60% of Burkina's export revenues and foreign exchange.[42] Understanding the context surrounding cotton issues was thus of particular interest to the analysis.

For example, according to the Boards of the African Development Bank (ADB) and the African Development Fund (ADF), between 2005 and 2009, the average annual real GDP growth rate in Burkina was 5%.[43] The cotton sector drove the growth despite the challenging international context of the food, fuel, and economic crisis and a significant drop in world
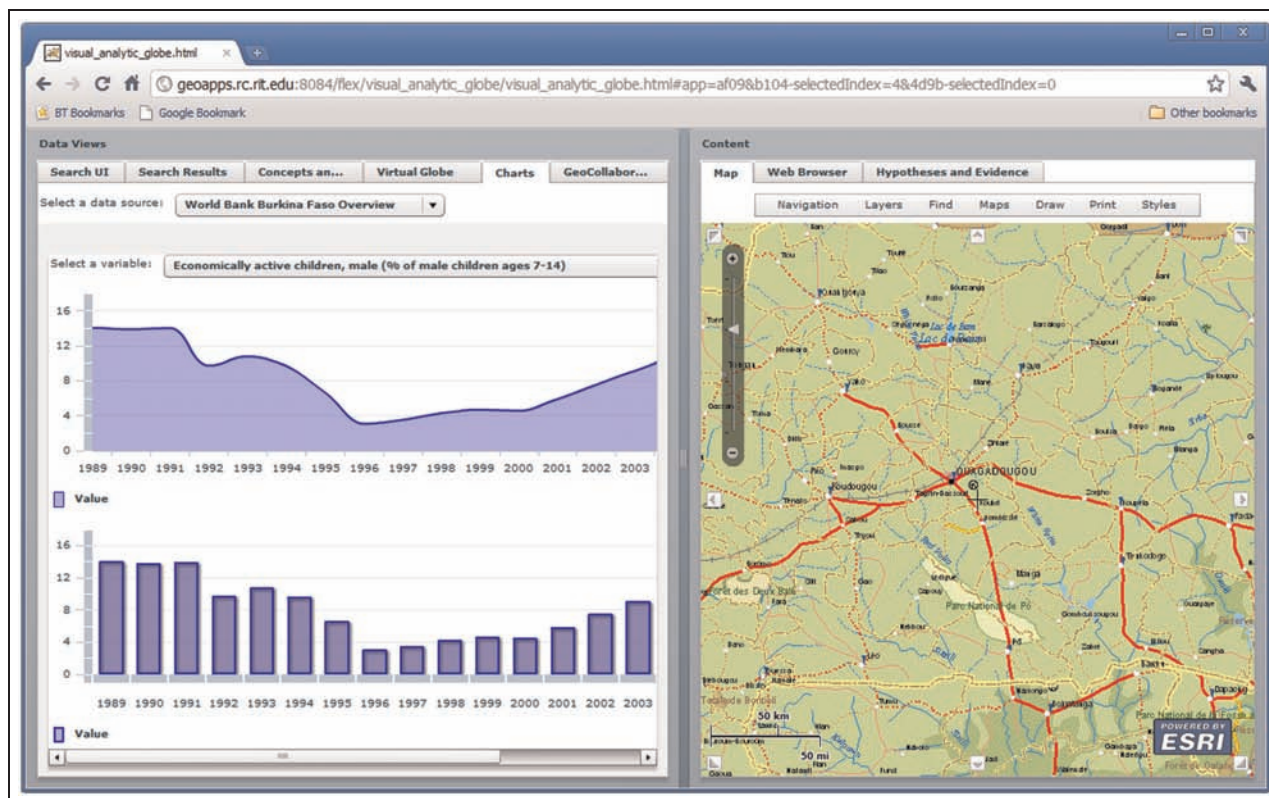
Figure 9. Integrated quantitative data representations. In this example, World Bank data related to economically active children in Burkina Faso are shown. Reviewing data such as these can help an analyst make sense of longer term social and economic trends in a country.
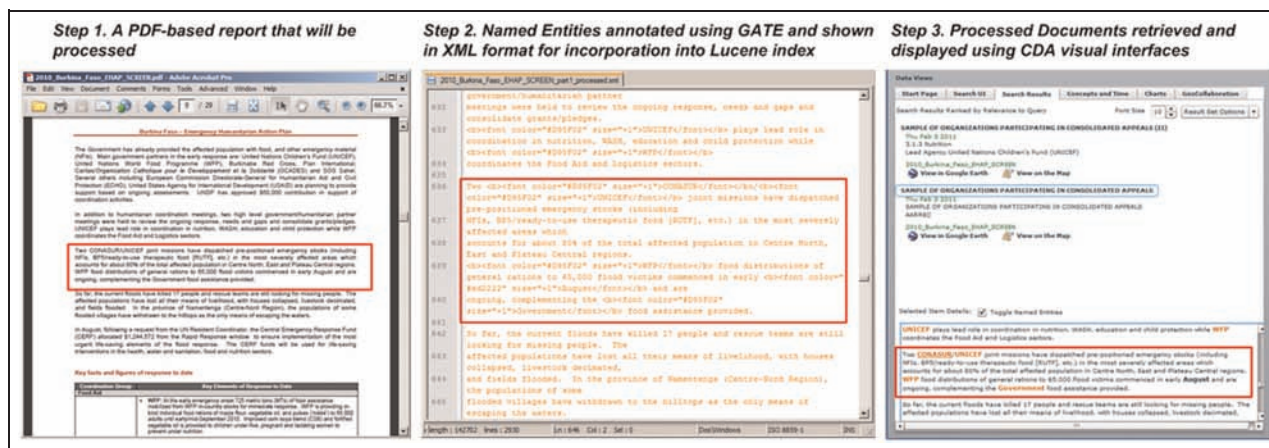
Figure 10. In this figure, a selection of text (outlined with a red box) from the 2010 Burkina Faso Emergency Humanitarian Action Plan published by UN OCHA is shown in various stages as it is processed for use as an archival document. This particular document was chosen as it provides valuable context information for Burkina Faso—a country that has been greatly impacted by the global economic crisis of 2008. The first step in the process is to run the PDF document (step 1) through the GATE program (not shown) to annotate named entities of interest in the text. Once the entities are annotated, the annotated text is stored in an XML-based template (step 2) that allows the document to be added to a customized Lucene-based search index, which can queried by an analyst, and return data artifacts such as named-entity annotated text that can be rendered in the CDA's existing visual interfaces (step 3) that are also used by the CDA's open-source information retrieval tools.

cotton prices between 2008 and 2009. In addition to pressures created by decreases in global cotton prices and demand, Burkina's cotton sector has also faced pressure because of subsidies given to US and other countries' cotton farmers that undercut cotton production.[44] In 2006, Burkina and other cotton-producing countries proposed that US cotton subsidies should be cut by 82.2% over a 2-year period as part of the World Trade Organization (WTO) Doha talks of reforming international trading systems by lowering barriers,[45] but the USA has yet to make a response.[46]

CDA support for understanding the Burkina Faso context. The CDA analytic framework helped develop this context in the following four ways. First, the news archive querying tools allowed analysts to quickly find relevant documents and activities between 2007 and 2009 (the approximate time period when the GEC was at its peak) (Figure 11).

Second, the named entity extraction and visual representation tools help to quickly discern organizations of interest such as the World Trade Organization, the Board of Governors of the African Development Bank (as seen in Figure 11), the World Food Programme (WFD), and locations of interest. For example, although Burkina Faso was the primary country of interest, understanding how Burkina Faso economically interacts with and is affected by other countries such as the USA is key to a contextual understanding of the cotton industry in Burkina Faso. Third, the integrated quantitative data representations (see Figure 9) helped to contextualize socio-economic trends in Burkina Faso. Finally, the integration of archival documents from UN-OCHA helped to make sense of how people were coping and adapting to economic fluctuations in the country such as diverting financial resources to purchase food at the expense of education and health, and switching to lower quality food.[47]

To summarize, the ability of the CDA analytic framework to contextualize complex situations, as demonstrated in the studies presented in this paper, helped in part to provide a key sense-making support role for the UN-OOSA RIVAF project. Specifically, the CDA assisted with understanding of the chain of effects that the GEC has had on poverty, livelihoods, and subsequent vulnerability to natural disasters and other stressors such as food insecurity.

Conclusions

Context is a difficult term to define and has many usages and meanings in different research and application domains. This research has defined a particular
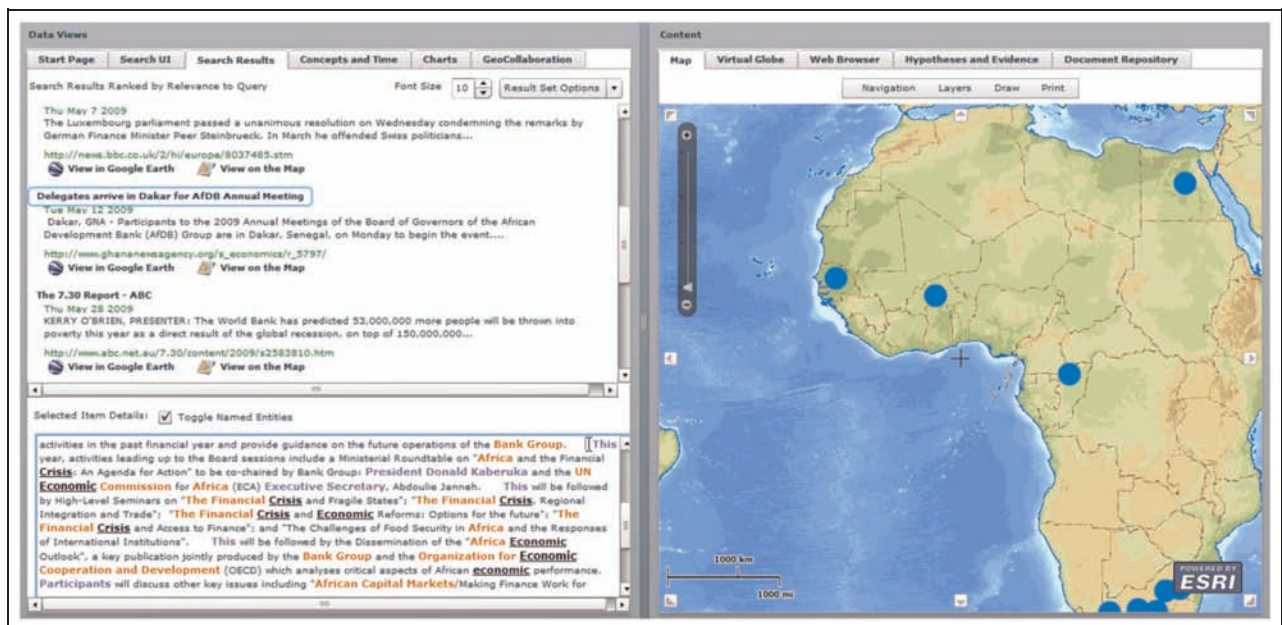
Figure 11. This figure shows how the CDA assisted with finding news stories from 2009 related to a regional organization within Africa (the Board of Governors of the African Development Bank) that was conducting round-table meetings and seminars and publishing reports related to African economic issues. In particular, the CDA's named entity recognition functions highlighted organizational names (as seen in orange on the bottom left of Figure 11). Items in bold, black, and underline are analyst query terms that were used to retrieve the news stories from Google News Archives.

type of context, called GHC, through the development of a GHC model that is based on relevant theoretical perspectives and computational representation strategies. The GHC model was the basis for developing a geovisual analytics environment designed to produce GHC from unstructured information sources within a challenging application domain, namely context analysis activities that underlie information foraging and sense-making in humanitarian crisis management.

While the evaluation of GHC as a conceptual framework presented here is a limited one, the limits are countered by the ecological validity of obtaining focused input from experts in international humanitarian relief. Evidence presented suggests that GHC is a viable framework for structuring situational reasoning tasks related to the assembling of GHC information. In particular, outcomes of the evaluation revealed that study participants thinking about the general notion of context matched closely with the local model semantics theoretical perspective as reflected in evaluation task outcomes and focus group discussions. Furthermore, participants were able to use the computational text processing and visual interface tools of the CDA to find relevant information and focus more of their attention on developing CAP reports. Most importantly, insights from implementation and the user studies have led to an adaptation of the methods and tools for real-world use by the UN.

Future research can incorporate the notion and evaluations of GHC promoted in this research to develop more robust geovisual analytics tools that are well suited to the information foraging and sense-making tasks of finding and interpreting information, identifying relationships, monitoring changes, and making decisions. Ideally, such efforts will lead to the development of visual analytics and other information systems that can effectively contextualize crisis situations from geographic, historical or any other dimension of interest.[48] Properly contextualizing crisis situations can lead to improved crisis mitigation, response, and coordination, and ultimately improve or save lives.

Funding

Declaration of conflicting interests

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the US Department of Homeland Security, United Nations Global Pulse or the United Nations Office of Outer Space Affairs.

References

1. Dey AK. Understanding and using context. *Pers Ubiquit Comput* 2001; 5(1): 4–7.
2. Cole M and Engeström Y. A cultural–historical approach to distributed cognition. In: Salomon G (ed.) *Distributed cognitions: psychological and educational considerations*. New York: Cambridge University Press, 1993, pp. 1–46.
3. Bradley N and Dunlop M. Toward a multidisciplinary model of context to support context-aware computing. *Hum-Comput Interact* 2005; 20(4): 403–446.
4. Context. Dictionary.com Unabridged (v. 1.1), http://dictionary.reference.com/browse/context (accessed 8 November 2007).
5. Brezillon P. Context in artificial intelligence: II. Key elements of contexts. *Comput Artif Intell* 1999; 18(5): 425–446.
6. Bouquet P, Giunchiglia F, van Harmelen F, et al. C-OWL: contextualizing ontologies. In: *Proceedings of the second international semantic web conference*, Sanibel Island, FL, USA: Springer, 2003 October 20-23, pp. 164–179.
7. Gahegan M and Pike W. A situated knowledge representation of geographical information. *Transactions in GIS* 2006; 10(5): 727–749.
8. Souza D, Salgado AC and Tedesco P. Towards a context ontology for geospatial data integration. In: *Workshop on reliability in decentralized distributed systems (RDDS)*, Montpellier, France, 2006 October 30, pp. 1576–1585. Berlin/Heidelberg: Springer.
9. Janowicz K. Kinds of contexts and their impact on semantic similarity measurement. In: *5th IEEE workshop on context modeling and reasoning (CoMoRea) at the 6th IEEE international conference on pervasive computing and communication (PerCom'08)*, Hong Kong, China, 17–21 March 2008, pp. 1–6.
10. Lenat D. The dimensions of context – space. CYCORP Report, CYCORP, Austin, TX, 1998 October 28.
11. Dourish P. What we talk about when we talk about context. *Pers Ubiquit Comput* 2004; 8: 19–30.
12. Greenberg S. Context as a dynamic construct. *Hum-Comput Interact* 2001; 16(2–4): 257–268.
13. McCarthy J. Generality in artificial intelligence. *Commun ACM* 1987; 30(12): 1030–1035.
14. Giunchiglia F. Contextual reasoning. *Epistemologia* 1993; XVI: 345–364.
15. Giunchiglia F and Bouquet P. Introduction to contextual reasoning. An artificial intelligence perspective. *Perspect Cogn Sci* 1997; 3: 138–159.
16. Bouquet P, Ghidini C, Giunchiglia F, et al. *Theories and uses of context in knowledge representation and reasoning*. Technical Report # 0110–28, October 2001. Istituto Trentino di Cultura, Trento, Italy.

17. Ghidini C and Giunchiglia F. Local models semantics, or contextual reasoning = locality + compatibility. *Artif Intell* 2001; 127(2): 221–259.

18. Peuquet D. *Representations of space and time*. Guilford Press, New York, USA, 2002.

19. Yuan M. Modelling semantical, spatial and temporal information in a GIS. In: Craglia M and Couclelis H (eds) *Geographic information research: bridging the Atlantic*. CRC Press, 1997, pp. 334–347, Bristol, PA USA.

20. Tomaszewski B. *A geovisual analytics approach for producing geo-historical context*. Doctoral Dissertation, Department of Geography, Penn State University, University Park, PA, 2009, 326 pp.

21. Strang T and Linnhoff-Popien C. A context modeling survey. In: *Workshop on advanced context modelling, reasoning and management associated with the sixth international conference on ubiquitous computing (UbiComp 2004)*, Nottingham, England, 2004 September 7-10, pp. 1–8.

22. Hobbs J and Pan F. Time ontology in OWL, http://www.w3.org/TR/owl-time/ (2006, accessed 28 August 2008).

23. Tomaszewski B. Producing geo-historical context from implicit sources: a geovisual analytics approach. *Cartogr J* 2008; 45(3): 165–181.

24. Smart P, Russell A and Shadbolt N. AKTiveSA: a technical demonstrator system for enhanced situation awareness. *Comput J* 2007; 50(6): 703–716.

25. Cunningham H. GATE, a general architecture for text engineering. *Comput Humanities* 2002; 36(2): 223–254.

26. Leake D, Maguitman A, Reichherzer T, et al. "Googling" from a concept map: towards automatic concept-map-based query formation. In: *Concept maps: theory, methodology, technology: Proceedings of the first international conference on concept mapping*, Universidad Pública de Navarra, Pamplona, Spain, 2004 September 5-9, pp. 1–8.

27. GeoVISTA Center. Context discovery application video, http://www.youtube.com/watch?v=TGCw9YlPqLc (2010, accessed 21 January 2012).

28. Wright W, Schroh D, Proulx P, et al. Advances in nSpace – the sandbox for analysis. In: *International conference on intelligence analysis*, McLean, VA, 2005 May 2–3, pp. 1–2.

29. Tomaszewski B and MacEachren AM. A distributed spatiotemporal cognition approach to visualization in support of coordinated group activity. In: *Proceedings of the 3rd international information systems for crisis response and management (ISCRAM) conference*, Information Systems for Crisis Response and Management (ISCRAM), Newark, NJ, 2006 May 16–17, pp. 347–351.

30. Ericsson K and Simon H. *Protocol analysis: verbal reports as data*. Rev. ed. Cambridge, MA: MIT Press, 1993.

31. Nielsen J. Ten usability heuristics, http://www.useit.com/papers/heuristic/heuristic_list.html (2005, accessed 25 August 2008).

32. Krueger R, Morgan D and King J. *Focus group kit*. SAGE, Thousand Oaks, CA, 1998.

33. Kessler F. Focus groups as a means of qualitatively assessing the U-Boat narrative. *Cartogr Int J Geogr Inf Geovis* 2000; 37(4): 33–60.

34. Tomaszewski B. Situation awareness and virtual globes: applications for disaster management. *Comput Geosci* 2011; 37: 86–92.

35. Tomaszewski B, Blanford J, Ross K, et al. Supporting geographically-aware web document foraging and sense-making. *Comput Environ Urban* 2011; 35(3): 192–201.

36. Tomaszewski B. Examining the effects of Global Economic Crisis (GEC) on poverty through natural disasters impacts – an overview of the UNOOSA RIVAF project. In: *Proceedings of the fourth United Nations international UN-SPIDER Bonn workshop on disaster management and space technology: "The 4C – Challenge: Communication – Coordination – Cooperation – Capacity Development"* (ed. A Froehlich), 2011. Bonn: Deutschen Zentrums für Luft- und Raumfahrt (DLR).

37. United Nations. Rapid Impact and Vulnerability Analysis Fund (RIVAF), http://www.unglobalpulse.org/projects/rapid-impact-and-vulnerability-analysis-fund-rivaf (2010, accessed 22 January 2011).

38. United Nations Office for Outerspace Affairs. About UN-SPIDER, http://www.oosa.unvienna.org/oosa/en/unspider/index.html (accessed 22 January 2011).

39. Environmental Systems Research Institute Inc. (ESRI). ArcGIS API for Flex. http://help.arcgis.com/en/webapi/flex/index.html (n.d., accessed 30 January 2011).

40. Google Inc. Google earth API, http://code.google.com/apis/earth/ (accessed 30 January 2011).

41. World Bank. 2009. Burkina Faso - Country assistance strategy for the period FY10-12. Washington D.C. - The Worldbank. http://documents.worldbank.org/curated/en/2009/08/10946534/burkina-faso-country-assistance-strategy-period-fy10-12

42. Amo Yartey C. Tackling Burkina Faso's cotton crisis. *IMF Survey Magazine: Countries & Regions*, 2008. http://www.imf.org/external/pubs/ft/survey/so/2008/CAR022508B.htm

43. African Development Bank. Burkina Faso: extension of the results-based country strategy paper from 2005–2009 to 2010–2011. 2010. http://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/Burkinafaso__CSP_eng2011.pdf

44. The Washington Times. (2008). Scenes from the world food crisis. Available: http://www.washingtontimes.com/news/2008/apr/28/scenes-from-the-world-food-crisis/ (accessed 20 January 2012).

45. The World Trade Organization (2012). The Doha Round. Available: http://www.wto.org/english/tratop_e/dda_e/dda_e.htm (accessed 20 January 2012).

46. Reuters Africa. (2009). US says no cotton trade deal yet but possible. Available: http://af.reuters.com/article/topNews/idAFJOE5060GU20090107 (accessed: 20 January 2012).

47. Office for the Coordination of Humanitarian Affairs (OCHA). 2010 West Africa consolidated appeal (2010), Geneva, Switzerland. see: http://ochaonline.un.org/cap2006/webpage.asp?Page=1880

48. Tomaszewski B, Robinson AC, Weaver C, et al. Geovisual analytics and crisis management. In: *Proceedings of the 4th international information systems for crisis response and management (ISCRAM) conference*, Information Systems for Crisis Response and Management (ISCRAM), Delft, the Netherlands, 2007, pp. 173–179.

Leveraging Geospatially-Oriented Social Media Communications in Disaster Response

Susan McClendon, Division of Information Technology, Wyoming Geographic Information Science Center, University of Wyoming, Laramie, WY, USA

Anthony C. Robinson, GeoVISTA Center, Department of Geography, The Pennsylvania State University, University Park, PA, USA

ABSTRACT

Geospatially-oriented social media communications have emerged as a common information resource to support crisis management. The research presented compares the capabilities of two popular systems used to collect and visualize such information - Project Epic's Tweak the Tweet (TtT) and Ushahidi. The research uses geospatially-oriented social media gathered by both projects during recent disasters to compare and contrast the frequency, content, and location components of contributed information to both systems. The authors compare how data was gathered and filtered, how spatial information was extracted and mapped, and the mechanisms by which the resulting synthesized information was shared with response and recovery organizations. In addition, the authors categorize the degree to which each platform in each disaster led to actions by first responders and emergency managers. Based on the results of the comparisons the authors identify key design considerations for future social media mapping tools to support crisis management.

Keywords: Communications, Crisis Management, Geographic Information, Geospatially-Oriented, Mashups, Social Media

INTRODUCTION

Crowd-sourced information has rapidly become an essential source of data in disaster response. Since the first well-documented efforts of citizen journalists on September 11th, 2001 and the use of internet blogs to collect information after the 2004 East Indian Ocean Tsunami, recent emergency response efforts have included mapping SMS messages after the Haiti earthquake in 2010. The Haiti earthquake represented a paradigm shift in the use of social media for disaster response, as multiple web-based platforms emerged to collect, refine, and disseminate crisis-related social media. The use of social media to gain real time information on the ground in a disaster has been driven by the rapid speed at which information can be

distributed, the cross-platform accessibility of information, and the ubiquity of social media worldwide (Vieweg, Hughes, Starbird, & Palen, 2010). The utility of this information has been enhanced by the creation of crisis maps based on location data extracted from social media communications (Liu & Palen, 2010; MacEachren et al., 2011).

In 2011 the American Red Cross conducted a survey that showed that 33% of citizens have used social media sites, including Facebook, Twitter, Flickr and SMS text messages/alerts to gain information about an emergency (American Red Cross, 2011). About half the respondents said they would contribute information during an emergency using social media channels. Statistics from the International Telecommunication Union reveal that in 2009 there were 4.6 billion mobile phone subscribers world-wide and 1.5 billion subscribers used mobile devices to access the internet (International Telecommunication Union, 2011). We can reasonably expect the use of social media in disaster response to increase in the future.

Research on the increasing use of social media in disaster response has emerged as a new focus in the field of crisis informatics (Anderson & Schram, 2011; Jennex, 2010). Extracting, categorizing, visualizing, and evaluating such information presents serious research challenges, including the problem of managing and extracting meaningful information from the large volume of contributions, applying the information to decision support workflows, and the development of formal information sharing protocols (Harvard Humanitarian Initiative, 2011). Recent work has focused on exploring the different roles social media sources may play among different types of crisis responders and decision makers (Lang, 2010; Reuter, Marx, & Pipek, 2012), and on the use of media gathered from recent disasters to characterize the utility of such information in crisis situations (Jennex, 2012). Mapping crowd sourced information in disaster response gained wide-scale media attention after the successful deployment of the Ushahidi Crisis Map during the 2010 Haiti earthquake (Starbird, 2011). There are a number of specific challenges involved in mapping social media communications, including the extraction of accurate location information, and the application of useful and usable cartographic representations to visually support situational awareness in crises. Research on the integration of Geographic Information Systems (GIS) and crowdsourced information from social media has focused more on the challenges of extracting action items and location information from social media feeds (MacEachren, et al., 2011) and less on the utility of the extracted information and the effectiveness of associated crisis maps to support emergency response.

Our research examines two applications that have leveraged geospatially-oriented social media during recent disasters; the Tweak the Tweet (TtT) project from Project Epic (Starbird & Palen, 2011) and Ushahidi (Okolloh, 2009), both of which have been used to create crisis maps of content collected from social media sources during recent disasters. For each application we examine collected data, information products, and evidence of subsequent response actions for two recent disasters; the 2011 Joplin tornado and 2010 Fourmile Canyon fire for Tweak the Tweet, and the 2010 Haiti earthquake and 2010 Gulf Oil Spill for Ushahidi. Other efforts have used crowdsourced information during recent disasters, including the open source Sahana platform (Currion, de Silva, & Van De Walle, 2007) and the collective effort of the Crisis Mappers Network (crisismappers.net). However, Sahana utilized data feeds directly from Ushahidi and TtT, and the Crisis Mappers Network is focused on connecting and empowering crisis collaborators, and does not offer its own specialized technology platform.

We begin with background information, including an overview of geospatially-oriented social media, followed by a brief history of TtT and Ushahidi. Next, we evaluate how each organization collected, processed and geo-located these social media communications and compare and contrast the cartographic representations and reporting capabilities of the resulting crisis maps. We then examine the effectiveness of each application by identifying examples of

actionable items used by military, government and non-government organizations that emerged from the use of these crisis maps. Finally, we conclude with key design considerations for future efforts to leverage geospatially-oriented social media in crisis informatics.

GEOSPATIALLY-ORIENTED SOCIAL MEDIA

Many social media sources, including Twitter, allow users to tag reports with coordinates to indicate their location on Earth (Li & Goodchild, 2010). This information is easy to process and represent on a map, but does not necessarily represent an "actionable" location. A more substantive challenge is associated with making use of textual descriptions of place (placenames and less-specific geographic features), like those often included in an SMS text message. Placenames in text can be geocoded to assign location coordinates, but placenames are usually associated with irregular areas (for example, the New York City metro area) at least as often as one might ever associate them with a specific coordinate location on earth (the centroid of the legal boundary of New York City). Determining less-specific geographic features is also of critical importance when using geospatially-oriented social media (Intagorn & Lerman, 2011).

Both platforms make use of social media collected from SMS and Twitter messages. SMS or "text messaging" is a short messaging service that allows for the storage and retrieval of short 160 character messages across global cellular telephone networks. Location information is not automatically attachable to SMS data, and must be inferred from the message itself. Twitter is a microblogging service that allows users to post messages up to 140 characters called Tweets via mobile phones or web accessible devices. Twitter users follow other users to see their tweets in a Twitter feed. The Twitter user community has developed linguistic markers to facilitate communication; including the @ symbol to address users (@username); the

RT abbreviation to represent a retweet (RT @ username); and # or hashtag to indicate keywords. Hashtags allow users to search Twitter feeds or to follow trends (Zappavigna, 2011). Location information can be added to a tweet using a phone's GPS capabilities, and can also be inferred from user profiles and mentions of placenames in messages themselves.

PROJECT EPIC AND TWEAK THE TWEET

Project Epic is research effort at the University of Colorado that aims to improve methods of public information gathering and dissemination during emergency situations. The project's mission is to couple computational methods with behavioral knowledge on how people develop information using social media in crisis situations (Palen et al., 2010). One Project Epic research project, Tweak the Tweet, was first presented in 2009 as a simple set of standardized communication practices coupled with a technology platform for making sense of crisis Tweets (Starbird, 2011; Starbird & Palen, 2011). TtT asks users to tweet using a crisis specific micro-syntax designed to enable real-time processing of Tweets. TtT features a web-based tool for collecting and visualizing contributed information using the Twitter API to continually-update a database. The categorized information is displayed on a simple map mashup using the Google Maps API.

The TtT micro-syntax is based on primary or main hashtags that can be used in any crisis situation and are designed to indicate the "who, what, and where" of the Twitter message content. For example, #name or #contact can be used to indicate "who"; #need, #shelter, #road, #open, #damaged can be used to indicate "what"; and #loc can used indicate "where". These hashtags are used in conjunction with an event tag to organize the crisis. Event tags can be spontaneously generated during an event, like #joplin or #tornado, or prescribed by the TtT micro-syntax like #4MileFire. Used together, the primary and event hashtags format meaningful machine readable tweets.

TtT was first deployed in the aftermath of the Haiti earthquake in January 2010 with the goal of having responders and agencies on the ground use the syntax. The first deployment did not have any associated mapping functionality, and the micro-syntax was not widely adopted by first responders or the public. Despite that, volunteers from Crisis Commons, TtT and other organizations tweeted or retweeted almost 3000 unique tweets formatted with the TtT syntax (Starbird & Palen, 2011). Since 2009, TtT has added a mapping component to their system design and the application has been deployed for over twenty major crises.

USHAHIDI

Ushahidi began as a non-profit African technology company that was developed to map incidents of violence in Kenya following elections in 2008. Ushahidi's mission is to develop platforms for sharing crisis information and personal narratives (Okolloh, 2009) and has since grown to develop tools to facilitate the democratization of information in broader contexts. The open source software tools developed by Ushahidi automate the collection of incident reports using cellular phones, email, and the web and facilitate the mapping of report locations in an interactive map mashup along with descriptive data to contextualize events.

Ushahidi offers three core products: the Ushahidi Platform, the SwiftRiver Platform, and Crowdmap. The Ushahidi Platform combines interactive mapping with the ability to capture real-time data streams from mobile messaging services and Twitter, and also supports email and web forms. It also provides spatial and temporal views of collected data. The SwiftRiver Platform allows for the real-time filtering and verification of data from these multiple data streams, including the ability to automatically categorize information based on semantic analysis, provide analytics and insight into user relationships and data trends, facilitate information validation and qualification, and it offers an interactive dashboard for monitoring and reporting purposes.

Crowdmap is a cloud-hosted solution designed to support rapid launches of both the Ushahidi and SwiftRiver platforms.

Since 2009, deployments of Ushahidi platforms have focused on election monitoring, reporting human rights violations, disease surveillance, wildlife tracking, and disaster response. Though there were several deployments of the Ushahidi platform prior to the 2010 Haiti earthquake, it was the Haiti crisis that brought Ushahidi international attention. Ushahidi adoption since the 2010 Haiti earthquake has seen significant growth. Since 2009, over 2 million people have visited Ushahidi sites, 600,000 of those during the Haitian and Chilean earthquakes in 2010. These deployments have gathered over 200,000 reports so far (Gosier, 2011).

DATA COLLECTION, MANAGEMENT, AND IDENTIFYING LOCATIONS

Both TtT and Ushahidi utilize technical and manual methods to collect, refine, and add meaning to data. The following sections describe how each platform is designed, how they manage data, and how they derive location information from collected social media reports.

Tweak the Tweet

After the initial launch of TtT during the 2010 Haiti earthquake, TtT refined its aims to promote *crowdfeeding* after analysis of results from the deployment in Haiti highlighted the difficulty of getting the crowd to adopt a micro-syntax for Twitter (Starbird & Palen, 2011). TtT promotes the monitoring of social media sites by volunteers (called voluntweeters) during a crisis and they disseminate information back into the crowd using the TtT micro-syntax. In addition, voluntweeters promote the use of the syntax through conversations with other responding organization volunteers and by posting instructions and links to TtT crisis maps on social media sites. For the events we

researched, TtT prescribed a micro-syntax with event tags including #boulderfire, #boulder, #4MileFire, #joplin, and #tornado. Used together, the primary and event hashtags can result in meaningful, machine-readable Tweets. Below are two examples from the Joplin Tornado and the Boulder Fourmile Canyon Fire:

- #offer diapers, paper towels, clothing, water at #loc 817 S. Main. #Joplin #contact 417-623-7927. #tornado;
- #boulderfire #loc VCA Pets Clinic 5290 Manhattan Circle #shelter cats & dogs #con 303-499-5335 #src http://bit.ly/bxzzav.

The TtT software platform utilizes the Twitter Streaming API to identify tweets based on the TtT micro-syntax and stores the tweets in a MySQL database, parsing the information into key-value pairs based on hashtags. The platform uses Google maps to map the tweet content after the location has been determined. At regular intervals a Ruby script parses the messages filtered by hashtags into a MySQL database and the script in turn updates a public Google spreadsheet. Because machine processing may miss meaningful data in the tweet, such as placenames and other locations, the TtT process uses a combination of automatic and manual processing by volunteers to populate data in the event spreadsheet. For this research we downloaded spreadsheets for the 2010 Boulder Fourmile Canyon fire and the 2011 Joplin tornado disasters. These TtT–hosted spreadsheets contain all the events collected for each disaster.

Ushahidi

In the hours following the 2010 Haiti earthquake, Ushahidi staff deployed the Ushahidi Haiti Crisis Map. Working in the United States, they gathered information from media reports and social media sources. Approximately 85% of Haitians had access to cellular telephones and the cellular telephone infrastructure, though damaged, was quickly repaired. Within days

a SMS short code number was set up in collaboration with phone companies and U.S. State Department resources and advertised through local radio stations. The messages received via SMS were sent to an automated system set up to facilitate message translation and mapping of the data by volunteers (Heinzelman & Waters, 2010).

Shortly before the 2010 Gulf Oil Spill, students at Tulane University began development of a crisis map to document oil refinery accidents using the Ushahidi platform. On the day the class presented the GIS map, the Deepwater Horizon oil rig exploded in the Gulf of Mexico (Dosemagen, 2010). The Louisiana Bucket Brigade, an environmental organization, worked with Tulane students to launch the Oil Spill Crisis Map to give Gulf residents a chance to contribute information about threats to their community and ecosystem from the oil spill. Data for the map was submitted via SMS, Email, Twitter and web forms. Citizens were encouraged to make a reports based on health issues, wildlife sightings, and other notable impacts they may witness in the region.

The Ushahidi API (Ushahidi, 2011) supports data exchange in XML (Extensible Markup Language) and JSON (JavaScript Object Notation). Ushahidi software supports PHP scripting and is designed to work with MySQL and is usually run on an Apache web server. Ushahidi software can be configured to work with common SMS gateway providers to process and deliver SMS messages, and it can be configured to use the Twitter Streaming API to process Tweets. Data can be exported from MySQL via a PHP script to a Google spreadsheet. The Ushahidi map template is designed with a link to download the raw data in a Google spreadsheet, but because the Ushahidi platform is open source and can be modified by the organizations that deploy the software, not all organizations include the ability to download the data. For the 2010 Haiti earthquake and the 2010 Gulf Oil Spill, we were able to download spreadsheets with data covering six months after the initial incidents.

Streaming Data and Scalability Challenges

Collecting data from social media communications like Twitter and SMS is difficult due to the large datasets that can be generated in a short amount of time. A key challenge has emerged in automating the extraction of useful and actionable data from such sources. In fact, applying structure to content using tweets with a micro-syntax to enhance computational automation was part of the original intent behind the TtT project. Challenges associated with filtering, managing, analyzing and translating large volumes of social media communications are being addressed through ongoing development of The SwiftRiver platform by Ushahidi.

During the first deployment of TtT for the 2010 Haiti earthquake, the syntax was not widely adopted by citizens and first responders, but the syntax was picked up by people who spontaneously volunteer during a crisis (Starbird & Palen, 2011). TtT efforts spurred a network of volunteers that helped give structure to the social media communications that were transpiring on Twitter during both the 2011 Joplin tornado and 2010 Boulder Fourmile Canyon fire crises. These volunteers adopted the TtT syntax and translated information from multiple sources using the syntax before tweeting it out to their followers. These followers were diverse, including media outlets, the American Red Cross, FEMA, and other relief organizations. This type of volunteerism was promoted to direct Twitter communications so that automatic filtering of Tweets would be more effective.

During the 2010 Haiti earthquake, Ushahidi enlisted volunteers to assist with handling the large volume of data. SMS messages began to flow at a rate of 1,000 to 2,000 a day and were passed directly from the cellular telephone provider to an automated system, designed by Ushahidi developers for coordinating volunteers. Volunteers manually translated the messages from Haitian Creole and then filtered and determined locations (Meier & Munro, 2010). The system supported message translation with a lead time of less than ten minutes.

EXTRACTING LOCATION INFORMATION

One of the most challenging aspects of using social media data during a disaster is extracting unambiguous and accurate location information. Locations are essential for determining if a message is actionable (Munro, 2011). Location can be determined in several ways, including processing location references like a place name or street address in message content; explicit coordinates derived from geo-location services from cellular phones; and extraction of location information in a user's social media account profile (Bellucci, Malizia, Diaz, & Aedo, 2010; Field & O'Brien, 2010). Table 1 illustrates examples of profile-derived location information shown in Tweets 5-8.

Location information included within a Twitter or SMS message as a text reference (e.g. a user mentions a specific place by name) must be extracted and geocoded to obtain coordinate information. Latitude and Longitude coordinates can also be included with messages as geospatial metadata. The process of manually or computationally assigning such metadata is called geo-tagging. Twitter included the ability to geo-tag tweets in 2009 (Bellucci, et al., 2010). Because of privacy concerns, social media applications and cellular phones usually require users to opt-in to enable geo-tagging. The SMS protocol does not incorporate geospatial metadata and typical messages sent from cellular phones via SMS will not contain location information (Munro, 2011). However, GeoSMS (geosms.wordpress.com), a location-enabled SMS standard, can embed geospatial metadata into a URI (Uniform Resource Identifier). MacEachren et al. (2011) notes that the proportion of users who enable geo-tagging is still small. However, geo-tagging alone is no guarantee the message content is meaningful. Of three geo-tagged examples in Table 1, two geo-tags are close to Joplin in Springfield, Missouri (1) and the nearby city of Miami, Oklahoma (2), while the location in Tweet 3 is in the Netherlands.

Table 1. Tweet examples from the 2011 Joplin Tornado

Tweet	Time	User	Tweet	Location
1	11.21 pm 23 May 2011	@sarahgracesitz	RT @shawncmatthews: Here is a great resource for donation centers #joplin #tornado #relief http://ow.ly/515zC	iPhone: 37.112511,-93.303925
2	11.18 pm 23 May 2011	@CajunTechie	If you want to donate clothing of all sizes to displaced residents in #Joplin #Missouri you can drop it by 702 Moffet #tornado	ÃœT: 36.874136,-94.873582
3	10.07 pm 23 May 2011	@pbdoetmee	Ronduit dramatisch fotowerk na de tornado hit-down in Joplin, Missouri, USA... http://bit.ly/jiw6Kg#indrukwekkend #joplin #tornado	51.953923,6.008155
4	11.35 pm 23 May 2011	@PeterKinder	Just confirmed I will be guest on @IngrahamAngle radio show Tuesday morn 5/24/11 9:30 CDT talking #Joplin #MO #tornado relief, update #pdk	iPhone: 0.000000,0.000000
5	10.48 pm 23 May 2011	@maryfranholm	RT @OzarksRedCross: #CBCO FB site says: CODE RED 4 blood donations 4 the #Joplin #tornado has been lifted. Thanks 2 so many of you donat ...	Lost in a good book
6	10.39 pm 23 May 2011	@Jeannie_Hartley	RT @OzarksRedCross: #RedCross update here: http://bit.ly/jZ3Ivp#Joplin #tornado	Universe
7	10.36 pm 23 May 2011	@wheelertweets	RT @Jeannie_Hartley: #Tornado #Joplin #mo @info4disasters @Redcross @kcredcross @1stAid4 @wheelertweets @jnicky63 @viequesbound @Lady1st ...	Tunis, Tunisia
8	11.02 pm 23 May 2011	@JoplinMoTornado	GOODNEWS: 7 people were rescued from the debris today! #joplin #tornado -G	Joplin, Mo
9	11.02 pm 23 May 2011	@DAOWENS44	RT @OzarksRedCross: #CBCO FB site says: CODE RED 4 blood donations 4 the #Joplin #tornado has been lifted. Thanks 2 so many of you donat ...	

The Ushahidi platform does not contain a mechanism to automatically geocode implicit location information, but the SwiftRiver Platform does incorporate tools that use natural language processing and a gazetteer to return coordinate locations based on place names. The Ushahidi platform will extract geospatial metadata from social media feeds if it exists. For the Ushahidi 2010 Haiti earthquake map, the majority of information gathered came via SMS and not geo-tagged. Location information from SMS messages was translated by volunteers who used a variety of resources to obtain coordinate locations from the translated messages. Many of the volunteers were originally from Haiti and used their own geographical knowledge of the region combined with Open Street Map to pinpoint extract coordinates (Heinzelman &

Waters, 2010). For the Ushahidi 2010 Gulf Oil Spill map we were unable to determine which specific methods were used to geo-locate implicit location information.

The software platform used by TtT extracts geospatial metadata using the Twitter API if such metadata exists. The software filters for location tags prescribed in the TtT micro-syntax or tags identifiable as spontaneously generated by the crowd that may include implicit location information, for example #loc or #lat and #long. These tags and the data after each tag were parsed into key-value pairs to populate the database. Location pairs, along with identified place names or event tags like Joplin or Boulder were geocoded using GeoKit (geokit.rubyforge. com), which can geocode textual information across a number of different geocoding services. Volunteers could review the resulting coordinate pairs, which were then entered into the database if approved.

COMPARING USHAHIDI AND TWEAK THE TWEET

Here we draw comparisons between Ushahidi and TtT in three key dimensions. First, we describe what types of data and variables are captured by each effort. Next, we compare the interactive mapping tools that each platform provides. We conclude our comparisons by characterizing how each platform has resulted in tangible actions by responders and emergency managers. We use four recent disasters in these comparisons. For Ushahidi, we explore its use in the 2010 Haiti earthquake and 2010 Gulf Oil Spill. For TtT we focus on the 2011 Joplin tornado and 2010 Boulder Fourmile Canyon fire. We were unable to find directly overlapping events for both platforms. Ushahidi deployments tend to focus on larger disasters rather than the localized events focused on by TtT. For one overlapping event, the 2010 Haiti earthquake, TtT had not yet implemented map-

ping tools, and in other overlapping events (like 2010 Pakistan floods) TtT has integrated their efforts with Ushahidi.

Raw Data

Data generated during a disaster from social media networks tend to be ephemeral, and if it is not collected during the disaster, it can be difficult to conduct related research after the fact. Collecting raw data from Twitter older than two weeks has become challenging due to changes in the Twitter API that forbid certain types of archiving. Here, we conduct our analysis using the spreadsheets gathered from each application and additional analytical results from the PeopleBrowsr (www.peoplebrowsr. com) service which provides 1000 days of social media content and social analytics for marketers (not including SMS). We did not include PeopleBrowsr analytics for the 2010 Haiti earthquake because that data collected was primarily from SMS, and PeopleBrowsr analytics are not available for the 2010 Boulder Fourmile Canyon fire due to a small number of relevant reports.

In Table 2 we list all fields we discovered in the TtT and Ushahidi spreadsheets and our interpretation of the definitions for each field type for each application. Common fields which we think share a common meaning across both platforms are highlighted. The Ushahidi platform has fewer fields (eight vs. twenty-five for TtT) and they do not vary between the two incidents. TtT has variation in field names and the number of fields. We note that it is difficult to differentiate between the terms Status, Actionable and Verified in the TtT fields.

In the content summary shown in Table 3 the Ushahidi field "Approved" always shows a rating of 100%. According to Ushahidi documentation all messages are "approved" once valid location coordinates are determined and an administrator approves the content. Reports that are not yet approved are not displayed. The

Table 2. Comparing data collected from TtT and Ushahidi

Field Type	Ushahidi Field Name	TtT Field Name Boulder[1] Joplin[2]	Definition - TtT	Definition - Ushahidi
Record ID	#	Record ID[1] / ID[2]	Unique identifier	Unique identifier
Event		Event[1]	Event Hashtag – used for Place location	
Categorization	Category	Report Type[1,2]	Primary Hashtag – only one allowed – used for Key legend in Web Maps	Multiple categories allowed – used for Web Map Category Filter in Legend
Report	Incident Title	Report [2]	Partial parsed tweet with hashtags removed for pop-up display	Report Title for Web Report
Details		Details [1]	Partial parsed tweet with hashtags removed for pop-up display	
Original Report	Description	Text[2]	Original tweet	Original Message (in original language and translated if necessary)
Date/Time Stamp	Incident Date	Time[1,2]	Tweet time stamp	Message time stamp
Date_Time		Date_Time [2]	Time contained in tweet message	
Info		Info[1]	Volunteer added comment	
Source		Source[1,2]	Twitter user	
Contact		Contact[1,2]	Name, number, web page or other contact info contained in tweet	
Completed		Complete[1,2]	Indication if report was acted upon	
Status		Status[1]	? All N/A	
Verification	Verified	Verified[1]	? All N/A	Corroborated via incident report credibility vote
Actionable		Actionable[1]	? All N/A	
Approved	Approved			Map location approved
Author		Tweet Author[1] / Author[2]	The author of the record in the spreadsheet or author of retweet	
Tweet		Tweet[1]	Original Tweet	
Photo URL		Photo URL[1], Photo[2]	URL to photo	
Video		Video [2]	URL to Video	
Location (Text)	Location	Location[1,2]	Parsed location string	Parsed location string
Mapped		Mapped[1]	? All N/A	
Longitude	Longitude	GPS Long[1,2]	Derived Longitude	Derived Longitude
Latitude	Latitude	GPS Lat[1,2]	Derived Latitude	Derived Latitude

"verified" field indicates a report is submitted by or corroborated by a trusted source or an administrator. Table 3 shows that 6% and 40% of the records were corroborated in the 2010 Haiti earthquake and 2010 Gulf Oil Spill events. Raw data from TtT did not reveal the meaning of the codes "Status", "Actionable", and "Verified."

The 2010 Gulf Oil Spill Ushahidi spreadsheet lists the first incident date eleven days after the Deepwater Horizon explosion. The total number of tweets with the #oilspill keyword from April 10th to October 18th, 2010 according to PeopleBrowsr, is 22,199. The Louisiana Bucket Brigade collected 2952 reports according to their spreadsheet, representing approximately 13% of the total Twitter traffic by the PeopleBrowsr estimate. Of note is that all 2952 reports were geo-located. Additionally, there were only 9 tweets on the day after the explosion and no tweets for the next 17 days. The traffic over six months highlights the extended nature of the disaster.

The 2011 Joplin tornado data starts the day after the tornado and ends 27 days after the tornado. According to PeopleBrowsr there were 333,387 total Twitter mentions of the #Joplin keyword from May 13 to June 13th, 2011 (Figure 1). TtT identified 504 tweets that were entered into the spreadsheet. This represents approximately 0.02% of the total Twitter traffic if the PeopleBrowsr estimates are correct.

This highlights the challenge associated with harvesting social media communications during temporally-limited crises. It is also interesting to note that 65% of the 504 records in the TtT spreadsheet for the 2011 Joplin tornado and 54% of the 522 records for the 2010 Boulder Fourmile Canyon fire included locations. Examination of the raw data reveals frequent communication between volunteers that was not mapped because it was irrelevant to the event itself.

Maps

Cartographic representation of crisis mapping represents another challenge in the use of social media communications for disaster response because of the need to display large volumes of data while avoiding information overload. This is complicated further by the fact that potential users of crisis maps, including citizens, responders, volunteers, journalists and managers will have different expectations influenced by their social and physical relation to the crisis event (Liu & Palen, 2010). Field and O'Brien (2010) recognize that given the growth of social media communications and the geospatial component integral to an interconnected world, good cartography is crucial for creating maps with a purpose that are more than one-dimensional.

To compare crisis maps created by TtT and Ushahidi we first look at how the data is displayed for each set of incident maps;

Table 3. Summary of Ushahidi and TtT spreadsheet content

	Incident	Incident Date	Reports	First Report Date	Last Report Date	% Verified	% Approved	% Actionable	% Complete	%LAT/ LONG
TtT	Joplin Tornado[2]	5/22/2011 5:34 PM	504	5/23/2011 12:11 AM	6/13/2011 11:10 PM	N/A[1]		N/A[2]	0.4%[2]	65%
	Boulder Fourmile Fire[1]	9/6/2010 10:00 AM	522	9/8/2010 5:50 PM	9/17/2010 9:33 PM			N/A[1]	N/A[1]	54%
Ushahidi	Haiti Earth-quake	1/12/2010 4:53 PM	3589	1/12/2010 4:08 AM	5/18/2010 4:26 PM	6%	100%			100%
	Gulf Oil Spill	4/10/2010 10:00 PM	2952	4/21/2010 1:44 PM	10/18/2010 10:07 PM	40%	100%			100%

Figure 1. Tweak the tweet's mapping interface showing data for the 2010 Boulder fire
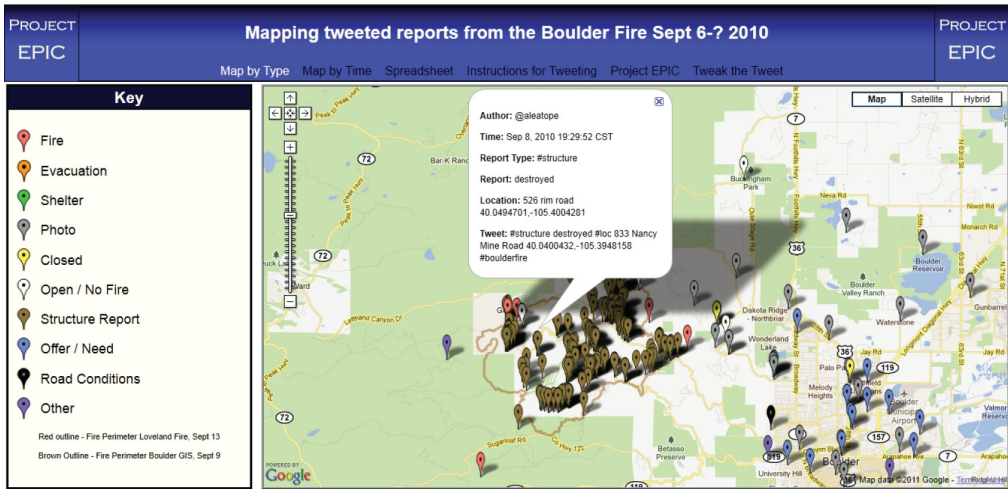
examine the layout, legends and tools for map interaction; and examine the map's reporting and temporal capabilities. Then we compare the overall cartographic qualities and utility of the representations between the TtT and Ushahidi maps.

TtT Maps

The TtT Boulder Fourmile Canyon Fire Map shown in Figure 1 is based on the TtT template and provides mechanisms to display both an incident map and a temporal map. Both map options use Google Maps as a base map with the standard Google map controls for map navigation and alternative base map layers. Incidents are symbolized using standard Google marker symbols based on report type, and these correspond to the primary hashtags prescribed by TtT for the crisis. An "Instructions for Tweeting" link contains information on how to correctly contribute Tweets using prescribed hashtags for the Boulder fire. There is also a link to a downloadable spreadsheet and the Project Epic home page. All the fields in the spreadsheet are presented in the popup bubble in the main map view with the exception of Event and GPS Lat/GPS Long fields. The legend is one-dimensional and does include sub-layers, though there is no

interaction to turn individual categories layers on and off. There is no reporting mechanism provided with the map other than the pop-up bubble for each specific report, and there is no direct mechanism for users to submit an incident report from the web map interface. One noteworthy feature is TtT's "Map by Time" capability, which provides polygons to represent the extent of a particular crisis and categorizes/colors individual incident reports by age.

In support of crisis response and recovery efforts for the 2011 Joplin Tornado, TtT offered a slightly different version of their mapping system. While the main template remains quite similar to previous versions (Figure 2), rather than using the default Google Maps marker symbols, incidents are plotted with simple color points based on the report type and correspondence to the primary hashtags prescribed by TtT for the crisis. There is an additional item in the pop-up bubble for "Pic" which can signal the attachment of a photo or video.

Ushahidi Maps

Ushahidi's basemap uses OpenStreetMap by default, and the map includes scale, latitude/longitude, and coordinate reference indicators as well as standard web map navigation tools

Figure 2. Tweak the tweet's mapping interface showing data from the 2011 Joplin tornado
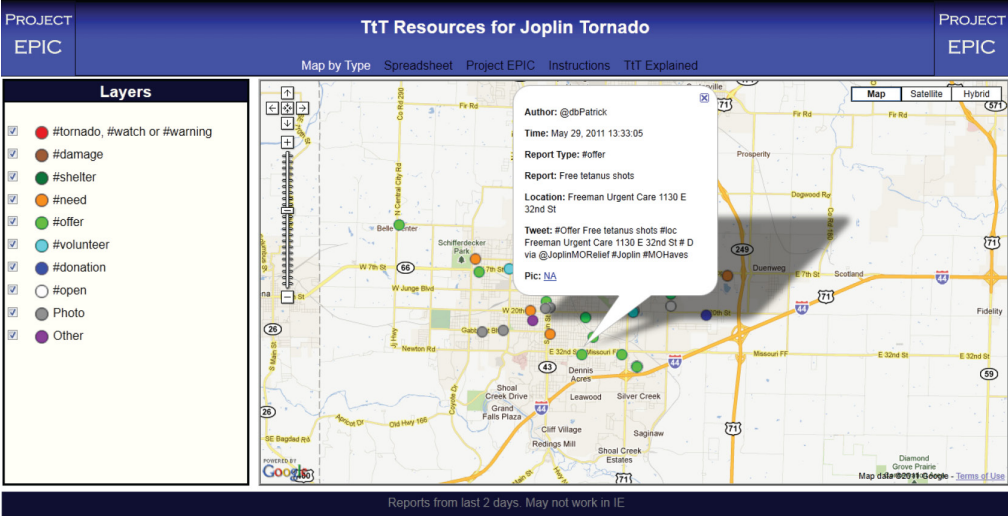
(Figure 3). A collapsible basemap selection tool in the upper right corner offers additional basemaps to show Google Streets, Satellite imagery, and Haiti Terrain maps in color and grayscale. The tool also includes overlays for Haiti Roads (major road only – shown by default), Haiti Obstacles, Haiti Fault Lines, Haiti Reports (shown by default), and Haiti MMI layers (Modified Mercalli Intensity scale) showing the epicenter and magnitudes of the earthquake and the major aftershocks. There is a filter control for the map and the default view shows clusters. There is an animated temporal tool and graph that will display the incident reports as they change over time based on a user specified date range. Below the temporal tool there are four more layers that can be overlaid on the map – reported aftershocks, medical facility locations, displacement zones, and damaged structures. The tool highlights how to submit a report in several places on its interface. Incidents can be reported via an SMS short code, sent to a generic email address, and or a submitted via a web form. The menu also includes links to a list of all reports, an alert link that allows a user to sign up for email or text alerts based on incident type and a location radius, and a link listing the organizations that

can be contacted to donate money, goods, and services, or to volunteer. Finally, the application includes a report search function and the ability to subscribe to an RSS feed of reports.

Incidents are symbolized in clusters and zooming in or out will adjust the clusters based on scale with a large cluster dispersing to smaller clusters as the scale gets larger. The legend is interactive and allows a user to expand each category (listed in French and English) and supports turning sub-layers on or off. Sub-layers within each category have unique symbol designs and are also clustered.

In support of response and recovery efforts for the 2011 Gulf Oil Spill, Ushahidi's Louisiana Bucket Brigade Map (Figure 4) uses the same basic template as the Haiti instance with some subtle changes in its design. The map uses Google Maps as a base map and offers no other base layer options. Additional thematic layers are listed in a separate layer control tool called "Other layers," rather than through the use of a category/filter layer control. The map header is customized to the organization and the incident reporting options are listed at the top of the page. Download options are available to acquire or subscribe to the data in multiple formats, including RSS, KML, JSON and CSV.

Figure 3. Example map from Ushahidi from the 2010 Haiti earthquake
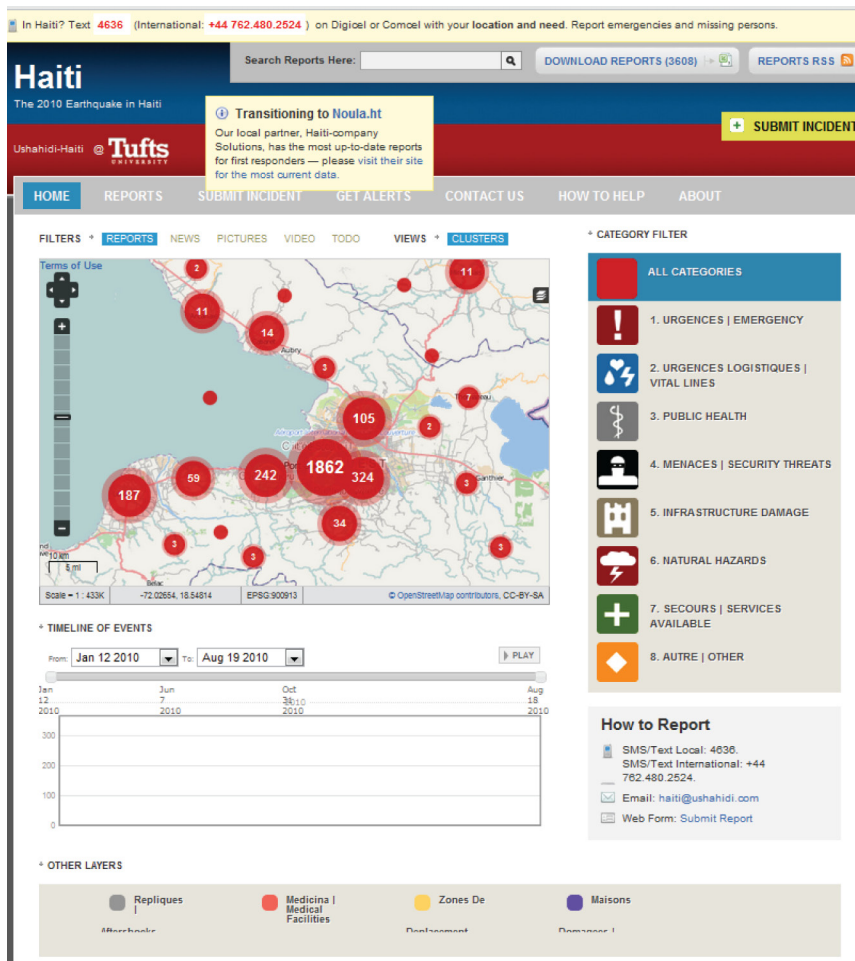
Discussion

Unlike the Ushahidi map, the icons for the layers in the legend are simple colored squares, but when selected some of them result in a more detailed icon marker appearing on the map.

Crisis maps created by Ushahidi and TtT are quite similar in terms of their core features. Both platforms utilize simple interactive map mashups and categorized point symbols to represent reports. Ushahidi has the ability to generalize dense sets of reports into aggregated symbols, making it scalable to larger datasets.

Both platforms have recently introduced temporal displays to highlight report frequency over time (frequency graph in Ushahidi and time-categorized markers in TtT). The overall map and interface aesthetic is significantly more refined in current implementations of Ushahidi, perhaps reflecting its relative maturity compared to TtT.

The creators of the TtT maps and Ushahidi share the same goal to create effective tools that enable managing of crowd sourced information to benefit those most affected by a disaster. In the case of TtT, the original intent

Figure 4. Ushahidi's Louisiana Bucket Brigade map from the 2011 Gulf oil spill
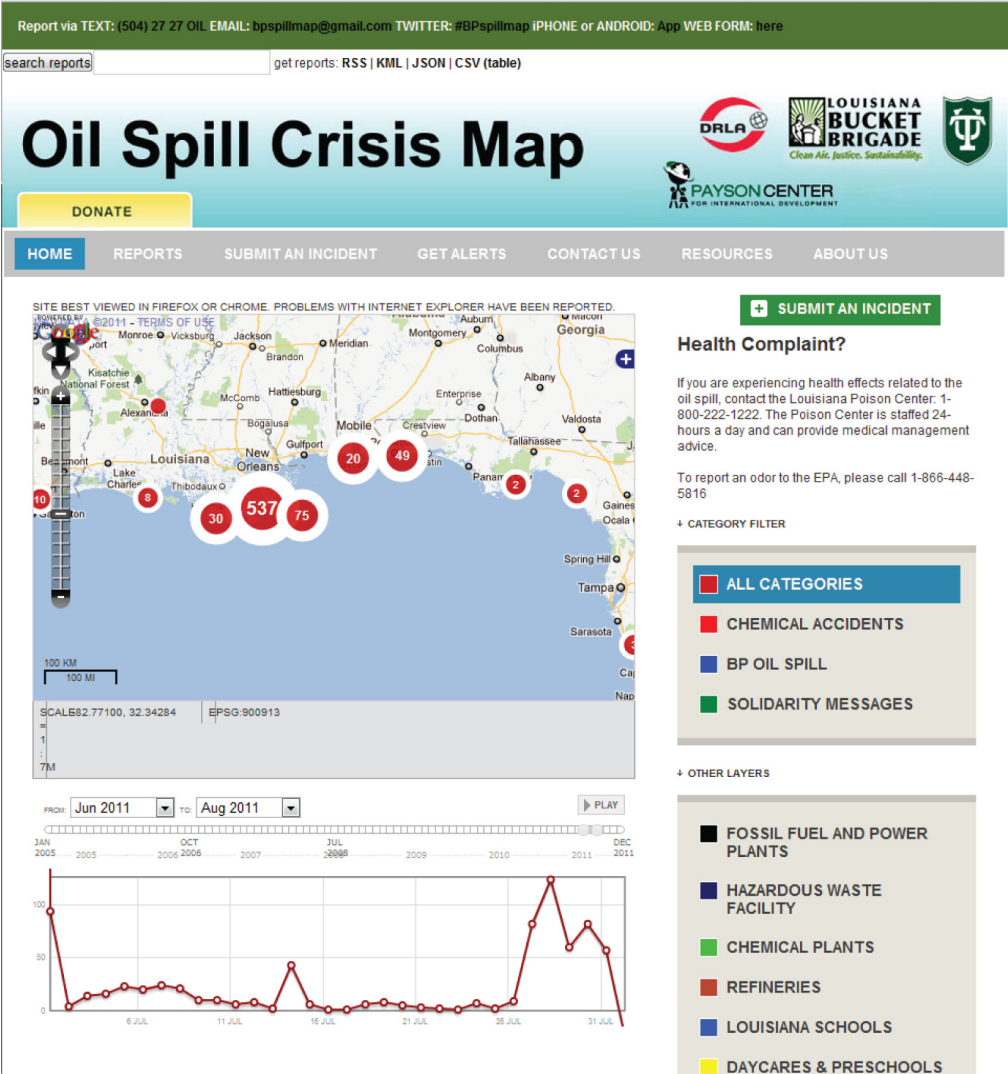
was to develop a micro-syntax for Twitter that would better enable automated processing of social media communications data generated during a disaster. Visualizing the data was not their primary focus. Ushahidi, on the other hand, began as an application designed to map election violence and human rights abuses, and the focus was not on the communications themselves, but on mapping and visualization of those communications to increase awareness through the media to the general public.

On the whole, neither platform supports significant geospatial analysis capabilities. Basic filtering controls are available to winnow the dataset, but there are no quantitative spatial analysis methods available to identify clusters or to compare current patterns to past patterns. Alerting functions are limited to the presence of new reports from a particular category, not based on a careful analysis of what is observed versus what is expected. A significant difference between platforms is that Ushahidi provides

alerting tools for users to "listen" for reports from a given area or matching a given set of thematic criteria.

Spatial data interoperability in both platforms is supported through spreadsheet downloads of raw data, making it possible for users to ingest collected information into a full-featured GIS if necessary.

Evidence of Action

Understanding the effectiveness of efforts like TtT and Ushahidi is a difficult task. We concentrated our efforts on identifying impacts from media reports, after action reports, and TtT and Ushahidi's own assessments.

Project Epic and TtT gained recognition for their efforts during the 2010 Boulder Fourmile Canyon fire when CNN ran a story which credited TtT for integrating crisis information through the use of volunteers and social media (Spellman, 2010). One article cites Project Epic's use of a map with geo-located tweets to tracking the fire movements from citizen reports (Orlando, 2011). Another news report (Petty, 2010) describes how TtT was used to gather and map data from the 2010 Boulder Fourmile Canyon fire, including information not provided through official emergency response channels. We did not find many media reports of the specific use of TtT after the 2011 Joplin tornado, but the use of the syntax was promoted by organizations like Crisis Commons.

Following the 2010 Haiti earthquake, Craig Fugate, Director of FEMA, tweeted that the Ushahidi Haiti Map was the most "comprehensive and up-to-date map available to humanitarian organizations" (Heinzelman & Waters, 2010). Newsweek's profile of the Ushahidi efforts after Haiti indicate that the crisis map resulted in saving lives (Ramirez, 2010). One after action report (UN-SPIDER, 2010) notes that the US Marines used Ushahidi to coordinate locations and direct relief efforts. They also indicate that data from Ushahidi was used to direct Coast Guard responders for search and rescue. The Ushahidi Blog (blog. ushahidi.com) highlights multiple examples

of action items and response efforts generated by Ushahidi, including food and water deliveries. Finally, a story from The Alumni Boom, a quarterly newsletter from the US Embassy in Ashgabat, Turkmenistan (Matsapayev, 2010) relates one very specific example:

"...crisis mappers learned of a clinic that needed diesel fuel in Haiti; after the clinic's location and specific needs were posted on a website called Ushahidi, the State Department and FEMA discovered the need and fuel was delivered."

The 2010 Gulf Oil Spill Ushahidi map represented a different use of this technology in disaster response. The oil spill was not as much a direct threat to lives as it was a threat to local economies and the environment. Use of the map has been primarily to raise awareness of the ongoing ecological disaster and to document the damage. One article highlighted the fact that the information collected would be useful in long-term recovery efforts and could also be used in future legal actions over long-term damage to ecosystems and livelihoods (Sutter, 2010).

A common thread throughout the reports we reviewed is that the impact of geospatial social media platforms on tangible emergency response actions is not yet well-defined. While both have received media attention and have clearly captured public interest, there are few specific examples of the information leading to different decision-making patterns, widespread allocation of resources, or information leading to the rescue of disaster victims.

FUTURE DESIGN CONSIDERATIONS AND CONCLUSIONS

The recent use of social media communications in disaster response is largely driven by volunteer organizations. Engagement with volunteers during the 2010 Haiti earthquake by TtT prompted TtT developers to focus research on a second layer of crowdsourcing: communication

between volunteers and response organizations (Starbird, 2011). By working with crisis volunteer organizations TtT has continued to promote the use of the micro-syntax after a disaster. Spurred by lessons learned from deployments of crisis mapping efforts like Ushahidi, the Standby Task Force was recently formed to organize volunteers and provide a dedicated technical interface to the humanitarian community to assist in dealing with new sources of information like social media.

A key design consideration going forward is to ensure effective mechanisms for disseminating and sharing information between first responders and crisis managers. A UN report suggests that efforts like Ushahidi and TtT may contribute to information overload, citing that a large percentage of the information gathered was already in the hands of relief organizations on the ground (UN-SPIDER, 2010). This does not take into consideration the value of mapping the information and providing crisis managers a mechanism to identify clusters where relief organizations and first responders could concentrate their efforts.

We have outlined several reasons why extracting the location from social media is difficult, and further research is needed in this area, especially in the role of automatic geocoding and disambiguation of text descriptions of place. Ushahidi point out that from 40,000 Haiti-related SMS messages only 5% were mapped (Norheim-Hagtun & Meier, 2010), but that does not mean that the other 95% did not contain any geospatial information. To support analytical reasoning and geospatial analysis we should be able to uncover patterns that reference physical and cultural regions, types of landforms, directional information and topological relations in addition to basic point locations. A range of recent research focuses on the challenges of extracting actionable information from social media. Munro (2011) proposes models to systematically identify actionable items using trending categories or topics, subwords, and spatiotemporal clusters. MacEachren et al. (2011) discuss how SensePlace2, a geovisual analytics application, includes a crawler ap-

plication designed to systematically query the Twitter API based on crisis-relevant keywords and phrases. SensePlace2 also mines the Tweets themselves to identify possible mentions of locations and geocodes those locations to explore which places are talked about in the context of a crisis. Vieweg et al. (2010) have recently described an automated methodology to detect messages that will be useful for situational awareness.

Credibility and verification of information is another area that needs to be addressed in future research. One report indicates that search teams found a high proportion of SMS reports about trapped wounded victims turned out to be coming from families wanting to recover their dead relatives (Harvard Humanitarian Initiative, 2011). Some recent research has focused on identifying ways in which credibility might be automatically assessed in Tweets by evaluating message content, user profile details, and message propagation (Castillo, Mendoza, & Poblete, 2011).

Finally, we must develop effective cartographic representation techniques to ensure the usability of web maps for crises. A particular challenge for crisis mapping is that there are a wide range of expectations and technical skills associated the diverse group of people that need to use crisis maps, including citizens, responders, volunteers, journalists and managers. Not all groups are equally equipped to evaluate the results of geographical analysis. Maps are likely to be seen as credible evidence, even when the underlying data is of unknown quality. The maps we reviewed from TtT and Ushahidi are simple, and therefore likely to be easily understood by people without substantial cartographic training. However, their simplicity limits their true utility in crisis situations where detecting emerging significant clusters of events, recognizing outliers when they appear, and predicting future outcomes all become essential geographical analysis products. Making these much more sophisticated products available to a wider group of novice map users remains a major research challenge.

There is no doubt that the contribution of social media communications during disaster has shifted the paradigm of emergency response to include at least a one-way social media dialog from those most affected. Mapping social media content provides a way to gather and visualize information from what can arguably considered the true first responders - the affected citizens who are the first to assess the situation and request assistance through social media. Data coming directly from the affected population can and should contribute to emergent situation awareness (Gunawan, Fitrianie, Brinkman, & Neerincx, 2012). Driven by volunteers and advances in web-based technology, the proliferation of this information has grown faster than the analytical capabilities of disaster management organizations and workflows. TtT has contributed a method to filter, automate and direct information from social media sources during a disaster and Ushahidi has proven to be an effective and widely adoptable platform for displaying geospatially-oriented social media communications. However, TtT and Ushahidi have only tackled simple location-related problems and provided only rudimentary situational awareness and mapping capabilities to visualize the social media communication stream. Future research must focus on applications that go beyond basic crowdsourcing to develop information collections, analytical tools, coordination of communications, and mapping visualization to support all phases of disaster management. Future platforms developed with the volunteer community in mind will need to incorporate social media as one piece of an overall strategy to support situational awareness and response and recovery featuring effective two-way communications with citizens through social media.

ACKNOWLEDGMENT

REFERENCES

American Red Cross. (2011). Social media in disasters. Retrieved October 15, 2011, from http://www.redcross.org/www-files/Documents/pdf/SocialMediainDisasters.pdf

Anderson, K. M., & Schram, A. (2011). *Design and implementation of a data analytics infrastructure in support of crisis informatics research (NIER track)*. Honolulu, HI: Waikiki. doi:10.1145/1985793.1985920.

Bellucci, A., Malizia, A., Diaz, P., & Aedo, I. (2010, May 2-5). Framing the design space for novel crisis-related mashups: The eStoryS example. In *Proceedings of the Information Systems for Crisis Management and Response (ISCRAM 2010)*, Seattle, WA.

Castillo, C., Mendoza, M., & Poblete, B. (2011). *Information credibility on twitter*. Hyderabad, India.

Currion, P., de Silva, C., & Van De Walle, B. (2007). Open source software for disaster management. *Communications of the ACM*, *50*(3), 61–65. doi:10.1145/1226736.1226768.

Dosemagen, S. (2010, 5/8/2010). *Ushahidi used to create oil spill crisis map*. Retrieved from http://blog.ushahidi.com/index.php/2010/05/08/labb/

Field, K., & O'Brien, J. (2010). Cartoblography: Experiments in using and organising the spatial context of micro-blogging. *Transactions in GIS*, *14*(1), 5–23. doi:10.1111/j.1467-9671.2010.01210.x.

Gosier, J. (2011). *Key deployments and lessons learned - part 1*. Retrieved from http://blog.ushahidi.com/index.php/2011/03/21/key-deployments-and-lessons-learned-part-1/

Gunawan, L. T., Fitrianie, S., Brinkman, W.-P., & Neerincx, M. A. (2012, April 22-25). Utilizing the potential of the affected population and prevalent mobile technology during disaster response: propositions from a literature survey. In *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2012)*, Vancouver, Canada.

Harvard Humanitarian Initiative. (2011). *Disaster relief 2.0: The future of information sharing in humanitarian emergencies*. Retrieved from http://issuu.com/unfoundation/docs/disaster_relief20_report

Heinzelman, J., & Waters, C. (2010). Crowdsourcing crisis information in disaster-affected Haiti. United States Institute of Peace, 252, 1-16.

Intagorn, S., & Lerman, K. (2011). Mining geospatial knowledge on the social web. *International Journal of Information Systems for Crisis Response and Management*, *3*(2), 33–47. doi:10.4018/jiscrm.2011040103.

International Telecommunication Union. (2011). *Key global telecom indicators for the world telecommunication service sector*. Retrieved 10/15/2011, 2011, from www.itu.int/ITU-D/ict/statistics/at_glance/KeyTelecom.html

Jennex, M. (2010). Implementing social media in crisis response using knowledge management. *International Journal of Information Systems for Crisis Response and Management*, *2*(4), 20–32. doi:10.4018/jiscrm.2010100102.

Jennex, M. (2012). Social media - viable for crisis response? Experience from the great San Diego/Southwest blackout. *International Journal of Information Systems for Crisis Response and Management*, *4*(2), 53–67. doi:10.4018/jiscrm.2012040104.

Lang, G. (2010). The use of social media in disaster situations: framework and cases. *International Journal of Information Systems for Crisis Response and Management*, *2*(1), 11–23. doi:10.4018/jiscrm.2010120402.

Li, L., & Goodchild, M. (2010). The role of social networks in emergency management: A research agenda. *International Journal of Information Systems for Crisis Response and Management*, *2*(4), 48–58. doi:10.4018/jiscrm.2010100104.

Liu, S. B., & Palen, L. (2010). The new cartographers: Crisis map mashups and the emergence of neogeographic practice. *Cartography and Geographic Information Science*, *37*(1), 69–90. doi:10.1559/152304010790588098.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., et al. (2011, October 23-28). SensePlace2: Geotwitter analytics support for situation awareness. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, Providence, RI.

Matsapayev, M. (2010). *The alumni bloom - Quarterly newsletter of the U.S. embassy in Ashgabat*. Retrieved from http://photos.state.gov/libraries/turkmenistan/49351/pdf/al-boom-2010-spring.pdf

Meier, P., & Munro, R. (2010). The unprecedented role of SMS in disaster response: Learning from Haiti. *SAIS Review*, *30*(2), 91–103.

Munro, R. (2011). Subword and spatiotemporal models for identifying actionable information in Haitian Kreyol. In *Proceedings of the 15th Conference on Computational Natural Language Learning (CoNLL '11)*, Portland, Oregon.

Norheim-Hagtun, I., & Meier, P. (2010). Crowdsourcing for crisis mapping in Haiti. *Innovations*, *5*(4), 81–89. doi:10.1162/INOV_a_00046.

Okolloh, O. (2009). Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action*, *59*(1), 65–70.

Orlando, J. (2011). Turning disaster response on its head. *Continuity Insights*. Retrieved from http://www.continuityinsights.com/articles/turning-disaster-response-on-its-head

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010, April 13-16). A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. In *Proceedings of the ACM-BCS Visions of Computer Science Conference*, Edinburgh, UK.

Petty, D. (2010, September 9). Evacuees use social media to keep up on Boulder wildfire disaster developments, *Denver Post*.

Ramirez, J. (2010). 'Ushahidi' technology saves lives in Haiti and Chile. *Newsweek*.

Reuter, C., Marx, A., & Pipek, V. (2012). Crisis management 2.0: Towards a systemization of social software use in crisis situations. *International Journal of Information Systems for Crisis Response and Management*, *4*(1), 1–16. doi:10.4018/jiscrm.2012010101.

Spellman, J. (2010). Heading off disaster, one tweet at a time. *CNN Tech*. Retrieved from www.cnn.com/2010/TECH/social.media/09/22/natural.disasters.social.media/index.html

Starbird, K. (2011). Digital volunteerism during disaster: Crowdsourcing information processing. In *Proceedings of the CHI '11 Workshop on Crowdsourcing and Human Computation*, Vancouver, BC.

Starbird, K., & Palen, L. (2011). "Voluntweeters": Self-organizing by digital volunteers in times of crisis. In *Proceedings of the SIGCHI Conference on Human factors in computing systems (CHI '11)*, Vancouver, BC, Canada.

Sutter, J. D. (2010). Citizens monitor Gulf Coast after oil spill. *CNN Tech*. Retrieved from articles.cnn.com/2010-05-06/tech/crowdsource.gulf.oil_1_oil-spill-gulf-coast-jeffrey-warren

UN-SPIDER. (2010). Lessons from Haiti. *Coordinates*, *6*(5), 27–31.

Ushahidi. (2011). *Ushahidi API*. Retrieved October 15, 2011, from http://wiki.ushahidi.com/doku.php?id=ushahidi_api

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human factors in computing systems (CHI '10)*, Atlanta, GA.

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, *13*(5), 788–806. doi:10.1177/1461444810385097.

Susan McClendon is part of the University of Wyoming's Information Technology Division and is the primary systems support staff for the Wyoming Geographic Information Science Center (WyGISC). Her efforts are currently focused on building a cloud based enterprise GIS cyberinfrastructure to support research efforts involving geospatial information and technology. Since graduating from Penn State in 2011 with a Master's in Geographic Information Systems, McClendon has begun exploring research opportunities in automating the integration of social media into crisis mapping tools and more recently on the integration of place-based social media data as a tool for GIS education in the classroom, including participatory sensing and citizen science.

Anthony C. Robinson is the Assistant Director for the Penn State Department of Geography's GeoVISTA research center. Robinson's research focuses on the science of interface and interaction design for geographic visualization software tools. He has developed interface design and usability assessment methods for integrating geographic visualization tools with work in epidemiology, crisis management, and homeland security. Robinson's recent research projects have focused on the design of map symbol standards for crisis management, developing tools for collecting and adding meaning to geographic information, and eye-tracking to design new geovisualization techniques. In addition, Robinson serves as the lead faculty member to direct Penn State's Master of GIS and post-baccalaureate GIS Certificate programs.

Leveraging Geospatially-Oriented Social Media Communications in Disaster Response

Susannah McClendon
GeoVISTA Center
Department of Geography
The Pennsylvania State University
susantbm@gmail.com

Anthony C. Robinson
GeoVISTA Center
Department of Geography
The Pennsylvania State University
arobinson@psu.edu

ABSTRACT

Geospatially-oriented social media communications have emerged as a common information resource to support crisis management. Our research compares the capabilities of two popular systems used to collect and visualize such information - Project Epic's Tweak the Tweet (TtT) and Ushahidi. Our research uses geospatially-oriented social media gathered by both projects during recent disasters to compare and contrast the frequency, content, and location components of contributed information to both systems. We compare how data was gathered and filtered, how spatial information was extracted and mapped, and the mechanisms by which the resulting synthesized information was shared with response and recovery organizations. In addition, we categorize the degree to which each platform in each disaster led to actions by first responders and emergency managers. Based on the results of our comparisons we identify key design considerations for future social media mapping tools to support crisis management.

Keywords

Geographic Information, Social Media, Crisis Management, Mashups.

INTRODUCTION

Crowd-sourced information has rapidly become an essential source of data in disaster response. Since the first well documented efforts of citizen journalists on September 11th, 2001 and the use of internet blogs to collect information after the 2004 East Indian Ocean Tsunami, recent emergency response efforts have included mapping SMS messages after the Haiti earthquake in 2010. The Haiti earthquake represented a paradigm shift in the use of social media for disaster response, as multiple web-based platforms emerged to collect, refine, and disseminate crisis-related social media. The use of social media to gain real time information on the ground in a disaster has been driven by the rapid speed at which information can be distributed, the cross-platform accessibility of information, and the ubiquity of social media worldwide (Vieweg, et al., 2010). The utility of this information has been enhanced by the creation of crisis maps based on location data extracted from social media communications (Liu and Palen, 2010, MacEachren, et al., 2011).

In 2011 the American Red Cross conducted a survey that showed that 33% of citizens have used social media sites, including Facebook, Twitter, Flickr and SMS text messages/alerts to gain information about an emergency (American Red Cross, 2011). About half the respondents said they would contribute information during an emergency using social media channels. Statistics from the International Telecommunication Union reveal that in 2009 there were 4.6 billion mobile phone subscribers world-wide and 1.5 billion subscribers used mobile devices to access the internet (International Telecommunication Union, 2011). We can reasonably expect the use of social media in disaster response to increase in the future.

Research on the increasing use of social media in disaster response has emerged as a new focus in the field of crisis informatics (Anderson and Schram, 2011). Extracting, categorizing, visualizing, and evaluating such information presents serious research challenges, including the problem of managing and extracting meaningful information from the large volume of contributions, applying the information to decision support workflows, and the development of formal information sharing protocols (Harvard Humanitarian Initiative, 2011). Mapping crowd sourced information in disaster response gained wide-scale media attention after the successful deployment of the Ushahidi Crisis Map during the 2010 Haiti earthquake (Starbird, 2011). There are a number

Proceedings of the 9ᵗʰ International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

2

of specific challenges involved in mapping social media communications, including the extraction of accurate location information, and the application of useful and usable cartographic representations to visually support situational awareness in crises. Research on the integration of Geographic Information Systems (GIS) and crowdsourced information from social media has focused more on the challenges of extracting action items and location information from social media feeds (MacEachren, et al., 2011) and less on the utility of the extracted information and the effectiveness of associated crisis maps to support emergency response.

Our research examines two applications that have leveraged geospatially-oriented social media during recent disasters; the Tweak the Tweet (TtT) project from Project Epic (Starbird and Palen, 2011) and Ushahidi (Okolloh, 2009), both of which have been used to create crisis maps of content collected from social media sources during recent disasters. For each application we examine collected data, information products, and evidence of subsequent response actions for two recent disasters; the 2011 Joplin tornado and 2010 Fourmile Canyon fire for Tweak the Tweet, and the 2010 Haiti earthquake and 2010 Gulf Oil Spill for Ushahidi. Other efforts have used crowdsourced information during recent disasters, including the open source Sahana platform (Currion, et al., 2007) and the collective effort of the Crisis Mappers Network (crisismappers.net). However, Sahana utilized data feeds directly from Ushahidi and TtT, and the Crisis Mappers Network is focused on connecting and empowering crisis collaborators, and does not offer their own specialized technology platform.

We begin with background information, including an overview of geospatially-oriented social media, followed by a brief history of TtT and Ushahidi. Next, we evaluate how each organization collected, processed and geo-located these social media communications and compare and contrast the cartographic representations and reporting capabilities of the resulting crisis maps. We then examine the effectiveness of each application by identifying examples of actionable items used by military, government and non-government organizations that emerged from the use of these crisis maps. Finally, we conclude with key design considerations for future efforts to leverage geospatially-oriented social media in crisis informatics.

GEOSPATIALLY-ORIENTED SOCIAL MEDIA

Many social media sources, including Twitter, allow users to tag reports with coordinates to indicate their location on Earth. This information is easy to process and represent on a map, but does not necessarily represent an "actionable" location. A more substantive challenge is associated with making use of textual descriptions of place (placenames and less-specific geographic features), like those often included in an SMS text message. Placenames in text can be geocoded to assign location coordinates, but placenames are usually associated with irregular areas (for example, the New York City metro area) at least as often as one might ever associate them with a specific coordinate location on earth (the centroid of the legal boundary of New York City). Determining less-specific geographic features is also of critical importance when using geospatially-oriented social media.

Both platforms make use of social media collected from SMS and Twitter messages. SMS or "text messaging" is a short messaging service that allows for the storage and retrieval of short 160 character messages across global cellular telephone networks. Location information is not automatically attachable to SMS data, and must be inferred from the message itself. Twitter is a microblogging service that allows users to post messages up to 140 characters called Tweets via mobile phones or web accessible devices. Twitter users follow other users to see their tweets in a Twitter feed. The Twitter user community has developed linguistic markers to facilitate communication; including the @ symbol to address users (@username); the RT abbreviation to represent a retweet (RT @username); and # or hashtag to indicate keywords. Hashtags allow users to search Twitter feeds or to follow trends (Zappavigna, 2011). Location information can be added to a tweet using a phone's GPS capabilities, and can also be inferred from user profiles and mentions of placenames in messages themselves.

PROJECT EPIC AND TWEAK THE TWEET

Project Epic is research effort at the University of Colorado that aims to improve methods of public information gathering and dissemination during emergency situations. The project's mission is to couple computational methods with behavioral knowledge on how people develop information using social media in crisis situations (Palen, et al., 2010). One Project Epic research project, Tweak the Tweet, was first presented in 2009 as a simple set of standardized communication practices coupled with a technology platform for making sense of crisis Tweets (Starbird, 2011, Starbird and Palen, 2011). TtT asks users to tweet using a crisis specific micro-syntax designed to enable real-time processing of Tweets. TtT features a web-based tool for collecting and visualizing contributed information using the Twitter API to continually-update a database. The categorized information is displayed on a simple map mashup using the Google Maps API.

Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

3

The TtT micro-syntax is based on primary or main hashtags that can be used in any crisis situation and are designed to indicate the "who, what, and where" of the Twitter message content. For example, #name or #contact can be used to indicate "who"; #need, #shelter, #road, #open, #damaged can be used to indicate "what"; and #loc can used indicate "where". These hashtags are used in conjunction with an event tag to organize the crisis. Event tags can be spontaneously generated during an event, like #joplin or #tornado, or prescribed by the TtT micro-syntax like #4MileFire. Used together, the primary and event hashtags format meaningful machine readable tweets.

TtT was first deployed in the aftermath of the Haiti earthquake in January 2010 with the goal of having responders and agencies on the ground use the syntax. The first deployment did not have any associated mapping functionality, and the micro-syntax was not widely adopted by first responders or the public. Despite that, volunteers from Crisis Commons, TtT and other organizations tweeted or retweeted almost 3000 unique tweets formatted with the TtT syntax (Starbird and Palen, 2011). Since 2009, TtT has added a mapping component to their system design and the application has been deployed for over twenty major crises.

USHAHIDI

Ushahidi began as a non-profit African technology company that was developed to map incidents of violence in Kenya following elections in 2008. Ushahidi's mission is to develop platforms for sharing crisis information and personal narratives (Okolloh, 2009) and has since grown to develop tools to facilitate the democratization of information in broader contexts. The open source software tools developed by Ushahidi automate the collection of incident reports using cellular phones, email, and the web and facilitate the mapping of report locations in an interactive map mashup along with descriptive data to contextualize events.

Ushahidi offers three core products: the Ushahidi Platform, the SwiftRiver Platform, and Crowdmap. The Ushahidi Platform combines interactive mapping with the ability to capture real-time data streams from mobile messaging services and Twitter, and also supports email and web forms. It also provides spatial and temporal views of collected data. The SwiftRiver Platform allows for the real-time filtering and verification of data from these multiple data streams, including the ability to automatically categorize information based on semantic analysis, provide analytics and insight into user relationships and data trends, facilitate information validation and qualification, and it offers an interactive dashboard for monitoring and reporting purposes. Crowdmap is a cloud-hosted solution designed to support rapid launches of both the Ushahidi and SwiftRiver platforms.

Since 2009, deployments of Ushahidi platforms have focused on election monitoring, reporting human rights violations, disease surveillance, wildlife tracking, and disaster response. Though there were several deployments of the Ushahidi platform prior to the 2010 Haiti earthquake, it was the Haiti crisis that brought Ushahidi international attention. Ushahidi adoption since the 2010 Haiti earthquake has seen significant growth.

DATA COLLECTION, MANAGEMENT, AND IDENTIFYING LOCATIONS

Both TtT and Ushahidi utilize technical and manual methods to collect, refine, and add meaning to data. The following sections describe how each platform is designed, how they manage data, and how they derive location information from collected social media reports.

Tweak the Tweet

After the initial launch of TtT during the 2010 Haiti earthquake, TtT refined its aims to promote *crowdfeeding* after analysis of results from the deployment in Haiti highlighted the difficulty of getting the crowd to adopt a micro-syntax for Twitter (Starbird and Palen, 2011). TtT promotes the monitoring of social media sites by volunteers (called voluntweeters) during a crisis and they disseminate information back into the crowd using the TtT micro-syntax. In addition, voluntweeters promote the use of the syntax through conversations with other responding organization volunteers and by posting instructions and links to TtT crisis maps on social media sites. For the events we researched, TtT prescribed a micro-syntax with event tags including #boulderfire, #boulder, #4MileFire, #joplin, and #tornado.

The TtT software platform utilizes the Twitter Streaming API to identify tweets based on the TtT micro-syntax and stores the tweets in a MySQL database, parsing the information into key-value pairs based on hashtags. The platform uses Google maps to map the tweet content after the location has been determined. At regular intervals a Ruby script parses the messages filtered by hashtags into a MySQL database and the script in turn updates a public Google spreadsheet. Because machine processing may miss meaningful data in the tweet, such as placenames and other locations, the TtT process uses a combination of automatic and manual processing by

Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

4

volunteers to populate data in the event spreadsheet. For this research we downloaded spreadsheets for the 2010 Boulder Fourmile Canyon fire and the 2011 Joplin tornado disasters. These TtT–hosted spreadsheets contain all the events collected for each disaster.

Ushahidi

In the hours following the 2010 Haiti earthquake, Ushahidi staff deployed the Ushahidi Haiti Crisis Map. Working in the United States, they gathered information from media reports and social media sources. Approximately 85% of Haitians had access to cellular telephones and the cellular telephone infrastructure, though damaged, was quickly repaired. Within days a SMS short code number was set up in collaboration with phone companies and U.S. State Department resources and advertised through local radio stations. The messages received via SMS were sent to an automated system set up to facilitate message translation and mapping of the data by volunteers (Heinzelman and Waters, 2010).

Shortly before the 2010 Gulf Oil Spill, students at Tulane University began development of a crisis map to document oil refinery accidents using the Ushahidi platform. On the day the class presented the GIS map, the Deepwater Horizon oil rig exploded in the Gulf of Mexico (Dosemagen, 2010). The Louisiana Bucket Brigade, an environmental organization, worked with Tulane students to launch the Oil Spill Crisis Map to give Gulf residents a chance to contribute information about threats to their community and ecosystem from the oil spill. Data for the map was submitted via SMS, Email, Twitter and web forms. Citizens were encouraged to make a reports based on health issues, wildlife sightings, and other notable impacts they may witness in the region.

The Ushahidi API (Ushahidi, 2011) supports data exchange in XML (Extensible Markup Language) and JSON (JavaScript Object Notation). Ushahidi software supports PHP scripting and is designed to work with MySQL and is usually run on an Apache web server. Ushahidi software can be configured to work with common SMS gateway providers to process and deliver SMS messages, and it can be configured to use the Twitter Streaming API to process Tweets. Data can be exported from MySQL via a PHP script to a Google spreadsheet. The Ushahidi map template is designed with a link to download the raw data in a Google spreadsheet, but because the Ushahidi platform is open source and can be modified by the organizations that deploy the software, not all organizations include the ability to download the data. For the 2010 Haiti earthquake and the 2010 Gulf Oil Spill, we were able to download spreadsheets with data covering six months after the initial incidents.

Streaming Data and Scalability Challenges

Collecting data from social media communications like Twitter and SMS is difficult due to the large datasets that can be generated in a short amount of time. A key challenge has emerged in automating the extraction of useful and actionable data from such sources. In fact, applying structure to content using tweets with a micro-syntax to enhance computational automation was part of the original intent behind the TtT project. Challenges associated with filtering, managing, analyzing and translating large volumes of social media communications are being addressed through ongoing development of The SwiftRiver platform by Ushahidi.

During the first deployment of TtT for the 2010 Haiti earthquake, the syntax was not widely adopted by citizens and first responders, but the syntax was picked up by people who spontaneously volunteer during a crisis (Starbird and Palen, 2011). TtT efforts spurred a network of volunteers that helped give structure to the social media communications that were transpiring on Twitter during both the 2011 Joplin tornado and 2010 Boulder Fourmile Canyon fire crises. These volunteers adopted the TtT syntax and translated information from multiple sources using the syntax before tweeting it out to their followers. These followers were diverse, including media outlets, the American Red Cross, FEMA, and other relief organizations. This type of volunteerism was promoted to direct Twitter communications so that automatic filtering of Tweets would be more effective.

During the 2010 Haiti earthquake, Ushahidi enlisted volunteers to assist with handling the large volume of data. SMS messages began to flow at a rate of 1,000 to 2,000 a day and were passed directly from the cellular telephone provider to an automated system, designed by Ushahidi developers for coordinating volunteers. Volunteers manually translated the messages from Haitian Creole and then filtered and determined locations (Meier and Munro, 2010). The system supported message translation with a lead time of less than ten minutes.

EXTRACTING LOCATION INFORMATION

One of the most challenging aspects of using social media data during a disaster is extracting unambiguous and accurate location information. Locations are essential for determining if a message is actionable (Munro, 2011). Location can be determined in several ways, including processing location references like a place name or street

Proceedings of the 9ᵗʰ International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

5

address in message content; explicit coordinates derived from geo-location services from cellular phones; and extraction of location information in a user's social media account profile (Bellucci, et al., 2010, Field and O'Brien, 2010). Table 1 illustrates examples of profile-derived location information shown in Tweets 5-8.

Tweet	Time	User	Tweet	Location
1	11.21 pm 23 May 2011	@sarahgracesitz	RT @shawncmatthews: Here is a great resource for donation centers #joplin #tornado #relief http://ow.ly/515zC	iPhone: 37.112511,-93.303925
2	11.18 pm 23 May 2011	@CajunTechie	If you want to donate clothing of all sizes to displaced residents in #Joplin #Missouri you can drop it by 702 Moffet #tornado	ÃœT: 36.874136,-94.873582
3	10.07 pm 23 May 2011	@pbdoetmee	Ronduit dramatisch fotowerk na de tornado hit-down in Joplin, Missouri, USA... http://bit.ly/jiw6Kg #indrukwekkend #joplin #tornado	51.953923,6.008155
4	11.35 pm 23 May 2011	@PeterKinder	Just confirmed I will be guest on @IngrahamAngle radio show Tuesday morn 5/24/11 9:30 CDT talking #Joplin #MO #tornado relief, update #pdk	iPhone: 0.000000,0.000000
5	10.48 pm 23 May 2011	@maryfranholm	RT @OzarksRedCross: #CBCO FB site says: CODE RED 4 blood donations 4 the #Joplin #tornado has been lifted. Thanks 2 so many of you donat ...	Lost in a good book
6	10.39 pm 23 May 2011	@Jeannie_Hartley	RT @OzarksRedCross: #RedCross update here: http://bit.ly/jZ3Ivp #Joplin #tornado	Universe
7	10.36 pm 23 May 2011	@wheelertweets	RT @Jeannie_Hartley: #Tornado #Joplin #mo @info4disasters @Redcross @kcredcross @1stAid4 @wheelertweets @jnicky63 @viequesbound @Lady1st ...	Tunis, Tunisia
8	11.02 pm 23 May 2011	@JoplinMoTornado	GOODNEWS: 7 people were rescued from the debris today! #joplin #tornado -G	Joplin, Mo
9	11.02 pm 23 May 2011	@DAOWENS44	RT @OzarksRedCross: #CBCO FB site says: CODE RED 4 blood donations 4 the #Joplin #tornado has been lifted. Thanks 2 so many of you donat ...	

Table 1. Tweet Examples from the 2011 Joplin Tornado.

Location information included within a Twitter or SMS message as a text reference (e.g. a user mentions a specific place by name) must be extracted and geocoded to obtain coordinate information. Latitude and Longitude coordinates can also be included with messages as geospatial metadata. The process of manually or computationally assigning such metadata is called geo-tagging. Twitter included the ability to geo-tag tweets in 2009 (Bellucci, et al., 2010). Because of privacy concerns, social media applications and cellular phones usually require users to opt-in to enable geo-tagging. The SMS protocol does not incorporate geospatial metadata and typical messages sent from cellular phones via SMS will not contain location information (Munro, 2011). However, GeoSMS (geosms.wordpress.com), a location-enabled SMS standard, can embed geospatial metadata into a URI (Uniform Resource Identifier). MacEachren, et al. (2011) notes that the proportion of users who enable geo-tagging is still small. However, geo-tagging alone is no guarantee the message content is meaningful. Of three geo-tagged examples in Table 1, two geo-tags are close to Joplin in Springfield, Missouri (1) and the nearby city of Miami, Oklahoma (2), while the location in Tweet 3 is in the Netherlands.

The Ushahidi platform does not contain a mechanism to automatically geocode implicit location information, but the SwiftRiver Platform does incorporate tools that use natural language processing and a gazetteer to return coordinate locations based on place names. The Ushahidi platform will extract geospatial metadata from social media feeds if it exists. For the Ushahidi 2010 Haiti earthquake map, the majority of information gathered came via SMS and not geo-tagged. Location information from SMS messages was translated by volunteers who used a variety of resources to obtain coordinate locations from the translated messages. Many of the volunteers were originally from Haiti and used their own geographical knowledge of the region combined with Open Street Map to pinpoint extract coordinates (Heinzelman and Waters, 2010). For the Ushahidi 2010 Gulf Oil Spill map we were unable to determine which specific methods were used to geo-locate implicit location information.

The software platform used by TtT extracts geospatial metadata using the Twitter API if such metadata exists. The software filters for location tags prescribed in the TtT micro-syntax or tags identifiable as spontaneously generated by the crowd that may include implicit location information, for example #loc or #lat and #long. These tags and the data after each tag were parsed into key-value pairs to populate the database. Location pairs, along with identified place names or event tags like Joplin or Boulder were geocoded using GeoKit (geokit.rubyforge.com), which can geocode textual information across a number of different geocoding services. Volunteers could review the resulting coordinate pairs, which were then entered into the database if approved.

Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

6

COMPARING USHAHIDI AND TWEAK THE TWEET

Here we draw comparisons between Ushahidi and TtT in three key dimensions. First, we describe what types of data and variables are captured by each effort. Next, we compare the interactive mapping tools that each platform provides. We conclude our comparisons by characterizing how each platform has resulted in tangible actions by responders and emergency managers. We use four recent disasters in these comparisons. For Ushahidi, we explore its use in the 2010 Haiti earthquake and 2010 Gulf Oil Spill. For TtT we focus on the 2011 Joplin tornado and 2010 Boulder Fourmile Canyon fire. We were unable to find directly overlapping events for both platforms. Ushahidi deployments tend to focus on larger disasters rather than the localized events focused on by TtT. For one overlapping event, the 2010 Haiti earthquake, TtT had not yet implemented mapping tools, and in other overlapping events (like 2010 Pakistan floods) TtT has integrated their efforts with Ushahidi.

Field type	Ushahidi Field Name	TtT Field Name Boulder[1] Joplin[2]	Definition - TtT	Definition - Ushahidi
Record ID	#	Record ID[1] / ID[2]	Unique identifier	Unique identifier
Event		Event[1]	Event Hashtag – used for Place location	
Categorization	Category	Report Type[1,2]	Primary Hashtag – only one allowed – used for Key legend in Web Maps	Multiple categories allowed – used for Web Map Category Filter in Legend
Report	Incident Title	Report[2]	Partial parsed tweet with hashtags removed for pop-up display	Report Title for Web Report
Details		Details[1]	Partial parsed tweet with hashtags removed for pop-up display	
Original Report	Description	Text[2]	Original tweet	Original Message (in original language and translated if necessary)
Date/Time Stamp	Incident Date	Time[1,2]	Tweet time stamp	Message time stamp
Date_Time		Date_Time[2]	Time contained in tweet message	
Info		Info[1]	Volunteer added comment	
Source		Source[1,2]	Twitter user	
Contact		Contact[1,2]	Name, number, web page or other contact info contained in tweet	
Completed		Complete[1,2]	Indication if report was acted upon	
Status		Status[1]	? All N/A	
Verification	Verified	Verified[1]	? All N/A	Corroborated via incident report credibility vote
Actionable		Actionable[1]	? All N/A	
Approved	Approved			Map location approved
Author		Tweet Author[1] / Author[2]	The author of the record in the spreadsheet or author of retweet	
Tweet		Tweet[1]	Original Tweet	
Photo URL		Photo URL[1], Photo[2]	URL to photo	
Video		Video[2]	URL to Video	
Location (Text)	Location	Location[1,2]	Parsed location string	Parsed location string
Mapped		Mapped[1]	? All N/A	
Longitude	Longitude	GPS Long[1,2]	Derived Longitude	Derived Longitude
Latitude	Latitude	GPS Lat[1,2]	Derived Latitude	Derived Latitude

Table 2. Comparing data collected from TtT and Ushahidi

Raw Data

Data generated during a disaster from social media networks tend to be ephemeral and if it is not collected during the disaster, it can be difficult to conduct related research after the fact. Collecting raw data from Twitter older than two weeks has become challenging due to changes in the Twitter API that forbid certain types of archiving. Here, we conduct our analysis using the spreadsheets gathered from each application and additional analytical results from the PeopleBrowsr (www.peoplebrowsr.com) service which provides 1000 days of social media content and social analytics for marketers (not including SMS). We did not include PeopleBrowsr analytics for the 2010 Haiti earthquake because that data collected was primarily from SMS, and PeopleBrowsr analytics are not available for the 2010 Boulder Fourmile Canyon fire due to a small number of reports.

In Table 2 we list all fields we discovered in the TtT and Ushahidi spreadsheets and our interpretation of the definitions for each field type for each application. Common fields which we think share a common meaning across both platforms are highlighted. The Ushahidi platform has fewer fields (eight vs. twenty-five for TtT) and

Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

7

they do not vary between the two incidents. TtT has variation in field names and the number of fields. We note that it is difficult to differentiate between the terms Status, Actionable and Verified in the TtT fields.

In the content summary shown in Table 3 the Ushahidi field "Approved" always shows a rating of 100%. According to Ushahidi documentation all messages are "approved" once valid location coordinates are determined and an administrator approves the content. Reports that are not yet approved are not displayed. The "verified" field indicates a report is submitted by or corroborated by a trusted source or an administrator. Table 3 shows that 6% and 40% of the records were corroborated in the 2010 Haiti earthquake and 2010 Gulf Oil Spill events. Raw data from TtT did not reveal the meaning of the codes "Status", "Actionable", and "Verified."

The 2010 Gulf Oil Spill Ushahidi spreadsheet lists the first incident date eleven days after the Deepwater Horizon explosion. The total number of tweets with the #oilspill keyword from April 10th to October 18th, 2010 according to PeopleBrowsr, is 22,199. The Louisiana Bucket Brigade collected 2952 reports according to their spreadsheet, representing approximately 13% of the total Twitter traffic by the PeopleBrowsr estimate. Of note is that all 2952 reports were geo-located. Additionally, there were only 9 tweets on the day after the explosion and no tweets for the next 17 days. The traffic over six months highlights the extended nature of the disaster.

The 2011 Joplin tornado data starts the day after the tornado and ends 27 days after the tornado. According to PeopleBrowsr there were 333,387 total Twitter mentions of the #Joplin keyword from May 13 to June 13th, 2011. TtT identified 504 tweets that were entered into the spreadsheet. This represents approximately 0.02% of the total Twitter traffic if the PeopleBrowsr estimates are correct. This highlights the challenge associated with harvesting social media communications during temporally-limited crises. It is also interesting to note that 65% of the 504 records in the TtT spreadsheet for the 2011 Joplin tornado and 54% of the 522 records for the 2010 Boulder Fourmile Canyon fire included locations. Examination of the raw data reveals frequent status communication between volunteers that was not mapped because it was not relevant to the event itself.

	Incident	Incident Date	Reports	First Report Date	Last Report Date	%Verified	%Approved	%Actionable	%Complete	%LAT/LONG
TtT	Joplin Tornado[2]	5/22/2011 5:34 PM	504	5/23/2011 12:11 AM	6/13/2011 11:10 PM	N/A[1]		N/A[2]	0.4%[2]	65%
	Boulder Fourmile Fire[1]	9/6/2010 10:00 AM	522	9/8/2010 5:50 PM	9/17/2010 9:33 PM			N/A[1]	N/A[1]	54%
Ushahidi	Haiti Earthquake	1/12/2010 4:53 PM	3589	1/12/2010 4:08 AM	5/18/2010 4:26 PM	6%	100%			100%
	Gulf Oil Spill	4/10/2010 10:00 PM	2952	4/21/2010 1:44 PM	10/18/2010 10:07 PM	40%	100%			100%

Table 3. Summary of Ushahidi and TtT Spreadsheet Content

Maps

Cartographic representation of crisis mapping represents another challenge in the use of social media communications for disaster response because of the need to display large volumes of data while avoiding information overload. This is complicated further by the fact that potential users of crisis maps, including citizens, responders, volunteers, journalists and managers will have different expectations influenced by their social and physical relation to the crisis event (Liu and Palen, 2010). Field and O'Brien (2010) recognize that given the growth of social media communications and the geospatial component integral to an interconnected world, good cartography is crucial for creating maps with a purpose that are more than one-dimensional.

Crisis maps created by Ushahidi and TtT are quite similar (Figure 1) in terms of their core features. Both platforms utilize simple interactive map mashups and categorized point symbols to represent reports. Ushahidi has the ability to generalize dense sets of reports into aggregated symbols, making it scalable to larger datasets. Both platforms have recently introduced temporal displays to highlight report frequency over time (frequency graph in Ushahidi and time-categorized markers in TtT). The overall map and interface aesthetic is significantly more refined in current implementations of Ushahidi, perhaps reflecting its relative maturity compared to TtT.

Neither platform supports significant geospatial analysis capabilities. Basic filtering controls are available to winnow the dataset, but there are no quantitative spatial analysis methods available to identify clusters or to compare current patterns to past patterns. A significant difference between platforms is that Ushahidi provides alerting tools for users to "listen" for reports from a given area or matching a given set of thematic criteria.

Spatial data interoperability in both platforms is supported through spreadsheet downloads of raw data, making it possible for users to ingest collected information into a full-featured GIS if necessary.

Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

8
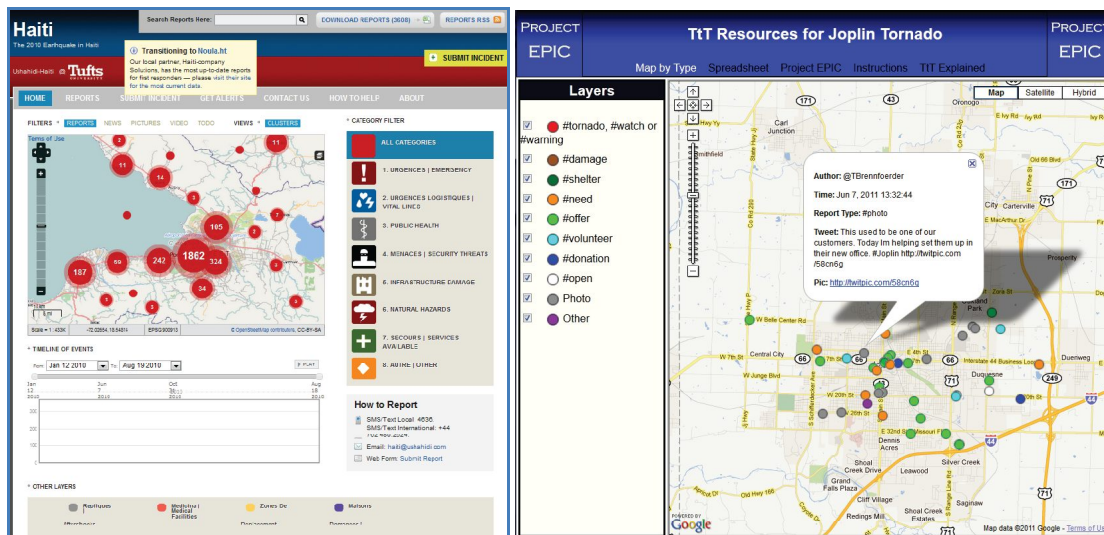
Figure 1. Examples of Ushahidi (left) and Tweak the Tweet (right) Mapping Interfaces

Evidence of Action

Understanding the effectiveness of efforts like TtT and Ushahidi is a difficult task. We concentrated our efforts on identifying impacts from media reports, after action reports, and TtT and Ushahidi's own assessments.

Project Epic and TtT gained recognition for their efforts during the 2010 Boulder Fourmile Canyon fire when CNN ran a story which credited TtT for integrating crisis information through the use of volunteers and social media (Spellman, 2010). One article cites Project Epic's use of a map with geo-located tweets to tracking the fire movements from citizen reports (Orlando, 2011). Another news report (Petty, 2010) describes how TtT was used to gather and map data from the 2010 Boulder Fourmile Canyon fire, including information not provided through official emergency response channels. We did not find many media reports of the specific use of TtT after the 2011 Joplin tornado, but the use of the syntax was promoted by organizations like Crisis Commons.

Following the 2010 Haiti earthquake, Craig Fugate, Director of FEMA, tweeted that the Ushahidi Haiti Map was the most "comprehensive and up-to-date map available to humanitarian organizations" (Heinzelman and Waters, 2010). Newsweek's profile of the Ushahidi efforts after Haiti indicate that the crisis map resulted in saving lives (Ramirez, 2010). One after action report (UN-SPIDER, 2010) notes that the US Marines used Ushahidi to coordinate locations and direct relief efforts. They also indicate that data from Ushahidi was used to direct Coast Guard responders for search and rescue. The Ushahidi Blog (blog.ushahidi.com) highlights multiple examples of action items and response efforts generated by Ushahidi, including food and water deliveries.

The 2010 Gulf Oil Spill Ushahidi map represented a different use of this technology in disaster response. The oil spill was not as much a direct threat to lives as it was a threat to local economies and the environment. Use of the map has been primarily to raise awareness of the ongoing ecological disaster and to document the damage. One article highlighted the fact that the information collected would be useful in long-term recovery efforts and could also be used in future legal actions over long-term damage to ecosystems and livelihoods (Sutter, 2010).

A common thread throughout the reports we reviewed is that the impact of geospatial social media platforms on tangible emergency response actions is not yet well-defined. While both have received media attention and have clearly captured public interest, there are few specific examples of the information leading to different decision-making patterns, widespread allocation of resources, or information leading to the rescue of disaster victims.

Future Design Considerations and Conclusions

The recent use of social media communications in disaster response is largely driven by volunteer organizations. Engagement with volunteers during the 2010 Haiti earthquake by TtT prompted TtT developers to focus research on a second layer of crowdsourcing: communication between volunteers and response organizations (Starbird, 2011). By working with crisis volunteer organizations TtT has continued to promote the use of the micro-syntax after a disaster. Spurred by lessons learned from deployments of crisis mapping efforts like Ushahidi, the Standby Task Force was recently formed to organize volunteers and provide a dedicated technical interface to the humanitarian community to assist in dealing with new sources of information like social media.

Proceedings of the 9ᵗʰ International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

9

A key design consideration going forward is to ensure effective mechanisms for disseminating and sharing information between first responders and crisis managers. A UN report suggests that efforts like Ushahidi and TtT may contribute to information overload, citing that a large percentage of the information gathered was already in the hands of relief organizations on the ground (UN-SPIDER, 2010). This does not take into consideration the value of mapping the information and providing crisis managers a mechanism to identify clusters where relief organizations and first responders could concentrate their efforts.

We have outlined several reasons why extracting the location from social media is difficult, and further research is needed in this area, especially in the role of automatic geocoding and disambiguation of text descriptions of place. Ushahidi point out that from 40,000 Haiti-related SMS messages only 5% were mapped (Norheim-Hagtun and Meier, 2010), but that does not mean that the other 95% did not contain any geospatial information. To support analytical reasoning and geospatial analysis we should be able to uncover patterns that reference physical and cultural regions, types of landforms, directional information and topological relations in addition to basic point locations. A range of recent research focuses on the challenges of extracting actionable information from social media. Munro (2011) proposes models to systematically identify actionable items using trending categories or topics, subwords, and spatiotemporal clusters. MacEachren, et al. (2011) discuss how SensePlace2, a geovisual analytics application, includes a crawler application designed to systematically query the Twitter API based on crisis-relevant keywords and phrases. Vieweg, et al. (2010) have recently described an automated methodology to detect messages that will be useful for situational awareness.

Credibility and verification of information is another area that needs to be addressed in future research. One report indicates that search teams found a high proportion of SMS reports about trapped wounded victims turned out to be coming from families wanting to recover their dead relatives (Harvard Humanitarian Initiative, 2011). Some recent research has focused on identifying ways in which credibility might be automatically assessed in Tweets by evaluating message content, user profile details, and message propagation (Castillo, et al., 2011).

Finally, we must develop effective cartographic representation techniques to ensure the usability of web maps for crises. A particular challenge for crisis mapping is that there are a wide range of expectations and technical skills associated the diverse group of people that need to use crisis maps, including citizens, responders, volunteers, journalists and managers. Not all groups are equally equipped to evaluate the results of geographical analysis. Maps are likely to be seen as credible evidence, even when the underlying data is of unknown quality.

There is no doubt that the contribution of social media communications during disaster has shifted the paradigm of emergency response to include at least a one-way social media dialog from those most affected. Mapping social media content provides a way to gather and visualize information from what can arguably considered the true first responders - the affected citizens who are the first to assess the situation and request assistance through social media. Driven by volunteers and advances in web-based technology, the proliferation of this information has grown faster than the analytical capabilities of disaster management organizations and workflows. TtT has contributed a method to filter, automate and direct information from social media sources during a disaster and Ushahidi has proven to be an effective and widely adoptable platform for displaying geospatially-oriented social media communications. However, TtT and Ushahidi have only tackled simple location-related problems and provided only rudimentary situational awareness and mapping capabilities to visualize the social media communication stream. Future research must focus on applications that go beyond basic crowdsourcing to develop information collections, analytical tools, coordination of communications, and mapping visualization to support all phases of disaster management. Future platforms developed with the volunteer community in mind will need to incorporate social media as one piece of an overall strategy to support situational awareness and response and recovery featuring effective two-way communications with citizens through social media.

ACKNOWLEDGEMENTS

REFERENCES

1. S. Vieweg, A.L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events: what twitter may contribute to situational awareness, in, ACM, Atlanta, Georgia, USA, 2010, pp. 1079-1088.

2. S.B. Liu, L. Palen, The new cartographers: crisis map mashups and the emergence of neogeographic practice, Cartography and Geographic Information Science, 37 (2010) 69-90.

Proceedings of the 9ᵗʰ International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

10

3. A.M. MacEachren, A. Jaiswal, A.C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, J. Blanford, SensePlace2: Geotwitter Analytics Support for Situation Awareness, in: IEEE Conference on Visual Analytics Science and Technology, Providence, RI, 2011.

4. American Red Cross, Social Media in Disasters, in: http://www.redcross.org/www-files/Documents/pdf/SocialMediainDisasters.pdf, accessed on 10/15/2011

5. International Telecommunication Union, Key Global Telecom Indicators for the World Telecommunication Service Sector, in: www.itu.int/ITU-D/ict/statistics/at_glance/KeyTelecom.html, accessed on 10/15/2011

6. K.M. Anderson, A. Schram, Design and implementation of a data analytics infrastructure in support of crisis informatics research (NIER track), in, ACM, Waikiki, Honolulu, HI, USA, 2011, pp. 844-847.

7. Harvard Humanitarian Initiative, Disaster Relief 2.0: The future of information sharing in humanitarian emergencies, in, 2011.

8. K. Starbird, Digital Volunteerism During Disaster: Crowdsourcing Information Processing, in: CHI '11 Workshop on Crowdsourcing and Human Computation, Vancouver, BC, 2011.

9. K. Starbird, L. Palen, "Voluntweeters": self-organizing by digital volunteers in times of crisis, in, ACM, Vancouver, BC, Canada, 2011, pp. 1071-1080.

10. O. Okolloh, Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information, Participatory Learning and Action, 59 (2009) 65-70.

11. P. Currion, C. de Silva, B. Van De Walle, Open source software for disaster management, Communications of the ACM, 50 (2007) 61-65.

12. M. Zappavigna, Ambient affiliation: A linguistic perspective on Twitter, New Media & Society, 13 (2011) 788-806.

13. L. Palen, K.M. Anderson, G. Mark, J. Martin, D. Sicker, M. Palmer, D. Grunwald, A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters, in: ACM-BCS Visions of Computer Science Conference, Edinburgh, UK, 2010, pp. 1-12.

14. J. Heinzelman, C. Waters, Crowdsourcing Crisis Information in Disaster-Affected Haiti, United States Institute of Peace, 252 (2010) 1-16.

15. S. Dosemagen, Ushahidi Used to Create Oil Spill Crisis Map, in: The Ushahidi Blog, 2010.

16. Ushahidi, Ushahidi API, in: http://wiki.ushahidi.com/doku.php?id=ushahidi_api, accessed on 10/15/2011

17. P. Meier, R. Munro, The Unprecedented Role of SMS in Disaster Response: Learning from Haiti, SAIS Review, 30 (2010) 91-103.

18. R. Munro, Subword and spatiotemporal models for identifying actionable information in Haitian Kreyol, in, Association for Computational Linguistics, Portland, Oregon, 2011, pp. 68-77.

19. A. Bellucci, A. Malizia, P. Diaz, I. Aedo, Framing the design space for novel crisis-related mashups: the eStoryS example, in: Information Systems for Crisis Management and Response (ISCRAM 2010), Seattle, WA, 2010, pp. 1-10.

20. K. Field, J. O'Brien, Cartoblography: experiments in using and organising the spatial context of micro-blogging, Transactions in GIS, 14 (2010) 5-23.

21. J. Spellman, Heading off disaster, one tweet at a time, in: CNN Tech, www.cnn.com/2010/TECH/social.media/09/22/natural.disasters.social.media/index.html,accessed on 10/15/2011

22. J. Orlando, Turning Disaster Response On Its Head, in: Continuity Insights, http://www.continuityinsights.com/articles/turning-disaster-response-on-its-head, accessed on 10/15/2011

23. D. Petty, Evacuees use social media to keep up on Boulder wildfire disaster developments, in: Denver Post, MediaNews, Denver, CO, 2010.

24. J. Ramirez, 'Ushahidi' Technology Saves Lives in Haiti and Chile, in: Newsweek, 2010.

25. UN-SPIDER, Lessons from Haiti, Coordinates, 6 (2010) 27-31.

26. J.D. Sutter, Citizens monitor Gulf Coast after oil spill, in: CNN Tech, articles.cnn.com/2010-05-06/tech/crowdsource.gulf.oil_1_oil-spill-gulf-coast-jeffrey-warren, accessed on 10/15/2011

27. I. Norheim-Hagtun, P. Meier, Crowdsourcing for Crisis Mapping in Haiti, innovations, 5 (2010) 81-89.

28. C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in, ACM, Hyderabad, India, 2011, pp. 675-684.

Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012
L. Rothkrantz, J. Ristvej and Z. Franco, eds.

11

Spatiotemporal crime analysis in U.S. law enforcement agencies: Current practices and unmet needs

Robert E. Roth [a,*], Kevin S. Ross [c], Benjamin G. Finch [b], Wei Luo [b], Alan M. MacEachren [b]

[a] University of Wisconsin-Madison, Geography, 550 N. Park Street, Madison, WI 53706, United States
[b] GeoVISTA Center, Department of Geography, The Pennsylvania University
[c] Nsite, LLC

ARTICLE INFO

ABSTRACT

This article compares the current states of science and practice regarding spatiotemporal (space + time) crime analysis within intermediate- to large-size law enforcement agencies in the Northeastern United States. The contributions of the presented research are two-fold. First, a comprehensive literature review was completed spanning the domains of Criminology/Crime Analysis and GIScience/Cartography to establish the current state of *science* on spatiotemporal crime analysis. This background review then was complemented with a set of interviews with personnel from seven intermediate- to large-size law enforcement agencies in the United States in order to establish the current state of *practice* of spatiotemporal crime analysis. The comparison of science and practice revealed a variety of insights into the current practice of spatiotemporal crime analysis as well as identified four broad, currently unmet needs: (1) improve access to externally maintained government datasets and allow for flexible and dynamic combination of these datasets; (2) place an emphasis on user interface design in order to improve the usability of crime mapping and analysis tools, (3) integrate geographic and temporal representations and analyses methods to better unlock insight into spatiotemporal criminal activity, and (4) improve support for strategic crime analysis and, ultimately, public safety policymaking and administration. The results of the interview study ultimately were used to inform the design and development of a spatiotemporal crime mapping application called *GeoVISTA CrimeViz*.

Published by Elsevier Inc.

1. Introduction: the analysis of information on criminal activity

Crime analysis describes the systematic collection, preparation, interpretation, and dissemination of information about criminal activity to support the mission of law enforcement (Boba, 2005). The goal of crime analysis is the unlocking of valuable insights from the collected crime information in order to assist law enforcement with criminal apprehension and crime prevention, to the end of improving the overall quality of life for community residents (O'Shea & Nicholls, 2003). Ideally, crime analysis draws upon both quantitative and qualitative approaches in order to understand criminal activity fully, integrating descriptive and inferential statistical analyses of crime incidents with text reports, information graphics, and prior experience to determine the appropriate response tactics, strategies, and broader policies (Gottlieb, Arenberg, & Singh, 1994; Osborne & Wernicke, 2003). Influenced by the Digital Revolution and associated

Information Age, research and development within crime analysis during the past two decades has emphasized the design of computer software that supports the assembly and interpretation of digitally-native crime information (Wilson, 2007). The research reported here focuses upon a critical subset of computing technologies designed to analyze the spatial and temporal (together *spatiotemporal*) components of crime information.

The field of *Geographic Information Science* (GIScience) and its technological counterpart *Geographic Information Systems* (GIS) describe the gamut of tools and techniques available to analyze geographically-referenced information (Goodchild, 1992). GIScience subsumes a variety of topics relevant to spatiotemporal crime analysis, which include geographic information collection (geocoding, GPS technology, remote sensing, and surveying), geographic information maintenance (geographic database management and multi-resolution databases), geographic information analysis (geocomputation, geographic data modeling, spatial analysis, and spatial statistics), geographic information representation (cartography and geographic visualization) and the use of geographic information and information products (geocollaboration, geovisual analytics, public participatory GIS, and spatial decision support systems) (for a general overview of these topics, see Longley, Goodchild, Maguire, & Rhind, 2005). The term *crime mapping* is used today to describe the

* Corresponding author.
E-mail addresses: reroth@wisc.edu (R.E. Roth), kevin@kross.com (K.S. Ross), bgf111@psu.edu (B.G. Finch), wul132@psu.edu (W. Luo), maceachren@psu.edu (A.M. MacEachren).

application of all GIScience tools and techniques for crime analysis (Getis et al., 2000), although its original use focused on applications of Cartography only (i.e., the representation of geospatial crime information in map form).

There is a substantial volume of work within GIScience examining the treatment of spatial and temporal components of information in conjunction (e.g., Andrienko, Andrienko, & Gatalsky, 2003; Hägerstrand, 1970; Langran, 1992; Peuquet, 1994; Sinton, 1978). Despite this research, there is little implementation of temporal analytical functionality in popular GIS software. Perhaps as a direct result, the analysis of the temporal component of crime has been identified as an under-supported function of crime analysis, with Ratcliffe (2009: 12) stating in an overview of current challenges to crime analysis that "At present, the most under-researched area of spatial criminology is that of spatio-temporal crime patterns." Existing reports on crime analysis indicate that spatiotemporal analysis and visualization often is limited in practice to the generation of one-off, static maps showing crime over a small period of time, usually the past 7-to-30 days (Lodha & Verma, 1999). Thus, the possible use cases for advanced spatiotemporal crime analysis remain undetermined and therefore the positive impacts of spatiotemporal crime analysis remain unrealized.

Here, we describe research to address directly this challenge of spatiotemporal crime analysis. The aim of our research was the identification of gaps between the spatiotemporal crime analysis techniques reported in the literature and the actual use of these techniques by law enforcement to combat crime. The primary contributions of the research are two-fold. We first completed a comprehensive background review to understand the current state of *science* in spatiotemporal crime analysis, disambiguating and synthesizing relevant research from the knowledge domains of Criminology/Crime Analysis and GIScience/Cartography. We then conducted a set of interviews with experts from seven intermediate- to large-size law enforcement agencies in the United States (daytime service populations of 125,000 to many millions) in order to compare the current state of *practice* in spatiotemporal crime analysis to the previously reviewed state of science. Such a critical comparison of science and practice is relevant to detectives, officers, and decision makers working in law enforcement as well as municipal, state, and federal administrators and policymakers working broadly in public safety. The interview study also served as the needs assessment stage for the design of a spatiotemporal crime mapping application called *GeoVISTA CrimeViz* (http://www.geovista.psu.edu/CrimeViz) developed in collaboration between the Penn State GeoVISTA Center and the Harrisburg (PA, USA) Bureau of Police (for details on the application, see Roth, 2011; Roth & Ross, 2009; Roth, Ross, Finch, Luo, & MacEachren, 2010). Therefore, we were interested in identifying the key crime analysis needs of law enforcement agencies that the *GeoVISTA CrimeViz* application must support, with a particular emphasis on those needs not currently supported by readily available spatiotemporal crime analysis software.

The article proceeds in four sections. In the following section, we synthesize background material from the domains of Criminology/Crime Analysis and GIScience/Cartography to establish the current state of science on spatiotemporal crime analysis. In the third section, our interview protocol and qualitative data analysis approach is described. We present the results and discuss the key findings of the interviews in the fourth section, providing an overview of the current state of practice to contrast with the background review. The fourth section is organized according to six key crime analysis needs identified from the background review: (1) geographic information, (2) cartographic representation, (3) cartographic interaction, (4) spatial analysis, (5) temporal analysis, and (6) map and analysis use. The fifth and final section contains our concluding remarks and lists several broad spatiotemporal crime analysis needs that currently are not fully support.

2. Background review: current state of science on crime analysis

A comprehensive review of existing literature was completed prior to the interview study in order to characterize the current state of science on crime analysis. The following review is organized into three sections: (1) a summary of the origins and purpose of crime analysis from the discipline of Criminology, with an emphasis on the types of crime analysis; (2) a summary of the different kinds of geographic information that may be collected to support crime analysis and the ways to represent this information cartographically (i.e., in map form); and (3) advanced statistical and computation techniques to analyze the spatial and temporal components of these information.

2.1. Origins and purpose of crime analysis

Crime analysis has its roots in 19th century London, where the first modern police department was established (Boba, 2005). August Vollmer, Police Chief of Berkeley (CA, USA) and founding professor of the UC-Berkeley School of Criminology, often is credited with the first application of crime analysis in the United States in the early 20th century, with other important early U.S. work conducted by the 'Chicago School' of sociologists (e.g., Shaw & McKay, 1942; Sutherland, 1934). Vollmer's student, O.W. Wilson, first defined the term 'crime analysis' in his recommendation of information analysis techniques to police departments in the 1950s and 1960s (Wilson & McLaren, 1977). The crime analysis capabilities of law enforcement agencies expanded through the 1970s and 1980s (Emig, Heck, & Kravitz, 1980), due in part to federal grants provided through the National Institute of Justice, a program of the United States Department of Justice. There also was increased interest at this time in crime analysis in academia; a review of this research is provided in Harries (1999).

Crime analysis therefore is informed by the discipline of *Criminology*, or the scientific study of the causes and control of crime and delinquent behavior, with the goal of understanding criminal activity, rehabilitating convicted criminals, and improving the quality of life within a community (Sutherland, Cressey, & Luckenbill, 1992). There are two popular criminological theories that emphasize the importance of spatiotemporal pattern and process (Cahill & Mulligan, 2007). Under *routine activity theory*, an individual criminal incident requires three conditions to occur concurrently in place: (1) presence of a motivated offender, (2) presence of a suitable target, and (3) absence of a proper guardian, law enforcement or otherwise (Cohen & Felson, 1979). The spatiotemporal dynamics of these three components can be analyzed both to identify locations of elevated crime risk and to prescribe the appropriate policing tactics to attenuate this crime risk (Bruce, 2008). In contrast, *social-disorganization theory* evaluates the ability of a community, or homogenous geographic unit, to combat negative community-level changes and enforce positive ones (Shaw & McKay, 1942). By analyzing the spatial and temporal differences in demographic and environment characteristics between stable and disrupted neighborhoods, long-term policing strategies can be developed and absent public policies can be established to prevent criminal activity in blighted communities (Sampson & Groves, 1989). Together, these two theories reveal the importance of spatial and temporal context during crime analysis (Wilcox, Land, & Hunt, 2003).

Boba (2005) describes five types of crime analyses, or the general applications of criminological theory and crime analysis techniques in support of the functions of law enforcement:

(1) *Criminal investigative analysis* describes the process of collecting and analyzing information about a criminal offender. Criminal investigative analysis often involves the construction of offender profiles from known information, which then allows for the inference of offender characteristics (e.g., personality type, social habits, and work habits) based on those profiles (Jackson & Bekerian, 1997); journey-to-crime analysis, described below, is

a spatial analysis technique that can be applied to build the geographic component of an offender profile.

(2) *Intelligence analysis* expands investigation of a single individual and single crime series to a larger crime syndicate, focusing upon identification of relationships among offenders, called *link analysis*. Intelligence analysis often is applied in the context of organized crime. By establishing the offender network, law enforcement can identify and target key players in the jurisdiction and diffuse crime from the top down (Innes, Fielding, & Cope, 2005). White and Roth (2010) describe the potential of harvesting the geographic information from microblogging and social networking services to build a spatially-anchored offender network for informing and structuring intelligence analysis, although noting potential ethical concerns of harvesting volunteered geographic information.

(3) *Tactical crime analysis* is the reactive investigation of recent crime spikes within a single jurisdiction or across multiple jurisdictions (Bruce, 2008). Tactical crime analysis examines key aspects of recent criminal activity (e.g., crime type, location, time, MO, suspect description) to identify overarching patterns that may explain the recent spike. Such analysis directly informs apprehension, suppression, and target hardening blue force tactics (Bruce & Ouellette, 2008). The application of tactical crime analysis is central to the *CompStat process*, where police captains are required to present statistical analyses and cartographic representations of recent crime in their jurisdiction during regularly scheduled meetings as a way to improve leadership accountability for recent crime spikes (Walsh, 2001); CompStat has the potential for application as a strong strategic tool as well (Weisburd, Mastrofski, McNally, & Greenspan, 2002).

(4) *Strategic crime analysis* is the analysis of crime and other police-related issues to identify long-term plans for reducing crime rates and improving the quality of life for a community. Strategic crime analysis embodies the concept of *problem-oriented policing* (Goldstein, 1979), which proactively seeks to understand the underlying causes of persistent criminal activity and to develop intervention strategies to attenuate this activity. There also is an important evaluation component of strategic crime analysis that determines how well previously applied intervention strategies worked to combat crime (Boba, 2001), the results of which may inform broader public policies. Such evaluation is the final step of the strategic crime analysis model recommended by Eck and Spelman (1987) called *SARA*: *S*canning, *A*nalysis, *R*esponses, and *A*ssessment.

(5) *Administrative crime analysis* presents interesting findings of crime research and analysis to audiences within police administration, city government officials, and citizens. Administrative crime analysis directly links the detectives, officers, and decision makers responding to criminal activity and the municipal, state, and federal administrators and policymakers responsible for broader issues in public safety. Such administrative activity includes the allocation of resources within the department, such as assigning cases to detectives, and any other internal collaboration (Zhao et al., 2006). This also includes the preparation of crime reports and graphics for use in court proceedings (Harries, 1999). Finally, administrative crime analysis includes the presentation of criminal activity for public consumption, through town hall meetings or websites, to the end of promoting dialogue about public policy (Rose, 2008).

2.2. Geographic information and cartographic representation in crime analysis

As introduced above, *crime mapping* describes the analysis of the geographic component of criminal activity, both at an individual level of analysis (e.g., investigative analysis of a single crime series) and

ecological level of analysis (e.g., comparative analysis across neighborhoods to identify communities with unusually high concentrations of crime) (Eck, Chainey, Cameron, Leitner, & Wilson, 2005). Law enforcement agencies are required to collect and maintain several different information sets to document criminal activity and to support crime analysis. Harries (1999) identifies three geographically-referenced information sets commonly maintained internally by municipal law enforcement agencies: (1) crime reports, (2) calls for service, and (3) vehicle recoveries. The *crime report* is the primary information set used by law enforcement agencies and includes both numerical and categorical information for indexing and searching of the record as well as a lengthy, textual narrative of the event compiled by the reporting officer. Many records management systems distinguish between *crime incidents*—which focus on attributes of the crime event such as location, time of day, and characteristics of the victim—and *arrests*—which focus on characteristics of the apprehended offender (Mamalian & La Vigne, 1999). Crime reports are organized according to *uniform crime reporting* (UCR) codes for comparison across municipalities and states. Although there is some variation in the exact coding scheme used across municipalities and states, the UCR code commonly includes a two digit *UCR primary* code indicating crime type and a two digit *UCR secondary* code indicating a discriminating condition within the primary crime type. Many municipalities also use the UCR system for indexing the *modus operandi* (MO), or method of committing the crime.

The *calls for service* information set indexes all requests for law enforcement services, typically submitted by phone, and is an order of magnitude larger than the crime incident information set, as most police dispatch does not lead directly to a reported incident or an arrest. Maps of calls for service are interpreted by crime analysts as the general 'demand' for police services within the municipality (Spelman, 1995). The *vehicle recoveries* information set, maintained primarily in larger municipalities, indexes the locations from which vehicles were reported as stolen and subsequently recovered (Chainey, Tompson, & Uhlig, 2008). A fourth information set maintained internally by some law enforcement agencies is the *field interview*, or information collected by officers from potential witnesses and offenders while on patrol (Osborne & Wernicke, 2003). Finally, Harries (1999) notes that agencies often utilize external information sources, which may include federal information like Census Bureau information, national crime information like the probation and missing persons lists as well as the sex offender registry, and volunteered information from microblogging and social networking services.

As the name implies, a principle task within crime mapping is the production of *cartographic representations* (i.e., maps) of the aforementioned kinds of geographic information collected on criminal activity. Literature on crime mapping uses alternative terminology from that common in GIScience (specifically within Cartography, e.g., Dent, 1999; Slocum et al., 2005) to describe the reference and thematic maps produced in support of crime analysis; translations between lexicons are provided below. Boba (2005) describes six types of crime maps generated to support the mission of law enforcement:

(1) *Single-symbol maps* use point symbols to represent the locations of features. In crime mapping, these are commonly referred to as *push pin maps*, drawing on the analog wall map solution used prior to the move to electronic information and GIS (for more on the etiology and evolution of push pin maps, see Wallace, 2011). In GIScience, this kind of map is referred to as a *one-to-one dot map*; when the symbology varies by color, shape, or central icon to represent a nominal difference in kind, the map sometimes is described as containing qualitative point symbols (Roth, 2010). One-to-many dot maps (i.e., *dot density maps*), where one dot represents multiple crimes, are not common in crime mapping, perhaps because of the potential misinterpretation of the meaning of a dot and the associated underestimation of total crime.

(2) *Graduated maps* are described by Boba (2005) as the use of either color or size to represent aggregated information. In GIScience, the use of a color gradient to represent aggregated information is called a *choropleth map*, while the use of size to represent aggregated information is called a *proportional symbol map* (MacEachren & DiBiase, 1991). For all choropleth maps and some proportional symbol maps, the information typically is aggregated to a set of relevant boundaries (i.e., enumeration units), such as police districts or beats. In the case of proportional symbol maps, the information also might be aggregated according to a set of point locations (e.g., apartment complexes, arenas, bars, stores) or linear features (e.g., street blocks).

(3) *Density maps* aggregate crime incidents to an arbitrary grid either directly or using a moving window smoothing function, with the frequency of each grid cell represented by color; these maps are referred to as *hot spot maps* in practice, although there is conflicting use of this term in the crime mapping literature (Chainey et al., 2008). In GIScience, this technique typically is called *isoline mapping* or *surface mapping* (Slocum, McMaster, Kessler, & Howard, 2005), although the actual isolines (i.e., lines of equal crime frequency) are rarely depicted on crime maps, with the underlying interpolation grid instead color tinted. Hot spot maps have the advantage over choropleth or proportional symbol maps in that they are not restricted by political units that have little impact on criminal activity, but they suffer more heavily from the *denominator dilemma*, as the underlying population typically is not known for arbitrary grid cells (i.e., the hot spots only may be indicating where the people are and not where criminal activity is elevated above average) (Ratcliffe, 2009).

(4) *Chart maps* show relative values within a single variable at the same time, such as the percentage of crime types by district. Examples include pie charts and stacked histograms that are placed directly on the map (Andrienko & Andrienko, 1999). The concept of a chart map can be extended to any form of *multivariate symbolization* (i.e., the representation of two or more variables in one map), rather than relative values within a single attribute only. Examples from Cartography include ray glyphs (Buja, Cook, & Swayne, 1996), star plot glyphs (Klippel, Hardisty, & Weaver, 2009), and Chernoff faces (Krygier & Wood, 2005).

(5) *Buffer maps* represent a distance zone around a feature or features of interest, such as a school or bar (e.g., Grubesic, Mack, & Murray, 2007). It is possible then to aggregate crime incidents within the buffer zone, representing the frequency using a color gradient (i.e., a buffer map/graduated map combination).

(6) *Interactive maps* leverage a digital environment to allow the map user to manipulate the mapped display according to his or her needs in real time. An interactive map is not a form of cartographic representation, as with the above map types listed by Boba (2005), but rather an additional aspect of a digital map that can be added at varying degrees to any static map (MacEachren, 1994). Thus, cartographic representation (i.e., maps) and *cartographic interaction* (i.e., user interfaces to these maps) are best considered as a fundamental duality within Cartography and GIScience, both of which requiring consideration during map design and development (Roth, 2011, 2012). MacEachren, Wachowicz, Edsall, Haug, and Masters (1999) further parse cartographic interaction into six interaction operators: (1) *focusing/filtering* (increasing or decreasing the detail of a selected subset of map objects; subsequent scholars have interpreted this operator as *filtering* or reducing the number of map objects in the display according to user imposed constraints), (2) *viewpoint manipulation* (panning, zooming, or changing the user's viewing angle of the map), (3) *brushing* (selecting a portion of the map display through direct manipulation of the map in order to perform some operation to the

highlighted features), (4) *sequencing* (dividing the crime information into a set of bins according to time intervals or an attribute of the information), (5) *colormap manipulation* (adjusting the map symbolization, including the map type, color scheme, classification scheme, etc.), and (6) *assignment* (associating a variable in the information set with a component of the map display).

An additional form of cartographic representation discussed by other scholars in crime analysis is the representation of time on maps. The cartographic representation of time focuses on visual depiction of entities and patterns, geographically; it can be used to monitor changing situations and to support more complex spatiotemporal analyses of crime information (which are treated in the subsequent subsection). Spatiotemporal phenomena can be represented by either static maps, which represent temporal change using one or several graphic(s), or animated maps, which represent temporal change in the phenomenon with temporal change in the map (Monmonier, 1990). Starting with the former, there are three general approaches to the representation of multiple points or intervals of time (or any other conceptually bivariate or multivariate representation) on static maps: (1) adjacent displays, (2) separable coincident displays, and (3) integral coincident displays (MacEachren, Brewer, & Pickle, 1998). *Adjacent displays*, or *small multiples*, represent each moment in time or interval of time on a separate map, producing a series of maps with the same spatial extent (Bertin, 1967, 1983; Tufte, 1983). A set of small multiples for crime incidents would divide the information set into a series of time intervals, with each interval receiving its own map and no crime incident occurring on two maps. In contrast, coincident displays juxtapose two or more time states or intervals in a single graphic; the map is termed *separable coincident* when each time period can be individually analyzed visually (e.g., crime incidents from the past 7 days in one color and incidents from the past 8–30 days in a second color) and *integral coincident* when only the difference between time periods can be analyzed visually (e.g., using color to represent the change in crime rates by district following a newly implemented policing tactic). The second general method for representing temporal change—*cartographic animation*—describes the display of individual maps (called frames) in rapid succession (DiBiase, MacEachren, Krygier, & Reeves, 1992). While several research applications of cartographic animation to crime analysis have been reported in the literature (Brunsdon, Carcoran, & Higgs, 2007; Lodha & Verma, 1999; Wolff & Asche, 2009), Ratcliffe (2009) notes that this has translated into little practical application due to a lack of easy-to-use cartographic animation tools and training.

2.3. Spatial and temporal analysis in crime mapping

In practice, the term crime mapping applies to the complete suite of GIScience tools and techniques when used to support crime analysis, including information assembly, spatial statistics, and geocomputation in addition to the aforementioned cartographic themes of representation and interaction (Harries, 1999). Although there is no established taxonomy of spatial analysis techniques for crime analysis, several methods are discussed regularly in the literature on crime analysis and crime mapping. A primary application of spatial statistics and geocomputation to crime analysis is for identification and interpretation of spatial clusters of crime incidents. The most straightforward calculation is *spatial autocorrelation*, which measures the departure from complete spatial randomness (CSR) observed in a distribution of incidents (Griffith, 1987); positive autocorrelation suggests a distribution in which spatially near objects are likely to be similar (i.e., clustered) and negative autocorrelation suggests a distribution in which near objects are likely to be dissimilar (i.e., a checkerboard pattern). Spatial autocorrelation indices such as Geary's C, Moran's I, and Getis's G provide a single value for the entire distribution; however, these

calculations have been extended to provide *local indicators of spatial autocorrelation* (LISA) that identify the location of clusters in the distribution, rather than simply reporting that a distribution is clustered (Anselin, 1995).

A second spatial analysis technique for the identification of clusters is the *spatial scan statistic* (Conley, Gahegan, & Macgill, 2005; Kulldorff, 1997; Openshaw, Charlton, Wymer, & Craft, 1987). A spatial scan statistic is a geocomputational routine that calculates a clustering metric (called the likelihood ratio) for a large number of distinct circular or elliptical sampling windows placed over a crime incident distribution; the output of these algorithms is a small subset of the sampling windows that have a significant number of incidents contained within them as compared to the area not within the window (Chen, Roth, Naito, Lengerich, & MacEachren, 2008). Numerous scholars in criminology and crime analysis have identified the potential of scan statistics for identifying clusters of elevated criminal activity (Chainey et al., 2008; Jefferis, 1998; LeBeau, 2000; Levine, 2006; Nakaya & Yano, 2010; Zeng, Chang, & Chen, 2004). Chen (2009) provides a useful discussion of the conceptual differences between spatial autocorrelation and spatial cluster measures, such as the spatial scan statistics.

Aside from geographic clustering methods, a method of specific interest to crime analysts is *journey-to-crime analysis*, which uses the locations of related crime incidents to determine the most likely areas of offender residence and to forecast the locations of future crimes (Brantingham & Brantingham, 1981). This technique also is referred to as geographic profiling (Rossmo & Velarde, 2008), although this term is being phased out of the literature due to the implication of police surveillance. Two additional, commonly applied spatial analyses are kernel density estimation and buffering, which primarily are applied to generate density maps and buffer maps respectively (described above).

Space and time are paramount to both tactical and strategic crime analysis, as indicated by the dominant theories on criminology described above. As with spatial analyses, temporal analyses and related information graphics primarily are employed for detection of temporal clusters in criminal activity. Modifications to the scan statistic are available to identify crime incident clusters in time alone or in space and time together (Block, 1995; Levine, 2006; Zeng et al., 2004); for these modifications, the scan is completed with a moving time window, rather than or in addition to a moving spatial catchment area. An alternative technique is the *cumulative summation* (*CUSUM*) *algorithm*, which also applies a sliding temporal window to detect aberrations in event activity, such as a spike in crime that is considerably higher than past incident rates (Hutwagner, Thompson, & Seeman, 2003; Maciejewski et al., 2010).

Aside from cluster analysis, there is a small amount of research within crime analysis on the use of temporal information graphics and statistical summaries to complete visually-based *trend analysis* (Chung, Chen, Chaboya, O'Toole, & Atabakhsh, 2005; Ratcliffe, 2004; Townsley, 2008). There also is work on predictive algorithms that attempt to forecast when future crime incidents will occur (Bowers, Johnson, & Pease, 2004); such research may be considered the temporal equivalent of journey-to-crime analysis. A final potentially useful temporal analysis specific to crime information is *aoristic analysis*, a technique for estimating an exact time stamp for a crime that occurred when the victim is not present (e.g., a burglary) based on the time windows of past crimes of the same crime type (Ratcliffe & McCullagh, 1998).

There are many software applications marketed for crime analysis that provide spatiotemporal analysis; most of these applications also support basic cartographic representations and cartographic interactions. Available software packages include: ATAC (Automated Tactical Analysis of Crime; Bair, 2000), Azavea HunchLab/Crime Spike Detector (Cheetham, 2010), CrimeStat (Levine, 2006), ESRI ArcGIS (http://www.Esri.com/software/arcgis/), GeoDa (Anselin, Syabri, & Kho, 2006), MapInfo (http://www.mapinfo.com), ReCAP (Brown, 1998), SaTScan (Kulldorff, 2010), STAC (Block, 1995), and STV (Buetow et al., 2003).

3. Method: needs assessment interviews

3.1. Participants

Seven law enforcement agencies in the United States participated in an interview study designed to assess the current practices and key unmet needs of spatiotemporal crime analysis. Law enforcement agencies were purposefully sampled based on two criteria: (1) the municipal law enforcement agency (six in total) had a daytime service population of 100,000 or greater (all participating law enforcement agencies ultimately had a daytime population of 125,000 or greater) and (2) the police headquarters was within a one day drive (~250 miles) of University Park, PA (the site of the research). One federal law enforcement agency was included in the study to provide a non-municipal perspective. Recruitment was completed via email, with contact information obtained through existing GeoVISTA Center contacts in law enforcement or through agency websites. The sample therefore is representative of intermediate- to large-size law enforcement agencies in the Northeastern United States. The generalizability of results may be limited beyond this context and caution must be applied in interpretation of results due to the relatively small sample of agencies at which interviews were conducted. Each responding law enforcement agency self-identified an individual most appropriate to discuss the spatiotemporal crime analysis practices across their agency. For two of the law enforcement agencies, it was necessary to interview a pair of individuals, as their responsibilities were split according to different internal units; thus, nine interview sessions were completed in total.

A background survey was administered at the start of each interview session to establish several characteristics of the interview participants. Two participants had no post-secondary education, three participants held a Bachelors degree, two participants held a Masters degree, one participant held a PhD, and one participant held a law degree (in addition to a BS in Criminal Justice); outside of the law degree, the degrees were in either Criminal Justice (5) or Geography (2). The participant sample was composed of a near even mixture of primarily producers of spatiotemporal information and associated information products (i.e., crime analysts and crime mappers) and primarily users of this spatiotemporal information and information products (i.e., administrators, detectives, officers, and decision-makers) (Table 1). The majority (7 of 9) of participants reported producing spatiotemporal information and associated information products at least monthly, with a large minority (4 of 9) completing this activity daily. The majority (7 of 9) reported using spatiotemporal information and associated information products at least weekly. Two high ranking officers stated that while they use spatiotemporal information and information products weekly, they never produce them, while two crime analysts stated that while they produce spatiotemporal information and information products weekly, they never use them for policing or decision making purposes. Four participants were sworn officers while the other five held civilian status.

3.2. Materials and procedure

Interviews vary on the degree of structure in their questioning (Robinson, 2009). Structured interviews include a series of focused

Table 1
Interview participant regularity of producing and using spatiotemporal information and associated information products.

Regularity of activity	Produce spatiotemporal info	Use spatiotemporal info.
Daily	4	2
Weekly	1	5
Monthly	2	0
Yearly	0	0
Rarely	2	2
Total	**9**	**9**

questions that typically prompt short and equally focused responses; all participants are asked the exact same set of questions in the same order. On the other end of the continuum, unstructured interviews include a set of broad discussion topics or general themes, with no preset order; these types of questions are exploratory in nature and typically prompt longer, open-ended responses that vary greatly from person to person. Many interview protocols follow a semi-structured approach, which starts with a set of focused questions but allows the interviewer to ask follow-up or probe questions as he or she sees fit and change the order of questioning if appropriate (for an example in crime mapping, see Ratcliffe, 2000).

At the end of the interview, participants were asked if there was anything else they would like to discuss before concluding or if they had any questions about the study. All interview sessions lasted between 60 and 75 minutes and were completed at the participant's work location in a private room. For consistency, the same project member acted as the interviewer for all nine interviews. The interviews were audio recorded for subsequent qualitative data analysis, as described in the following subsection.

The interview protocol for the needs assessment proceeded in six sections; a summary of the interview questions is included in Table 2. Each interview session began with an introduction to the project and

Table 2
A summary of the interview questions.

Introduction	
Background	
1	What is your agency, department, or organization, job title, and responsibilities at this position?
2	Are you a sworn officer or a civilian?
3	Please describe your prior education and formal training?
4	Please describe any previous employment relevant to crime mapping and analysis?
5	How frequently do you produce spatiotemporal information and associated information products (maps, analyses, etc.) in your daily work?
6	How frequently do you use spatiotemporal information and associated information products (maps, analyses, etc.) in your daily work?
Information	
7	Please list the types of spatial or temporal phenomena for which your agency collects information.
8	For each collected information set, describe its: format, number of entities/records, geographic and temporal resolution, scale of analysis and mapping.
9	Does your agency use any external information sources?
10	Is the information your agency collects text/report-based or entered into a table or database?
11	Are there any information sets not collected by your agency that would be useful in crime mapping and analysis?
Mapping and Analysis	
12	Please describe the kinds of maps produced by your agency.
13	Please list the reference or basemap information your agency uses on these maps.
14	What spatial analyses or data transformations does your agency apply to the collected raw information?
15	What temporal analyses or models does your agency apply to the collected raw information?
16	Does your agency aggregate your point incident information in space or time?
17	Does your agency filter your point incident information prior to mapping?
18	Does your agency represent the temporal component of your information directly on maps?
Use	
19	How are maps and analyses used in a tactical way at your agency?
20	How are maps and analyses used in a strategic way at your agency?
21	What is the workflow from generation of maps and analyses to usage of these information products at your agency?
22	Please describe a successful use of mapping and analysis at your agency?
23	Please describe an unsuccessful use of mapping and analysis at your agency?
Do you have any last questions or comments before we conclude the session?	

an overview of the goals of the needs assessment; here, participants were informed that they did not have to respond to all questions, particularly if the question was irrelevant or sensitive. Participants then were asked two sets of brief, structured questions. Participants first responded to structured questions about their general background in law enforcement and their overall experience producing and >using spatiotemporal information and associated information products in support of crime analysis (summarized in the previous subsection). After the background questioning, participants were asked a set of structured questions about characteristics of the geographic information that their agency collects and maintains, as well as any external geographic information sources that their agency leverages.

Following the structured portion of the interview, the participants were asked two rounds of semi-structured questions. The first round of semi-structured questioning focused upon the current crime mapping practices from an information producer perspective, asking about the types of maps that are generated and the types of spatial and temporal analyses that are applied to the information. The second round of semi-structured questioning focused upon the current crime mapping practices from an information user perspective, asking about tactical and strategic uses of crime mapping, the general crime analysis workflow, and examples of successes and failures when using maps and analyses to support the mission of law enforcement.

3.3. Qualitative data analysis

Qualitative data analysis (QDA) describes the systematic interpretation of qualitative information, such as text reports, websites, photos, maps, and field observations (Dey, 1993; Miles & Huberman, 1994). A review of work using qualitative data analysis on electronic government information, government information products, and government information use is provided by Yildiz (2007), with a multitude of examples published more recently in Government Information Quarterly outside of the domains of law enforcement and public safety. In the most robust form of QDA, the documents in the set are decomposed to their smallest unit of analysis and a series of codes are applied to the units by several independent coders, with the coding then compared across coders to ensure reliability in interpretation of the document set.

Transcription of the audio recordings was completed using Transana, with the transcripts then unitized at the statement level in Microsoft Excel for margin coding (Bertrand, Brown, & Ward, 1992). The above background review on the current state of science in spatiotemporal crime analysis was used to identify six key themes: geographic information (G), cartographic representation (R), cartographic interaction (I), spatial analysis (S), temporal analysis (T), and map and analysis use (U). These key themes are areas in which law enforcement agencies may have an unmet spatiotemporal analysis need, defined as a resource or feature required by the targeted end user to complete their work and thus represents a disconnect between the current states of science and practice in spatiotemporal crime analysis. Thirty-one individual codes then were identified from the above background review within these six needs; each code was marked during margin coding to distinguish needs that were met by existing software (+) from those that were not met (−) at the time of the interview. Table 3 lists the six higher level categories, each of the 31 codes across these categories, and the source of the individual code from the above background review; Table 4 lists the frequency of each code across the nine transcripts. A total of 515 codes identifying user needs were applied to the nine transcripts, an average of 57.2 codes per transcript.

Two coders with expertise in GIScience and training in crime analysis were hired to apply independently the same 31-part coding scheme used in the initial coding, with code reliability assessed using the inter-rater reliability score described by Robinson (2008). The two coders achieved inter-coder reliability scores of 93.2% and 87.6% against the initial margin coding, indicating a high degree of reliability in the interpretation and application of the coding scheme, particularly considering the large number of codes in the coding scheme. Differences in coding were reconciled for reporting through discussion among the coders and a third project member. Statements were sorted according to the assigned code and summarized using the synoptic style of reporting described by Monmonier and Gluck (1994) and Roth (2009). Crime analysis needs within the six higher level categories are summarized in the following section.

4. Results and discussion: current state of practice

4.1. Geographic information

Codes included in the geographic information (G) category indicate statements about the geographic information sets leveraged to support crime analysis. Five codes were included under the geographic information (G) category based upon the above background review: (G1) crime reports (incidents plus arrests), (G2) calls for service, (G3) vehicle recoveries, (G4) field interviews, and (G5) any external information sources not collected or maintained by the law enforcement agency itself. The most frequently discussed geographic information sets include crime reports (average = 6.6) and external information sources (average = 6.3), with participants identifying external information sources as an unmet need (average = 1.9) slightly more frequently than crime reports (average = 1.8). Participants rarely discussed calls for service (average = 1.6), vehicle recoveries (average = 0.8), and field interviews (average = 0.2).

Overall, participants indicated that crime reports are the primary geographically-referenced information collected and used at their law enforcement agencies. Discussion centered almost exclusively on crime reports describing incidents, rather than arrests. The number of crime incident records collected per year by the interviewed agencies ranges from approximately 7,000 to 2.5 million, indicating a need for user interfaces to scale to increasingly large and complex information sets. All participants described a similar set of core attributes captured in their crime incident reports: crime type (by UCR code), address, date and time (often with precision to the minute, except in the cases of burglary when a time range is given), MO, suspect and victim description, and a text narrative. Surprisingly, one participant noted that his/her agency did not regularly geocode (i.e., convert the listed address to spatial coordinates) their crime incident reports for mapping and analysis, instead geocoding only a small grouping of crime incident reports if an association is suspected. One participant also noted that his/her agency also captures information on location type, such as "parking lot, convenience store, restaurant, street, sidewalk"; while this information is not geographic in the sense of absolute coordinates, it is highly relevant to spatiotemporal crime analysis as it provides important geographic context for understanding the crime setting.

Most participants indicated that their agency leverages externally maintained geographic information sources. One participant stated that "we have gone out and tried to collect as many datasets as we can find that may or may not be useful to us, just so we know where they are at and what we have access to." Two important geographic information sets mentioned repeatedly were parole/probation records and registered sex offender records maintained at the state level, both of which include the home address of the offenders. Departments that have access to this information emphasized its utility and those that do not have access acknowledged their desire to acquire it. Other information sets include DMV (Department of Motor Vehicles) records and infrastructure information from the City's GIS department. One external geographic information set that is not used regularly is the federal census, with one participant stating that "I had to jump through hoops just to break it down by district and section within the police department" and a second stating that "the census data is about nine years old now [and] just isn't accurate…the census doesn't really mean much of anything to us." This contradicts descriptions of crime analysis in the

Table 3
The coding scheme applied for QDA of the needs assessment study. The categories of needs and individual codes were derived from the background review.

ID	Name	Source
Geographic information: Statements about the information sets used and their characteristics		
G1	Crime reports (incidents and arrests)	Harries (1999)
G2	Calls for service	Harries (1999)
G3	Vehicle recoveries	Harries (1999)
G4	Field interviews	Osborne & Wernicki (2003)
G5	External information sources	Harries (1999)
Cartographic representation: Statements about the way information sets are mapped		
R1	Push pin maps (i.e., one-to-one dot or single-symbol maps)	Boba (2005)
R2	Choropleth maps (i.e., graduated maps using color)	Boba (2005)
R3	Proportional symbol maps (i.e., graduated maps using size)	Boba (2005)
R4	Hot spot maps (i.e., density maps)	Boba (2005)
R5	Multivariate symbolization (i.e., chart maps)	Boba (2005)
R6	Buffer maps	Boba (2005)
R7	Maps representing time	Monmonier (1990)
R8	Reference or basemap symbolization	Dent (1999)
Cartographic interaction: Statements about the way in which maps are manipulated		
I1	Focusing/filtering	MacEachren et al. (1999)
I2	Viewpoint manipulation	MacEachren et al. (1999)
I3	Brushing	MacEachren et al. (1999)
I4	Sequencing	MacEachren et al. (1999)
I5	Colormap manipulation	MacEachren et al. (1999)
I6	Assignment	MacEachren et al. (1999)
Spatial analysis: Statements about applied spatial statistics and geocomputation		
S1	Spatial autocorrelation measures	Griffith (1987); Anselin (1995)
S2	Spatial scan statistics	Openshaw et al. (1987)
S3	Journey–to–crime analysis (i.e., geographic profiling)	Brantingham & Brantingham (1981)
Temporal Analysis: Statements about applied temporal transformations and models		
T1	Temporal and spatiotemporal cluster analysis	Zeng et al. (2004)
T2	Trend analysis	Ratcliffe (2004)
T3	Predictive analysis	Bowers (2004)
T4	Aoristic analysis	Ratcliffe & McCullagh (1998)
Map & analysis use: Statements about the use of maps and analysis to support law enforcement		
U1	Criminal investigative analysis	Boba (2005)
U2	Intelligence analysis	Boba (2005)
U3	Tactical crime analysis	Boba (2005)
U4	Strategic crime analysis	Boba (2005)
U5	Administrative analysis	Boba (2005)

Table 4
Frequency of codes applied for QDA of the needs assessment study. *Total* describes the total number of statements given the code, while *Avg* divides this total by the sample size (n = 9).

ID	Name	Have		Need		All	
		Total	Avg	Total	Avg	Total	Avg
G1	Crime reports	43	4.8	16	1.8	59	6.6
G2	Calls for service	9	1.0	5	0.6	14	1.6
G3	Vehicle recoveries	5	0.6	2	0.2	7	0.8
G4	Field interviews	2	0.2	0	0.0	2	0.2
G5	External information sources	40	4.4	17	1.9	57	6.3
Total geographic information (G)		**99**	**11.0**	**40**	**4.4**	**139**	**15.4**
R1	Push pin maps	26	2.9	6	0.7	32	3.6
R2	Choropleth maps	13	1.4	3	0.3	16	1.8
R3	Proportional symbol maps	7	0.8	2	0.2	9	1.0
R4	Hot spot maps	15	1.7	2	0.2	17	1.9
R5	Multivariate symbolization	1	0.1	1	0.1	2	0.2
R6	Buffer maps	11	1.2	0	0.0	11	1.2
R7	Maps representing time	33	3.7	9	1.0	42	4.7
R8	Reference or basemap symbolization	29	3.2	8	0.9	37	4.1
Total cartographic representation (R)		**135**	**15.0**	**31**	**3.4**	**166**	**18.4**
I1	Focusing/filtering	18	2.0	10	1.1	28	3.1
I2	Viewpoint manipulation	5	0.6	1	0.1	6	0.7
I3	Brushing	11	1.2	1	0.1	12	1.3
I4	Sequencing	3	0.3	2	0.2	5	0.6
I5	Colormap manipulation	1	0.1	0	0.0	1	0.1
I6	Assignment	3	0.3	0	0.0	3	0.3
Total cartographic interaction (I)		**41**	**4.6**	**14**	**1.6**	**55**	**6.1**
S1	Spatial autocorrelation measures	0	0.0	2	0.2	2	0.2
S2	Spatial scan statistics	4	0.4	1	0.1	5	0.6
S3	Journey-to-crime analysis	7	0.8	2	0.2	9	1.0
Total spatial analysis (S)		**11**	**1.2**	**5**	**0.6**	**16**	**1.8**
T1	Temporal & spatiotemporal cluster analysis	3	0.3	0	0.0	3	0.3
T2	Trend analysis	23	2.6	5	0.6	28	3.1
T3	Predictive analysis	4	0.4	0	0.0	4	0.4
T4	Aoristic analysis	3	0.3	0	0.0	3	0.3
Total temporal analyses (T)		**33**	**3.7**	**5**	**0.6**	**38**	**4.2**
U1	Criminal investigative analysis	3	0.3	2	0.2	5	0.6
U2	Intelligence analysis	8	0.9	3	0.3	11	1.2
U3	Tactical crime analysis	34	3.8	3	0.3	37	4.1
U4	Strategic crime analysis	19	2.1	7	0.8	26	2.9
U5	Administrative analysis	20	2.2	2	0.2	22	2.4
Total map & analysis uses (U)		**84**	**9.3**	**17**	**1.9**	**101**	**11.2**
Total		**403**	**44.8**	**112**	**12.4**	**515**	**57.2**

literature, where the integration of census information for strategic crime analysis is highly recommended (e.g., Cahill & Mulligan, 2007); the recent release of the 2010 census may alleviate the latter participant's concern, at least for the small window of time that the census is current enough for the purpose of crime analysis. Many of the participants stated that their agencies use volunteered geographic information collected from social networking websites such as Facebook and MySpace, primarily for link analysis; none of the participants recalled using volunteered geographic information posted to microblogging websites like Twitter, but stated that they would like to do so if simple methods were available. While most law enforcement agencies appear willing and able to synthesize a large amount of external information sources, it is important to note that participants knew little about if or how their internally maintained information are shared with other agencies within their municipality or other law enforcement agencies in neighboring cities.

4.2. Cartographic representation

Codes included in the cartographic representation (R) category indicate statements about the way in which the collected geographic information are represented in map form to support crime analysis. Eight codes are included under the cartographic representation (R) category based upon the above background review: (R1) push pin maps (i.e., one-to-one dot or single-symbol maps), (R2) choropleth maps (i.e., graduated maps by color), (R3) proportional symbol maps (i.e., graduated maps by size), (R4) hot spot maps (i.e., density or surface maps), (R5) multivariate maps, (R6) buffer maps, (R7) maps representing time, and (R8) aspects of the underlying basemap. On average, cartographic representation was the most mentioned of the six key themes, indicating that the design of the map remains as important—or more so, as suggested by the code frequency—as more technically complex spatial and temporal analyses. The most frequently discussed cartographic representation forms include maps representing time (average = 4.7), basemaps (average = 4.1), and push pin maps (average = 3.6). Less discussion was elicited concerning hot spot maps (average = 1.9) and choropleth maps (average = 1.8). Participants infrequently identified buffer maps (average = 1.2), proportional symbol maps (average = 1.0), and multivariate maps (average = 0.2) as key needs, either met or unmet.

Of the set of crime map types identified by Boba (2005), push pin maps were by far the most commonly identified as a core need by participants. Participants indicated that the primary explanation for the frequent employment of push pin maps was their simplicity, but their simplicity in interpretation by map users rather than their simplicity in creation by mapmakers. One participant stated that "for the most part, simpler is better for the [officers] on the street" and second stated that "a lot of times we want to do more fancy and sophisticated analytical maps, but [high ranking officials] want to see pin maps, so of course we have to do pin maps." Such statements suggest that a large number of the information users at law enforcement agencies currently are incapable or unwilling to utilize more complex cartographic representations in support of criminal investigation and resource allocation. Interestingly, at least one agency still commonly adds push pins to printed wall maps manually for serial tracking and collaborative decision-making.

Despite its relatively small amount of overall discussion within the cartographic representation (R) category, participants identified the hot spot map (i.e., density or surface map) as the preferred cartographic representation technique for large volume crimes where aggregation is necessary. Several participants noted that the generation of hot spot maps is growing in popularity, with one participant stating that "the latest and greatest thing that people like to see is a hot spot map" and a second stating that "it was something that my analysts saw at a conference, [so] we started making hot spot maps." Several of the participants responsible for producing hot spot maps indicated that they primarily use the kernel density estimation (KDE) function in the 3D Analyst extension to Esri's ArcGIS, which uses a moving window (i.e., a kernel) or multiple pixels to generate a crime estimation for the central pixel. However, several participants were cautious about inappropriately using hot spot maps, fearing that officers and detectives "don't understand them." One agency specifically avoided the use of KDE because they considered it misleading, as their map users did not realize that the shading is a smoothed result of a search window and not an aggregate of crime incidents within the specific pixel. Interestingly, several participants identified hexagons, rather than squares, as the preferred tessellation (i.e., cell shape), as the representation leads to naturally shaped hot spots that are easier to interpret and use for allocating patrol.

Participants generally considered choropleth maps as inferior to hot spot maps for crime analysis, as choropleth maps aggregate crime information to political or jurisdictional units that have little impact on patterns of criminal activity. Only one law enforcement agency regularly generated choropleths instead of hot spot maps for tactical crime analysis. However, choropleth maps are generated regularly for strategic and administrative crime analysis; one participant stated that "with more strategic or long-term maps, then we will do a choropleth map" and a second stated that choropleth mapping is "more of the administrative work" that he/she does. Other maps that were created on occasion by a subset of departments include graduated symbol maps (graduated by point locations and, at one department, by line segment), flow maps (primarily to connect the location of vehicle thefts versus recoveries, but also to connect crime incidents in a serial; never scaled to show the volume of flow), and buffer maps. Most mapmaking is completed using Esri's ArcMap, although one agency exclusively uses the Microsoft MapPoint software to produce push pin maps.

Surprisingly, cartographic representation of time was the most discussed of the themes included in the cartographic representation (R) category; maps representing time also were the most frequent cartographic representation form listed as a need that currently is unmet. Coloring the pins on a push pin map according to the date of the incident—a separable coincident technique—was identified by participants as the primary method for representing time on maps. One participant stated that "I will have 28 days in color, the previous 28 days in grey, and the current 7 days will be in purple", a second participant stated that he/she will use "a color ramp to show thirty incidents over a two month period…the initial incident may be a white dot and it progresses to red over time", and a third participant stated that he/she would produce "a map of the last 30 days, [with] halos around the different periods of times." Participants identified the last seven days, the last month, and the last two months as common time periods used for temporal colored push pin maps. Several departments also apply color to represent cyclical temporal patterns, such as different days of the week (e.g., 'Monday', 'Tuesday', 'Wednesday') or shifts ('8 am–4 pm', '4 pm–12 am', '6 pm–2 am', '12 am–8 am'). Several participants also remembered isolated times that they had created animations when specifically prompted in the interview, with frames typically forwarded manually using Microsoft Powerpoint; one participant did report using Adobe Flash once to generate an animation. While these participants stated that the animations were extremely well received, they also stated this was not an approach they typically completed due to the perceived time-consuming nature of constructing the animation, making animation a key unmet cartographic representation need.

Participants agreed that a street network with labels is the primary basemap or reference information included on their crime maps. Other infrastructure information like building footprints and parcels may be included for large scale maps. Participants noted that points of interest (e.g., schools, parks, police stations, bars, restaurants, bus stops) may be included, but typically only upon special request or for the generation of buffer maps. Most agencies have access to

aerial imagery, but generally only include it upon special request (particularly for court maps).

4.3. Cartographic interaction

Codes included in the cartographic interaction (I) category indicate statements about the way in which the generated cartographic representations are manipulated through user interfaces. Six codes are included under the cartographic interaction (I) category based upon the above background review: (I1) focusing/filtering, (I2) viewpoint manipulation, (I3) brushing, (I5) colormap manipulation, and (I6) assignment. Focusing/filtering was identified as the biggest need (average = 3.1) among the MacEachren et al. (1999) cartographic interaction operators, with brushing (average = 1.3) also garnering some discussion. Viewpoint manipulation (average = 0.7), sequencing (average = 0.6), assignment (average = 0.3), and colormap manipulation (average = 0.1) were discussed infrequently.

Participants stated that most of the exploratory crime analysis leveraging cartographic interaction is completed with desktop GIS software designed for other purposes. While some law enforcement agencies have generated customized interface widgets providing some cartographic interaction in real-time, this still is limited to a subset of interaction operators and a subset of agencies. Across the nine participants, cartographic interaction was performed only by the crime analysts responsible for producing crime maps and analyses. Participants from agencies that hold CompStat meetings noted that high-ranking officers may include interactive maps in their presentations, manipulating the cartographic representation in real-time; however, in all reported examples "analysts are in the back driving that." Thus, a transparently usable interface, providing a subset of core interaction operators through an intuitive interface design, may fill a key unmet need for the consumers of crime maps, such as administrators, detectives, supervisors, and other decision makers.

Participants identified focusing/filtering as the most needed cartographic interaction operator, with one participant acknowledging that crime analysts "filter continuously, every time they make a map they filter." Most participants have to perform focusing/filtering queries using a series of *nested dialog windows*, which hide interface features in a set of windows that must be activated in sequence by users, limiting the *usability* (i.e., ease-of-use) of the application and making exploratory crime analysis difficult. One recommended solution is the inclusion of *persistent dialog windows* housing focusing/filtering controls that remain visible until minimized by the user; this was deemed particularly appropriate for common filtering attributes such as UCR and MO. Participants also identified brushing as a commonly employed cartographic interaction operator, noting that brushing typically is provided on digital push pin maps in order to retrieve additional information about the selected crime incident. Only one participant described an application currently in use at his/her agency that uses brushing for linked highlighting across multiple information graphics (a desktop mashup between ArcGIS and the ATAC system).

Participants indicated that the other cartographic interaction operators are employed infrequently. Participants from agencies that hold CompStat meetings noted that the sequence operator sometimes is applied during the meeting, as they have to bin their crime incident information across multiple attributes for each weekly or bimonthly meeting. However, the generated maps almost never are animated across the temporal bins generated by the sequence operator. Participants stated that viewpoint manipulation often is available only in a discrete fashion (i.e., no continuous panning across the map extent or zooming across scales), as analysts have the extent of each district preset in ArcMap and toggle between individual districts and the full extent. Participants indicated that assignment and colormap manipulation are applied rarely.

Interestingly, participant discussion on cartographic interaction revealed a split on the potential utility of web maps and web mapping services, such as Google Maps. One participant was excited about the potential of such services, stating that "a lot of folks are looking at the Google Maps…I think Google Maps has been really useful in getting law-enforcement to use these types of things because prior to Google Maps, agencies didn't even know these things were accessible" while a second was concerned not about the interactivity of these services, but about the underlying information quality, stating that "if you go to Google Maps, or something like that, you don't know how old those maps are or what actually changed, so some of that information when you physically get out there could be bad information." Interestingly, one participant stated that he/she had experimented with using Google Earth because of the potential for sharing interactive maps via .kml files; this participant stopped doing this, however, because the intended users did not have Google Earth installed on their work machine (and did not have security permission to do so) and therefore could not access the maps.

4.4. Spatial analysis

Codes included in the spatial analysis (S) category indicate statements about the spatial statistics and geocomputational routines that are applied in support of crime analysis. Three broad codes were included under the spatial analysis (S) category based upon the above background review: (S1) measures of spatial autocorrelation, (S2) spatial scan statistics, and (S3) journey-to-crime analysis; other spatial analyses were considered in the original coding scheme, but were dropped because they were not identified during the interviews. Spatial analysis was by far the least identified need during the transcript analysis, with an overall average of only 4.2 statements per transcript; journey-to-crime analysis was identified most frequently as a spatial analysis need (average = 1.0), followed by spatial scan statistics (average = 0.6) and spatial autocorrelation measures (average = 0.2).

The overall low amount of discussion on spatial analysis needs revealed a large and unexpected disconnect between practice and science, where the application of spatial transformations and spatial models is frequently reported and highly recommended. One possible explanation for this disconnect that came up in several of the interviews was that there is a lack of relevant expertise within the municipal law enforcement agencies in terms of both understanding how to apply the spatial analysis techniques and how to interpret their results. There were only two total references to spatial analysis by sworn officers, with one participant stating that he/she "leave[s] that up to the analyst to do because they are a lot more familiar with those type of things that I am." There also was a general notion communicated by many of the participants that crime analyst units are undermanned and even misused, forcing the crime analysts to respond to specific, often basic requests rather than providing them with the autonomy to complete more advanced spatial analyses. One participant noted his/her agency "is very short on manpower… so it becomes very difficult [to complete such analyses]." This participant went on to add that "I don't want to say it's a waste of time, but they just don't have the time to focus on this."

The two most commonly applied spatial analyses—kernel density estimation and buffering—are completed to generate an output map. Interestingly, participants from two different agencies commonly apply these two spatial analyses upon request during their CompStat meetings, illustrating the potential of providing cartographic interfaces to computational processes in support of exploration and reasoning in real-time. Several participants stated that they occasionally conduct journey-to-crime analyses using the geographic mean calculation in the CrimeStat application (Levine, 2006) or the animal movement extension in ArcGIS. None of the participants calculated spatial autocorrelation statistics, with most participants seeing limited value in metrics that do not provide local indicators of crime clusters. One participant had applied the spatial scan statistic routines provided in SaTScan

(Kulldorff, 2010) and the LISA statistics provided in GeoDa (Anselin et al., 2006) several times, but noted that such application "is unusual… [crime analysts] don't use GeoDa, they don't even use SaTScan." While other participants hinted at the need to automate the identification of crime incident clusters, no other participants were aware of spatial scan statistics or other methods of spatial cluster analysis, stating that cluster identification is completed visually at their agencies.

4.5. Temporal analysis

Codes included in the temporal analysis (T) category indicate statements about the temporal transformations and models that are applied in support of crime analysis. The temporal analysis (T) category includes four codes drawn from the above review: (T1) temporal and spatiotemporal cluster analysis, (T2) trend analysis, (T3) predictive analysis, and (T4) aoristic analysis. Trend analysis using information graphics was identified by participants as the largest temporal analysis need (average = 3.1). Predictive analysis (average = 0.4), temporal and spatiotemporal cluster analysis (average = 0.3), and aoristic analysis (average = 3.1) were discussed infrequently.

There was extreme variation in the temporal analyses applied to the crime information across the participating law enforcement agencies. Most of the participants indicated that their agencies apply very little temporal analyses in support of crime analysis. These participants primarily rely upon trend analysis that does not have a linked cartographic component to it. Participants stated that the temporal information graphics used for trend analysis typically are generated as one-offs in Microsoft Excel or Crystal Reports and are restricted to incidents in the past one or two months, indicating an emphasis on tactical over strategic crime analysis (see the following subsection). Interactive integration of these temporal graphics with map views was identified as a key unmet need. Participants noted that there is little systematic interpretation of the time-series graphics, with one participant saying "there is quite a bit of qualitative feel to it." Most participants stated that their agencies generate temporal composite graphics that show cyclical crime patterns, such as peaks by day of the week or time of the day; at several agencies, however, these graphics only are created when a crime analyst notices a potential cyclical pattern when reading the individual crime reports, rather than generating the composite graphics regularly to identify patterns without sifting through the full incident narrative. For these agencies, the application of temporal or spatiotemporal clustering analysis, temporal prediction algorithms, and aoristic analysis was minimal.

In contrast, participants from two of the law enforcement agencies indicated that they regularly apply sophisticated temporal analyses, offering initial insight into potential use case scenarios of spatiotemporal crime analysis. One agency takes full advantage of the ATAC software (Bair, 2000). ATAC automates the generation of temporal information graphics for trend analysis and provides a suite of statistics to help with the interpretation of these graphics. ATAC also has an aoristic analysis feature that provides a time split for incidents logged with a time interval. Further, ATAC includes a temporal prediction feature, although the participant using ATAC stated "I don't understand the [technique] very well" and that "this technique didn't seem to work very well for me." Finally, as mentioned above, the ATAC software is fully linked with ArcMap, allowing for real-time, interactive exploration of criminal activity in both space and time.

A second agency employs the HunchlLab/Crime Spike Detector software developed by Azavea (Cheetham, 2010), referred to internally as SpikeStat. The purpose of SpikeStat is to automate the identification of spatiotemporal hot spots of criminal activity. The participant reported that the software uses a spatiotemporal scan statistic to identify crimes spikes in both space and time. Interestingly, the implemented spatiotemporal scan statistic uses different window sizes based upon the type of crime under investigation because, as the participant noted, "you want to have a smaller search radius for thefts and a bigger one

for homicide." The locations of these spikes then are compared against the set of police jurisdictions in order to send an alert to the appropriate commanding officer. The crime analyst using SpikeStat was pleased with the results, stating that it works "like an early warning system."

4.6. Map and analysis use

Codes included in the map and analysis use (U) category indicate statements about the way in which spatiotemporal mapping and analysis techniques are used in support of crime analysis. The map and analysis use (U) category includes a code for each of Boba's (2005) five types of crime analyses reviewed above: (U1) criminal investigative analysis, (U2) intelligence analysis, (U3) tactical crime analysis (U4) strategic crime analysis, and (U5) administrative analysis. Support for tactical crime analysis was identified by participants as the overall largest need (average = 4.1), but strategic crime analysis was identified as the largest unmet need (overall average = 2.9; unmet average = 0.8). Participants also identified support for administrative analysis as an important need (average = 2.4); support for intelligence analysis (average = 1.2) and criminal investigate analysis (average = 0.6) were less frequently discussed.

An interesting characteristic of the current practice of crime analysis was revealed when comparing discussion on tactical versus strategic crime analysis. Most of the participating law enforcement agencies primarily conduct tactical crime analysis, applying spatiotemporal analyses and producing output crime maps in order to react to the most recent crime spikes. One participant stated that "tactically is probably how we most often use our maps" and went on to say "a lot of what we do is more just support of day-to-day functions of the police department, so you don't see [strategic crime analysis] a lot…we are more tactical." A second participant stated that "currently the way that we are utilizing [crime analysis] is to attack specific problems" and went on to say "by the time we start gathering the clustering, we have attacked the clustering, and once that clustering is eradicated, we then move on; we do not have the time or luxury to see if the clustering started nine months ago." Participants indicated that the purpose of the tactical analysis is to adjust blue force patrolling in response to recent criminal activity; one participant stated that the "most common use [of tactical crime analysis] is for patrol deployment" while a second stated "we do tactical analysis for patrol." However, at least one participant lamented this focus of crime analysis, stating "that is why we play catch-up most of the time, because it is all reactionary."

While all but one participant could think of at least a single example of strategic mapping completed at their agency, only three of the interviewed agencies considered themselves positioned to complete strategic crime analysis on a regular basis. One key barrier to strategic analysis identified by the participants is that many agencies are undermanned, a similar barrier preventing more sophisticated spatial analyses. With regard to conducting strategic crime analysis, one participant stated that "allocating resources is a big deal" while a second participant stated "if they are overloaded with requests they won't have time to do [strategic analysis]", and a third stating that his/her agency conducts strategic analysis "whenever they get grant money or extra money." Thus, it appears as though law enforcement agencies currently are in need of a cartographic interface that supports rapid and straightforward strategic analyses in addition to tactical analyses.

A second, and likely related, barrier to strategic analysis identified by several of the participants is accountability, as enacting intervention programs based on strategic analysis requires a long-term commitment from decision makers. As one participant noted "with strategic projects, it is a lot easier to ignore them than actually go out and initiate projects with them because it takes a lot of time and effort to do a [strategic] project … if we are not being held accountable, it is a lot easier for some people not to do it." The best examples of cartographic representations, and associated user interfaces,

that support strategic crime analysis were provided by participants whose agencies hold regular CompStat meetings; four of the seven participating agencies hold CompStat meetings, with three of these four agencies conducting regular or semi-regular strategic crime analysis. These participants noted that the primary purpose of the CompStat process is to increase accountability, which has a direct tactical goal of reacting to recent crime spikes, but also should support the long-term strategic goal of improving the quality of life of a community.

Beyond Boba's (2005) tactical and strategic forms of crime analysis, participants also were able to provide numerous examples of administrative analysis. Almost all participants indicated that their agencies generate maps for court proceedings, with several agencies balancing their budgets by charging for the preparation of these maps. Many participants also described map-based reports or simple online mapping websites that are maintained by their agency for the public consumption of spatiotemporal crime information. Participants provided few examples of criminal investigative analysis and intelligence analysis during the interviews. Examples of criminal investigative analysis primarily referenced applications of the journey-to-crime analysis described above. Examples of intelligence analysis primarily referenced the extraction of geographic information from social networking applications to identify the location and relationships of a gang or other organized crime syndicate.

5. Conclusion and outlook: unmet needs for spatiotemporal crime analysis

This article provides a snapshot and comparison of spatiotemporal crime analysis science and practice, with the aim of revealing major disconnects and currently unmet needs. This research contributes to our knowledge of spatiotemporal crime analysis in two ways. We first completed a comprehensive background review across the domains of Criminology/Crime Analysis and GIScience/Cartography in order to characterize the current science of spatiotemporal crime analysis. We then conducted a set of interviews with seven law enforcement agencies in order to compare our background review to the current practice of spatiotemporal crime analysis; again, the insights elicited from the interviews are specific to intermediate- to large-size law enforcement agencies located in the Northeastern United States. The comparison between science and practice was completed across six themes relevant to spatiotemporal crime analysis: (1) geographic information, (2) cartographic representation, (3) cartographic interaction, (4) spatial analysis, (5) temporal analysis, and (6) map and analysis use. Importantly, the comparison between the background review and interview responses revealed several broad, unmet needs for spatiotemporal crime analysis in United States law enforcement agencies, each of which span across several or all of these six themes:

(1) *Expand and combine geographic information sources*: All of the participating law enforcement agencies indicated the need to acquire geographic information from additional sources. Participants noted two internal or government information sources, which are compiled in a consistent and top-down manner: parole/probation records and registered sex offender records. However, many intriguing comments were offered from participants with regards to external information sources. Law enforcement personnel need applications that allow for fast and flexible combination of internal and external information sources, an approach described in information science as a *mashup* or, regarding online applications, *Web 2.0* technologies (O'Reilly, 2007; Roth et al., 2008). They also require cartographic representations and cartographic interactions that scale to the growing size of these information sets, particularly volunteered geographic information sources such as Facebook and Twitter. Further, participants indicated that they are not fully aware how their internally maintained information sets are used by other agencies at the municipal, state, or federal level. Greater coordination across agencies involved in law enforcement and public safety would act to refine the database schema to better support diverse information uses, promote transparency and collaboration across agencies, and remove overlap in collection and maintenance efforts.

(2) *Improve the usability of crime mapping and analysis tools*: Acquisition of additional information sources means little if this information cannot be made usable through mapping and analysis techniques that are both easy to perform and comprehend. The above background review yielded a large number of techniques regarding crime mapping and analysis that support the mission of law enforcement. However, based on our empirical results, only a portion of this spatiotemporal crime analysis toolkit regularly is put to use in intermediate- to large-size law enforcement agencies; cartographic representation is limited primarily to pushpin and hotspot maps, cartographic interaction is limited primarily to the focusing/filtering and brushing, spatial analysis is surprisingly limited altogether, and temporal analysis exhibits a large amount of variation across agencies. Rather than continuing the pursuit of novel spatiotemporal crime mapping and analysis tools and techniques to add to those reviewed above, researchers perhaps instead should be investigating how to make existing tools and techniques *transparently usable* (i.e., immediately can be used by law enforcement with little training) (Robinson, Roth, & MacEachren, 2011). The topic of usability is one that has received minimal attention within crime analysis and spatial criminology, but one that is of fundamental importance considering that most-to-all law enforcement personnel are not formally trained in spatiotemporal crime analysis and have little experience interpreting complex cartographic representations and spatiotemporal analytical results. By placing an emphasis on the design of the user interfaces to the mapping and analysis techniques—rather than the techniques themselves, and perhaps even limiting their sophistication depending on the use case scenario—a greater number of law enforcement personnel can integrate spatiotemporal analysis into their workflows. Pervasive use of highly usable interfaces to simplified spatiotemporal mapping and analysis techniques also may promote buy-in within the agency to allow dedicated crime analysts to spend their time performing more sophisticated mapping and analysis techniques.

(3) *Integrate geographic and temporal representations and analyses*: Criminal activity has prominent spatial and temporal components that must be treated in concert during the analysis of crime information in order to glean the maximum amount of insight in to the identified pattern. Feedback elicited from the interview study supports Ratcliffe's (2009: 12) assessment that "At present, the most under-researched area of spatial criminology is that of spatio-temporal crime patterns." Across the seven participating law enforcement agencies, there were numerous positive examples of crime analysis treating the geographic (at least regarding mapping; less so for spatial analysis) or the temporal component individually. However, the representation of time on maps was identified as the primary unmet need regarding cartographic representation, with participants indicating the need for cartographic animation in particular. Further, only a single law enforcement agency described a use case scenario that considered both space and time together (the ATAC-ArcGIS desktop mashup). Research on representation, interaction, and analysis techniques that are explicitly spatiotemporal appears to be the most fruitful avenue for crime analysis moving forward, again with a mind towards development of transparently usable interfaces to these spatiotemporal techniques.

(4) Improve support for strategic crime analysis: An emphasis on tactical crime analysis of recent criminal activity, while

understandable from a practical perspective, privileges the victim of the crime, as the goal is to ameliorate the damages incurred quickly and ensure that justice is served. However, strategic crime analysis across longer time periods is needed to better understand the offenders participating in the criminal activity, the second condition under routine activity theory required for the occurrence of a crime (the law enforcement guardian being the third). All participating law enforcement agencies emphasized that it only is through such long-term, strategic spatiotemporal analysis of criminal activity that institutionalized criminal activity may be mitigated and blighted communities may be revitalized. Such an emphasis ultimately requires and reinforces better public safety policymaking and administration as well. Yet, participants noted that resources, tools, and training for strategic spatiotemporal analysis are lacking. In a period during which resources towards law enforcement and public safety are in decline, it is increasingly important to provide law enforcement agencies with crime mapping and analysis tools that are affordable, intuitive, and useful so that these agencies can improve the efficiency of their spatiotemporal crime analysis work and therefore dedicate additional time towards strategic crime analysis.

As stated in the Introduction, the interview study acted as the needs assessment stage for design and development of a spatiotemporal crime mapping application called *GeoVISTA CrimeViz* (http://www.geovista.psu.edu/CrimeViz), a project completed in collaboration between the Penn State GeoVISTA Center and the Harrisburg (PA, USA) Bureau of Police. The key unmet needs identified through the interviews directly informed the conceptual design of the *GeoVISTA CrimeViz* application, providing positive evidence for speaking with the targeted end users about their core needs prior to development. Important design elements of *GeoVISTA CrimeViz* drawn from the interviews include: a web-based architecture for real-time loading of internal and external information sets, persistent interface controls and help documentation to improve the transparent usability of the application for use by all personnel within the Harrisburg Bureau of Police, multiple geographic and temporal representations that are live-linked for coordinated interaction, and contextual geographic information layers and advanced spatiotemporal analyses oriented towards strategic crime analysis. Since the initial needs assessment interview study reported here, we have completed several interface evaluation-refinement loops with the Harrisburg Bureau of Police following a user-centered design approach. The application was transitioned into use by the Harrisburg Bureau of Police for spatiotemporal crime analysis in 2012.

Acknowledgments

References

Andrienko, G. L., & Andrienko, N. V. (1999). Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4), 355–374.
Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Exploratory spatio-temporal visualization: An analytical review. *Journal of Visual Languages and Computing*, 14, 503–541.
Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, 27, 93–115.
Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, 38, 5–22.
Bair, S. (2000). ATAC: A tool for tactical crime analysis. *Crime Mapping News*, 2, 9.
Bertin, J. (1983). Semiology of graphics: Diagrams, networks, maps. Madison, WI: University of Wisconsin Press.
Bertrand, J. T., Brown, J. E., & Ward, V. M. (1992). Techniques for analyzing focus group data. *Evaluation Review*, 16(2), 198–209.
Block, C. R. (1995). *STAC hot spot areas: A statistical tool for law enforcement decisions.* Washington, D.C.: National Institute of Justice.
Boba, R. (2001). *Introductory guide to crime analysis and mapping.* Washington, DC: Office of Community Oriented Policing Services.
Boba, R. (2005). *Crime analysis and crime mapping.* Thousand Oaks, CA: Sage.
Bowers, K. J., Johnson, S. D., & Pease, K. (2004). Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*, 44, 641–658.
Brantingham, P. L., & Brantingham, P. J. (1981). Notes on the geometry of crime. In P. J. Brantingham, & P. L. Brantingham (Eds.), *Environmental Criminology* (pp. 27–54). Prospect Heights, IL: Waveland Press.
Brown, D. (1998). The Regional Crime Analysis Program (ReCAP): A framework for mining data to catch criminals. *Paper presented at the International Conference on Systems, Man, and Cybernetics, San Diego, CA.*
Bruce, C. W. (2008). *Police strategies and tactics: What every analyst shoud know.* International Association of Crime Analysts.
Bruce, C., & Ouellette, N. (2008). Closing the Gap Between Analysis and Response. *The Police Chief*, 75(9), 30–32.
Brunsdon, C., Carcoran, J., & Higgs, G. (2007). Visualising space and time in crime patterns: A comparison of methods. *Computers, Environment and Urban Systems*, 31, 52–75.
Buetow, T., Chaboya, L., O'Toole, C., Cushna, T., Daspit, D., Petersen, T., et al. (2003). A spatio temporal visualizer for law enforcement. *Lecture Notes in Computer Science*, 181–194.
Buja, A., Cook, D., & Swayne, D. F. (1996). Interactive high-dimension data visualization. *Journal of Computational and Graphical Statistics*, 5(1), 78–99.
Cahill, M., & Mulligan, G. (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), 174–193.
Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21, 4–28.
Cheetham, R. (2010). HunchLab: Spatial data mining for intelligence-driven policing. *Paper presented at the Annual Meeting of the Association of American Geographers, Washington, DC.*
Chen, J. (2009). *Exploratory learning from space-attribute aggregated data: A geovisual analytics approach.* (Ph.D.). University Park, PA: The Pennsylvania State University.
Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J., & MacEachren, A. M. (2008). Enhancing spatial scan statistic interpretation with geovisual analytics: An analysis of US cervical cancer mortality. *International Journal of Health Geographics*, 7, 57.
Chung, W., Chen, H., Chaboya, L. G., O'Toole, C. D., & Atabakhsh, H. (2005). Evaluating event visualization: A usability study of COPLINK spatio-temporal visualizer. *International Journal of Human-Computer Studies*, 62, 127–157.
Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44, 588–608.
Conley, J., Gahegan, M., & Macgill, J. (2005). A genetic approach to detecting clusters in point data sets. *Geographic Analysis*, 37, 286–314.
Dent, B. D. (1999). *Cartography: Thematic map design.* Boston, MA: McGraw-Hill.
Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists.* London, England: Routledge.
DiBiase, D., MacEachren, A. M., Krygier, J. B., & Reeves, C. (1992). Animation and the role of map design in scientific visualization. *Cartographic and Geographic Information Systems*, 19((4), 201–214 (265-266).
Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., & Wilson, R. E. (2005). *Mapping Crime: Understanding Hot Spots.* Washington, D.C.: National Institute of Justice.
Eck, J., & Spelman, W. (1987). *Problem-solving: problem-oriented policing in Newport News.* Police Executive Research Forum.
Emig, M. N., Heck, R. O., & Kravitz, M. (1980). *Crime analysis–A selected bibliography.* Rockville, MD: National Institute of Justice.
Getis, A., Drummy, P., Gartin, J., Gorr, W., Harries, K., Rogerson, P., et al. (2000). Geographic information science and crime analysis. *URISA Journal*, 12(2), 7–14.
Goldstein, H. (1979). Improving policing: A problem-oriented approach. *Crime & Delinquency*, 25(2), 236–258.
Goodchild, M. F. (1992). Geographic information science. *International Journal of Geographical Information Science*, 6(1), 31–45.
Gottlieb, S., Arenberg, S., & Singh, R. (1994). *Crime analysis: From first report to final arrest.* Montclair, CA: Alpha Publishing.
Griffith, D. A. (1987). *Spatial Autocorrelation: A Primer.* Washington, D.C.: The Association of American Geographers.
Grubesic, T. H., Mack, E., & Murray, A. T. (2007). Geographic Exclusion: Spatial analysis for evaluating the implications of Megan's Law. *Social Science Computer Review*, 25(2), 143–152.
Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science Association*, 24(1), 155–168.
Harries, K. (1999). *Mapping crime: Principle and practice.* Washington, D.C.: National Institute of Justice, Crime Mapping Research Center.
Hutwagner, L. C., Thompson, W. W., & Seeman, G. M. (2003). The bioterrorism preparedness and response early aberration reporting system. *Journal of Urban Health*, 80(2), 89–96.
Innes, M., Fielding, N., & Cope, N. (2005). 'The appliance of science?': The theory and practice of crime intelligence analysis. *British Journal of Criminology*, 45(1), 39.
Jackson, J. L., & Bekerian, D. A. (1997). Does offender profiling have a role to play? In J. L. Jackson, & D. A. Bekerian (Eds.), *Offender profiling: theory, research and practice.* West Sussex: John Wiley & Sons.
Jefferis, E. S. (1998). *A multi-method exploration of crime hot spots: SaTScan results.* Washington, D.C.: National Institute of Justice, Crime Mapping Research Center.
Klippel, A., Hardisty, F., & Weaver, C. (2009). Star Plots: How shape characteristics influence classification tasks. *Cartography and Geographic Information Science*, 36(2), 149–163.

Krygier, J., & Wood, D. (2005). *Making maps: A visual guide to map design for GIS.* New York, NY, USA: The Guilford Press.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, 26, 1481–1496.

Kulldorff, M. (2010). SaTScan v9.0: Software for the spatial and space-time scan statistics. Information Management Services.

Langran, G. E. (1992). *Time in geographic information systems.* Bristol, PA: Taylor & Francis.

LeBeau, J. L. (2000). *Demonstrating the analytical utility of GIS for police operations.* Rockville, MD: US Department of Justice, National Institute of Justice.

Levine, N. (2006). Crime mapping and the CrimeStat program. *Geographic Analysis*, 38, 41–56.

Lodha, S. K., & Verma, A. (1999). Animations of crime maps using Virtual Reality Modeling Language. *Western Criminology Review*, 1(2).

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science.* West Sussex, England: John Wiley & Sons.

MacEachren, A. M. (1994). Visualization in modern cartography: Setting the agenda. In A. M. MacEachren, & D. R. F. Taylor (Eds.), *Visualization in modern cartography* (pp. 1–12). Oxford, England: Pergamon.

MacEachren, A. M., Brewer, C. A., & Pickle, L. W. (1998). Visualizing georeferenced data: Representing reliability of health statistics. *Environment and Planning A*, 30, 1547–1561.

MacEachren, A. M., & DiBiase, D. (1991). Animated maps of aggregate data: Conceptual and practical problems. *Cartographic and Geographic Information Science*, 18(4), 221–229.

MacEachren, A. M., Wachowicz, M., Edsall, R., Haug, D., & Masters, R. (1999). Constructing knowledge from multivariate spatiotemporal data: Integrating geographical visualization with knowledge discovery in database methods. *International Journal of Geographical Information Science*, 13(4), 311–334.

Maciejewski, R., Rudolph, S., Hafen, R., Abusalah, A., Yakout, M., Ouzzani, M., et al. (2010). A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16(2), 205–220.

Mamalian, C. A., & La Vigne, N. G. (1999). *Research preview: The use of computerized crime mapping by law enforcement: Survey results.* Washington, DC: U.S. Department of Justice, National Institute of Justice.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Monmonier, M. (1990). Strategies for the visualization of geographic time-series data. *Cartographica*, 27(1), 30–45.

Monmonier, M., & Gluck, M. (1994). Focus groups for design improvement in dynamic cartography. *Cartography and Geographic Information Science*, 21(1), 37–47.

Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density esimtation and scan statistics. *Transactions in GIS*, 14(3), 223–239.

Openshaw, S., Charlton, M., Wymer, C., & Craft, A. (1987). A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets. *International Journal of Geographical Information Science*, 1, 335–358.

O'Reilly, T. (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 17.

Osborne, D., & Wernicke, S. (2003). *Introduction to crime analysis: Basic resources for criminal justice practice.* Binghampton, NY: Haworth Press.

O'Shea, T. C., & Nicholls, K. (2003). *Crime analysis in America: Findings and recommendations.* Washington, D.C.: Office of Community Oriented Policing Services, U.S. Department of Justice.

Peuquet, D. J. (1994). It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, 84(3), 441–461.

Ratcliffe, J. (2000). Implementing and integrating crime mapping into a police intelligence environment. *International Journal of Police Science & Management*, 2(4), 313–323.

Ratcliffe, J. H. (2004). The Hotspot Matrix: A framework for the spatio-temporal targeting of crime reduction. *Police Practice and Research*, 5(1), 5–23.

Ratcliffe, J. H. (2009). Crime Mapping: Spatial and temporal challenges. In A. R. Piquero, & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 5–24). New York City, NY: Springer Science.

Ratcliffe, J. H., & McCullagh, M. J. (1998). Aoristic crime analysis. *International Journal of Geographic Information Science*, 12(7), 751–764.

Robinson, A. C. (2008). Collaborative synthesis of visual analytic results. *Paper presented at the Visual Analytics Science and Techology, Columbus, OH.*

Robinson, A. C. (2009). *Needs assessment for the design of information synthesis visual analytics tools.*

Robinson, A. C., Roth, R. E., & MacEachren, A. M. (2011). Designing a web-based learning portal for geographic visualization and analysis in public health. *Health Informatics*, 17, 191–208.

Rose, S. (2008). Community Safety Mapping Online System: Mapping reassurance using survey data. In Chainey, S. Tompson, L. (Eds.), *Crime Mapping Case Studies: Practice and Research* (pp. 93–102). John Wiley & Sons, Ltd.

Rossmo, D. K., & Velarde, L. (2008). Geographic profiling analysis: principles, methods and applications. In S. Chainey, & L. Tompson (Eds.), *Crime Mapping Case Studies: Practice and Research* (pp. 35–43). John Wiley & Sons, Ltd.

Roth, R. E. (2009). A qualitative approach to understanding the role of geographic information uncertainty during decision making. *Cartographic and Geographic Information Science*, 36(4), 315–330.

Roth, R. E. (2010). Dot density maps The Encyclopedia of Geography. Thousand Oaks, CA: SAGE, 787–790.

Roth, R. E. (2011). *Interacting with Maps: The science and practice of cartographic interaction.* (PhD). University Park: The Pennsylvania State University.

Roth, R. E. (2012). Cartographic interaction primitives: Framework and synthesis. *The Cartographic Journal*, 49(4), 376–395.

Roth, R. E., Robinson, A., Stryker, M., MacEachren, A. M., Lengerich, E. J., & Koua, E. (2008). Web-based geovisualizaiton and geocollaboration: Applications to public health. *Paper presented at the Joint Statistical Meeting, Invited Session on Web Mapping, Denver, CO.*

Roth, R. E., & Ross, K. S. (2009). Extending the Google Maps API for event animation mashups. *Cartographic Perspectives*, 64, 21–40.

Roth, R. E., Ross, K. S., Finch, B. G., Luo, W., & MacEachren, A. M. (2010). A user-centered approach for designing and developing spatiotemporal crime analysis tools. *Paper presented at GIScience 2010, Zurich, Switzerland.*

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology*, 94(4), 774–802.

Shaw, C., & McKay, H. (1942). *Juvenile deliquency and urban areas.* Chicago, IL: University of Chicago Press.

Sinton, D. (1978). The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. *Paper presented at the Harvard papers in GIS #7, Cambridge, Massachusetts.*

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2005). *Thematic cartography and geographic visualization* (2nd ed.). Upper Saddle River, NJ, USA: Pearson Prentice Hall.

Spelman, W. (1995). Criminal careers of public places. In J. E. Eck, & D. Weisburd (Eds.), *Crime and Place.* Monsey, NY: Criminal Justice Press.

Sutherland, E. (1934). *Principles of Criminology* (2nd ed.). Philadelphia, PA: J.B. Lippincott.

Sutherland, E. H., Cressey, D. R., & Luckenbill, D. F. (Eds.). (1992). *Principles of Criminology* (11th ed.). Landhan, MD: AltaMira Press.

Townsley, M. (2008). Visualising space time patterns in crime: The Hotspot Plot. *Crime Patterns and Analysis*, 1(1), 61–74.

Tufte, E. (1983). *The visual display of quantitative information* (2nd ed.). Cheshire, Connecticut: Graphics Press LLC.

Wallace, T. R. (2011). A new map sign typology for the GeoWeb. *Paper presented at the International Cartographic Conference, Paris, France.*

Walsh, W. (2001). COMPSTAT: An analysis of an emerging police managerial paradigm. *Policing: An International Journal of Police Strategies & Management*, 24(3), 347–362.

Weisburd, D., Mastrofski, S., McNally, A., & Greenspan, R. (2002). Reforming to preserve: COMPSTAT and strategic problem solving in American policing. *Criminology & Public Policy*, 2, 421–456.

White, J. J. D., & Roth, R. E. (2010). TwitterHitter: Geovisual analytics for harvesting insight from volunteered geographic information. Paper presented at GIScience 2010, Zurich, Switzerland.

Wilcox, P., Land, K. C., & Hunt, S. (2003). *Criminal circumstance: A dynamic multi-contextual criminal opportunity theory.* New York, NY: Walter de Gruyter.

Wilson, R. E. (2007). The impact of software on crime mapping: An introduction to a special journal issue of Social Science Computing Review on crime mapping. *Social Science Computer Review*, 25(2), 135–142.

Wilson, O. W., & McLaren, R. C. (1977). *Police Adminstration* (4th ed.). New York, NY: McGraw-Hill.

Wolff, M., & Asche, H. (2009). Geovisualization approaches for spatio-temporal crime scene analysis - Towards 4D crime mapping. *Lecture Notes in Computer Science*, 5718, 78–89.

Yildiz, M. (2007). E-Government research: Reviewing the literature, limitations, and ways forward. *Government Information Quarterly*, 24(3), 646–665.

Zeng, D., Chang, W., & Chen, H. (2004). A comparative study of spatio-temporal hotspot analysis techniques in security informatics. *Paper presented at the Intelligent Transportation Systems Conference, Washington, DC.*

Zhao, J. L., Bi, H. H., Chen, H., Zeng, D. D., Lin, C., & Chau, M. (2006). Process-driven collaboration support for intra-agency crime analysis. *Decision Support Systems*, 41, 616–633.

Robert E. Roth is an Assistant Professor of Geography in the University of Wisconsin–Madison, Department of Geography. Additional information about Roth's teaching and research interest can be viewed at http://www.geography.wisc.edu/faculty/roth.

Kevin S. Ross is a Systems Design Architect at Nsite LLC. Additional information about Ross can be viewed at http://ksross.com.

Benjamin G. Finch is an undergraduate student majoring in Geography, with a specialization in GIScience, and a student intern in the Penn State GeoVISTA Center.

Wei Luo is a PhD candidate in the Penn State Department of Geography and a researcher in the Penn State GeoVISTA Center. Additional information about Luo can be viewed at http://www.personal.psu.edu/wul132/.

Alan M. MacEachren directs the GeoVISTA Center (http://www.geovista.psu.edu) and is a Professor of Geography and Affiliate Professor of Information Sciences and Technology at Penn State University. For details, see http://www.geovista.psu.edu/members/maceachren/.
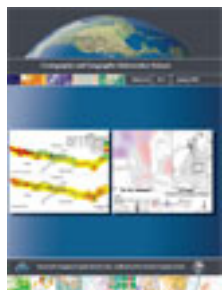
Cartography and Geographic Information Science

Symbol Store: sharing map symbols for emergency management

Anthony C. Robinson [a] , Scott Pezanowski [a] , Sarah Troedson [a] , Raechel Bianchetti [a]
, Justine Blanford [a] , Joshua Stevens [a] , Elaine Guidero [a] , Robert E. Roth [b] & Alan M. MacEachren [a]

[a] Department of Geography, GeoVISTA Center , The Pennsylvania State University , University Park , PA , 16802 , USA

[b] Department of Geography , University of Wisconsin-Madison , Madison , WI , 53706 , USA
Published online: 03 Jun 2013.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Symbol Store: sharing map symbols for emergency management

Anthony C. Robinson[a]*, Scott Pezanowski[a], Sarah Troedson[a], Raechel Bianchetti[a], Justine Blanford[a], Joshua Stevens[a], Elaine Guidero[a], Robert E. Roth[b] and Alan M. MacEachren[a]

[a]*Department of Geography, GeoVISTA Center, The Pennsylvania State University, University Park, PA 16802, USA;* [b]*Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA*

Maps are a primary means for supporting information sharing and collaboration in emergency management and crisis situations. While a variety of formalized map symbol standards for emergency contexts exist, they have not been widely adopted by mapmakers. Informal symbol conventions are commonly used within emergency management stakeholder groups, but until now there has not been a flexible mechanism for discovering, sharing, and previewing these symbol sets among mapmakers. In this paper, we describe the design and development of the Symbol Store, a visually enabled, web-based interactive tool intended to help mapmakers share point symbols. The Symbol Store allows users to browse symbols by keyword, category tags, and contributors. It also allows for symbols to be previewed on realistic maps prior to download. An initial prototype of the Symbol Store was evaluated by flood mapping experts from the State of California, and the results of this user study led to multiple refinements now implemented in the public version of Symbol Store located at www.symbolstore.org.

Keywords: map symbology; emergency management; web applications; standards

1. Introduction

Maps offer a critical form of communication and function as important analytical tools for distributing information in emergencies when it becomes necessary for stakeholders to develop a common operational picture (COP). Maps are also essential in other phases of emergency management that include preparedness, recovery, and mitigation activities. During emergencies, all tasks are time sensitive, and the speed with which one can read and interpret a map can signal the difference between lives and property saved or lost. Supporting rapid and efficient consumption of geographic data can be significantly assisted by having map design and symbology standards that users can learn and apply in advance of an emergency event. Additionally, when teams of collaborators from different organizations come together in an emergency operations center (EOC), they can quickly learn how to read and interpret maps by learning from an existing map design and symbol standard.

This paper describes the design, development, and initial evaluation of a new, web-based mechanism for searching for and retrieving point symbols to support digital cartography. We present a novel system called the Symbol Store, based on a long-term, iterative study of mapmakers' symbol needs at the US Department of Homeland Security (DHS) (see previous work in Robinson, Roth, and MacEachren, 2010, 2011; Robinson et al. 2012). The Symbol Store can help aid interoperability in emergency management situations where multiple local, state, and federal authorities collaborate on the

development of a wide range of mapping products to develop situational awareness and marshal resources. Emergency management mapping tasks frequently include pre-disaster planning activities and strategic recovery efforts which can also stand to benefit from mechanisms that improve symbol interoperability. Currently, there are few elegant or efficient means for such stakeholders to access and share each other's methods for representing features on maps.

In the following sections, we describe relevant prior work on symbol standardization and previous means for sharing map symbology. Then we describe the key design goals and features of the Symbol Store. This is followed by a section describing an initial user evaluation conducted with flood mapping experts to test core features of the Symbol Store. Finally, we conclude with planned next steps to further refine and exploit the unique capabilities of the Symbol Store.

2. Background

Our focus in this research is on sharing point symbols for maps. While line and polygon features are also quite important, we have focused on symbol interoperability first with respect to point features, which are among the most common types of symbols used in emergency management tasks. These tasks can include disaster mitigation and response planning activities, as well as direct response and situational assessment mapping in the immediate aftermath of a disaster (Cutter 2003). They can also include long-term planning and

---

*Corresponding author. Email: arobinson@psu.edu

remediation mapping efforts as communities engage in recovery efforts. As we discuss below, the majority of existing symbol standards focus on point symbols entirely or in large part. In addition, funding support for this research from DHS was directed specifically toward the task of exploring point symbol interoperability since it represents their most pressing area of need.

Two primary influences for our work include past efforts to develop map symbol standards and existing approaches for sharing map symbols.

2.1. Symbol standards

In recent decades there have been a wide range of symbol standardization efforts that have resulted in the design and dissemination of point symbol references for use in common mapping contexts. These efforts include the development of the US Military Standard MIL-STD-2525 (Department of Defense 2008), which prescribes a set of representations and modifiers to support map interoperability in military planning and combat situations. As a result of its adoption by NATO, this standard is used and modified by other stakeholders as well in a wide range of countries and contexts. For example, NATO countries frequently support humanitarian relief missions with military-led assistance, and non-governmental agencies engaged in relief efforts will often make use of the same symbology to ensure interoperability.

Symbol standards for non-military purposes include the US ANSI INCITS 41-2006 (ANSI 2006) Homeland Security Mapping Standard which prescribes a set of point symbols, symbol outlines, and graphical modifiers to support emergency management interoperability. The ANSI set was designed to function as a single standard used by local, state, and federal stakeholders in US domestic emergency management situations. Despite concerted efforts to develop the ANSI standard for point-symbol symbology for exclusive use across DHS and with DHS partners (Dymon 2003; Dymon and Mbobi 2005), the standard has not achieved widespread adoption within DHS, or outside DHS in commercial tools like WebEOC (www.esi911.com) that are commonly used by state and local emergency management organizations. In previous research, we found that DHS mapmakers had symbolization needs that were unmet by the ANSI set, that many ANSI symbols were too graphically intricate to use in many situations, and that task-specific (but non-official) symbol sets already existed that were more commonly used to support emergency management mapmaking (Robinson, Roth, and MacEachren, 2011). Furthermore, recent usability and utility evaluation of these symbols revealed that they have serious weaknesses when they are used to support emergency management tasks (Akella 2009).

Other symbol standards exist for emergency mapping and related contexts, including an international standard for humanitarian demining symbols (Kostelnick et al. 2008) and a recent homeland security symbol standardization project in Canada (Sondheim, Charmley, and Leeming 2010), which led to the development of a web repository (www.EMSymbology.org) where those symbols could be downloaded. Despite the availability of these official and many non-official but widely utilized symbol sets (such as those included with common Geographic Information Systems (GIS) software packages) no single symbol standard has been widely adopted for general use across the full range of emergency situations in the US or elsewhere.

One potential reason for lack of widespread adoption of symbol standards is that their development tends to receive a great deal of attention and resources, while their dissemination, implementation, and revision does not. Emerging needs for new and modified representations are not easily met through official standardization processes. In discussing the need for symbol standardization in the 1970s, Arthur Robinson suggested that while standards are necessary and useful, they must remain open-ended so that evolution can occur as mapping requirements change (Robinson 1973). New efforts to design symbols by MapBox (www.mapbox.com/maki) and the Noun Project (www.thenounproject.com) are responding to evolutionary needs to develop simple point-of-interest symbols for web maps for the former project, and to develop a common visual language to cross language barriers for the latter. The Map Icons Collection (www.mapicons.nicolasmollet.com) allows users to contribute point symbol designs for use in Google Map mashups. New approaches for supporting symbol sharing and interoperability need to be flexible enough to support new symbol sets like these and integrate them with existing methods to ensure the best representations are used for any given mapping context.

2.2. Current methods for sharing symbols

Map symbols are most commonly shared today via digital means through common GIS software packages. While mapmakers are always able to create their own symbols, the default palettes available in GIS software see wide adoption, although the original meaning of a particular symbol may be ignored in part or whole as mapmakers reinterpret symbols for use in new representational contexts (Robinson, Roth, and MacEachren, 2011). In terms of their formal implementation, outside of the MIL-STD-2525 example noted above, where military units require its use and mapmakers as well as map readers are trained to use the standard, the other standards we reviewed may have formal endorsement but no firm requirements for mapmakers to actually use them.

In terms of tools for sharing symbols, Esri's ArcGIS supports the use of *.style files which can catalog and describe the representations used on maps in ArcGIS. We found when studying the use of ANSI symbols and other symbology at DHS that DHS mapmakers frequently made use of *.style files to curate their own task-specific (but not official) symbol collections, and in some cases would share these with other

mapmakers. Recent development by the makers of Ortelius, a cartography software package for use in Apple iOS, includes the ability to share Ortelius-compatible map symbols online via a dedicated symbol management tool (Saligoe-Simmel 2010). Most officially sanctioned map symbol standards are disseminated by government websites and shared in Esri-compatible formats. Our work with mapmakers at DHS indicated that the primary means for discovering symbols was by using web search engines to find new *.style files and fonts that include symbol markers.

We began our work on map symbology for emergency management in 2009 by researching the use and adoption of the ANSI symbol standard, identifying other symbols in use, and developing key user requirements for map symbology standards (Robinson, Roth, and MacEachren, 2011) at DHS. We followed this work with the development and evaluation of a new, more flexible and iterative process for creating map symbol standards for DHS (Robinson et al. 2012). A primary result discovered through this prior research was the need for a new platform and model for supporting symbol interoperability. While current tools like ArcGIS and Ortelius allow users to share pre-defined symbol sets using Esri *.style files or other special formats, and web repositories like the Noun Project and EMSymbology.org exist for individual symbol sets there remains a need for flexible, visually enabled tools to support dynamic symbol sharing and discovery. We believe the next step involves moving beyond simply retrieving symbols via the web, to support users who wish to contribute symbols to an evolving repository, to support users who want to search for appropriate symbols using keywords and other metadata (something not supported on sites like www.EMSymbology.org for example), to preview symbols on realistic maps, and to allow communities of mapmakers to iteratively rate and refine symbol collections to create new *de facto* map symbol standards. To fill this gap, we have developed the web-based Symbol Store to provide a usable and useful tool for contributing, browsing, rating, and assembling customized symbol palettes to support mapmakers at DHS and beyond. The following sections describe our progress so far in meeting that goal.

3. Symbol store

To support map symbol interoperability we have designed a web-based prototype tool for discovering, sharing, and retrieving map symbols called the Symbol Store. The following sections describe Symbol Store's core design objectives, its technical underpinnings, our demonstrated progress toward implementing a working prototype, and the results of a user evaluation to test the first working prototype to suggest future refinements.

3.1. Design objectives

Based on the results of our prior work to study the utility of the ANSI standard for DHS mapmakers and to design a new process to developing more flexible symbol standards, we developed four core system functionality targets for shaping the design of our first Symbol Store prototype:

3.1.1. Search for and retrieve symbols

The most basic design goal for Symbol Store is to support keyword searches for symbols in use by agencies across DHS (and potentially wider audiences as the tool becomes open to other groups as well). Symbols retrieved via keyword search can be collected and downloaded as an ESRI *.style file or in other common formats for immediate application.

3.1.2. Preview symbols on realistic maps

After selecting a subset of symbols from the Symbol Store search interface, users can preview their symbols on a variety of realistic base maps. The map preview feature in Symbol Store provides a range of basic design controls to allow users to change the map scale, feature density, labeling, coloring, and other common map design aspects in order to preview the suitability of their symbols prior to downloading them.

3.1.3. Browse for symbols

Apart from searching for specific symbols, users can browse symbols by time (most recent uploads, for example), contributor (symbols from a specific agency, for example), and symbol categories (all symbols corresponding to infrastructure, for example). This supports flexible means for discovery, for instances in which keyword searches are not as efficient or effective. It also allows popularity measures such as subjective ratings to become one of the means through which new symbols can be discovered and disseminated.

3.1.4. Share symbols

Users can contribute symbols to Symbol Store by uploading an Esri *.style file and associated fonts through the Symbol Store interface. After uploading symbols, users are able to tag individual symbols or groups of symbols to assign keywords, category names, and other important metadata information.

3.2. System architecture

To accomplish our design objectives we designed an interactive web interface to encourage use by a wide range of users across multiple platforms. Simple and effective interoperability is essential to support DHS users coming from a diverse set of organizations, and to ensure more

widespread adoption of formal standards by other related emergency management communities in the longer term.

The Symbol Store interface (Figure 1) runs in a standard web browser using the Adobe Flash plugin. The interface itself was constructed using Adobe Flash Catalyst, an interface development environment that integrates with Adobe Flex, which we used to connect the Symbol Store interface to server-side components. Symbol Store is comprised of four main components, illustrated in Figure 2 and includes: 1) the Flex and ActionScript User Interface (UI); 2) the .NET CSharp web service middleware; 3) the storage system of an Apache Lucene Index (apache.lucene.org) and a Structured Query Language (SQL) Server Relational Database Management System (DB); 4) and an instance of Esri ArcGIS Server to produce live, interactive map previews with selected symbols.

A Lucene index is used to store the text metadata about symbols in the Symbol Store, including Symbol name, symbol description, keywords, user, symbol set, and other features. This allows for text searches to be performed when searching for information. When a text based search is performed, the Lucene index is queried and pointers to the symbols are returned. Next the DB is
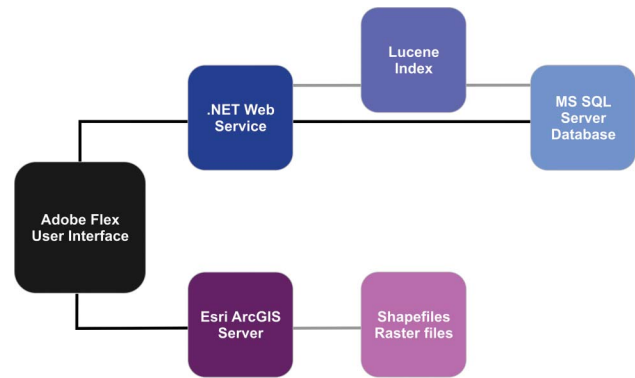
Figure 2. Architecture of Symbol Store. Symbols with metadata (e.g., keywords, set, categories, date of upload, ratings) are stored in a database and indexed using the Lucene Index. Map previews are stored and served using Esri ArcGIS Server 10.

queried to retrieve those symbols that matched the Lucene index query. Users are then able to create a customized symbol set by selecting individual symbols, adding them to their symbol cart and downloading the newly created style file containing the symbols.
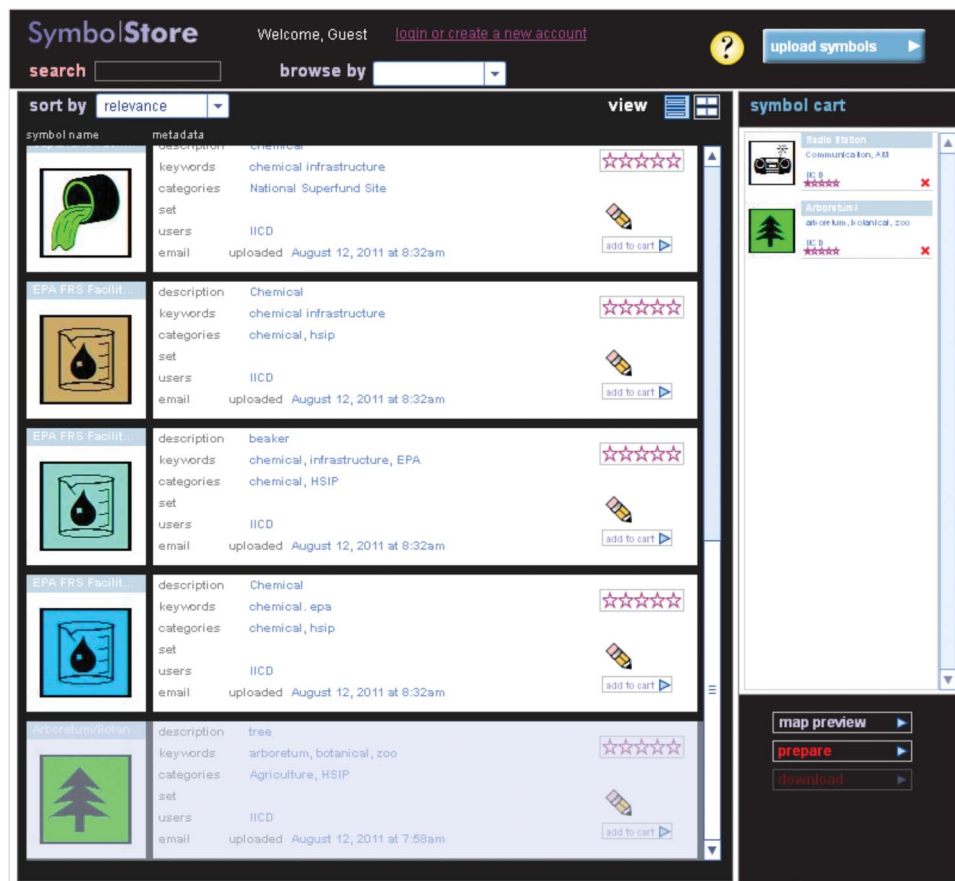
Figure 1. This screen capture shows the initial prototype Symbol Store interface with symbols contributed from Infrastructure Information Collection Division (IICD).

In addition to obtaining symbols to match their needs, users can contribute symbols to the store by uploading either Esri *.style files or TrueType font files. When these files are uploaded, symbols are stored in a SQL Server Database table record and an image thumbnail of the symbol is created using Esri's ArcObjects tools. Once symbols have been uploaded to the system and stored in the database, the user is presented with a metadata editing interface where they have the option to add metadata for the symbol collection as a whole and/or for individual symbols. Symbols are easily retrieved through an SQL query when the user decides to download symbols from their symbol cart.

The final component of the Symbol Store allows for the preview of symbols on a map. The Symbol Store map preview tool allows users to preview the symbols that they have placed in their shopping cart. The map preview has been crafted to allow users to choose a base map style and scale, and then adjust many of the symbol properties, such as size, density, and labeling. This allows users to compare different potential symbol sets and choose the set that is the most legible given their choice in cartographic design. Three main control types have been designed to support on-the-fly visualization of symbol properties. The base map type can be chosen using a button feature. Horizontal sliders are used to adjust for symbol size, map scale, point density, and label size. Finally, a user's choice in labeling or not labeling the features uses a checkbox control.

The base map is comprised of several ArcGIS shapefiles representing transportation, political boundaries, and water bodies. The features were acquired from the US National Atlas for national and regional scales, and several state-level agencies for the state, county, and city level data. For state-level data, we chose Washington State as the example, and have used data from the Washington Department of Transportation, the Washington Department of Environmental Quality, and King County's GIS Center. To display map point symbols, we have created several levels of point density for each of the scales by generating random point symbols.

3.3. Symbol store prototype features

The initial Symbol Store prototype underwent multiple rounds of internal evaluation and refinement to satisfy our stated goals to support symbol search and retrieval, symbol browsing, symbol sharing, and preview symbols on realistic maps.

The initial prototype Symbol Store search interface is shown in Figure 2. Each symbol includes metadata to show its description, a list of relevant keywords, relevant thematic categories, membership in a formal symbol standard (e.g., ANSI), users of this symbol (e.g., different divisions within DHS), contact information for the person responsible for uploading the symbol, and the date the symbol was uploaded. Users can select symbols for download or preview by adding them to the cart shown on the right of the screen in Figure 2. Once users are satisfied with the selected set of symbols, they can prepare and download the set as an Esri *.style file.

Symbols can be easily contributed to the initial prototype Symbol Store by way of a specialized upload and metadata creation interface we have implemented (Figure 3). Users can select an Esri *.style file and associated font files and upload these to the Symbol Store database. Once the Symbol Store has processed these files, a secondary interface appears (shown at the bottom of Figure 3) so that users can add metadata to the contributed symbols. For the entire collection, users can edit the contributing agency, assign symbol categories associated with the set, identify the username of the person contributing the set, and name the official symbol standard to which the symbols belong (if applicable). For individual symbols (or smaller groups of multiple symbols, selected by clicking checkboxes in the metadata editing interface), users can add descriptions, keywords, categories, users (such as agencies or departments), and ratings (from 1 star to 5 stars). Users can also edit symbol metadata once the symbols have been published in the Symbol Store.

The map preview feature (Figure 4), inspired by popular web tools for cartographic design like ColorBrewer (Harrower and Brewer 2003) and TypeBrewer (Sheesley 2007), which provide example maps to help users choose among design options, will allow users to visually evaluate the symbols that are currently in their symbol cart in a realistic map design context. The map preview tool in Symbol Store allows users to change the type of map background between common base map options, to change the color of the base map between color or black and white, and to adjust symbol sizes, the map scale, feature density, and labeling. These common design parameters are adjustable interactively to help users rapidly evaluate the symbols they have chosen against a variety of realistic map design constraints. One goal here is to decrease the amount of time mapmakers spend creating multiple design iterations to develop a particular map.

4. Evaluation

In addition to multiple rounds of internal refinement of the Symbol Store prototype, including substantial informal feedback from DHS stakeholders in monthly project meetings, we have completed an initial formative user evaluation to inform our next steps.

We recruited six mapmakers from California's Department of Water Resources (DWR) who regularly engage in emergency management mapping centered on floods and other water-related emergency events. DWR staff are also engaged in an effort to develop an internal standard for symbology, so this user group would potentially benefit from a tool like the Symbol Store when identifying candidate symbols from other sets. Our goals in this study

Figure 3.   The Symbol Store symbol upload and metadata editing interface.

were to evaluate the utility of the Symbol Store for daily DWR mapmaking needs, to help develop their own standard set of symbols, and to characterize how well the Symbol Store could support symbol sharing for DWR once they have created a standard that could be shared.

4.1.   Evaluation procedures

Our evaluation activities were composed of a task analysis to test key functions in the Symbol Store and an online survey to capture and elicit user feedback. We developed six tasks for DWR mapmakers to complete. In general terms, the six tasks included: uploading symbols, searching for symbols, selecting symbols to preview, previewing symbols on realistic maps, downloading symbols, and testing a downloaded Esri *.style file. After completing these

tasks with the Symbol Store prototype, users were asked to rate the usability and utility of the Symbol Store in an online survey. Many of the survey questions also prompted users to provide qualitative feedback to elaborate upon their opinions and suggest specific new features or bugs to fix.

This user study was aimed at formative assessment (Buttenfield 1999) of the Symbol Store prototype in order to establish which steps to take going forward to add/remove key features and to develop a baseline understanding of the tool's usability.

4.2.   Task analysis results

The following sections describe each of the six tasks our study participants completed and discuss survey feedback gained from questions related to each task. All
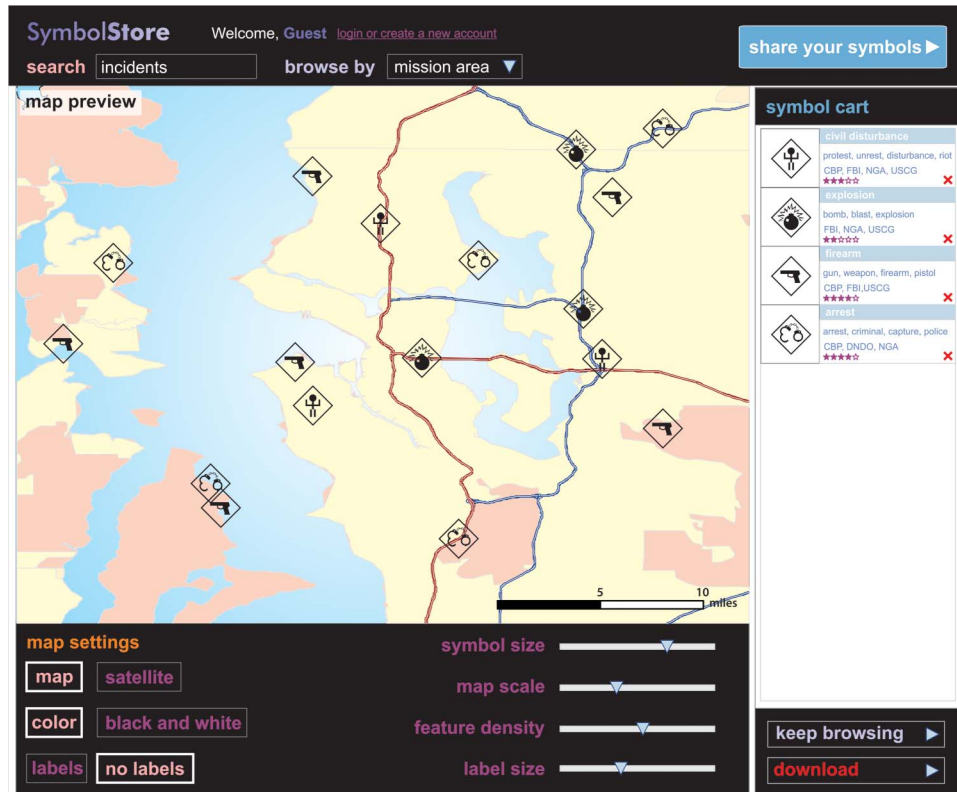
Figure 4. The map preview function (shown here in a design mockup) in the Symbol Store allows users to select from different map types and to control other symbol and map design attributes to preview symbols placed in the symbol cart.

survey questions aside from two multiple choice prompts used a five-point agreement scale; strongly disagree – 1, disagree – 2, neither agree nor disagree – 3, agree – 4, and strongly agree – 5. Average ratings are used in the following sections according to this scale to note participant agreement level with prompts about the usability and utility of features and functions in the Symbol Store (Figure 5). All questions were crafted in a manner to allow participants to agree or disagree with the premise of the prompt. A reference document showing the full task descriptions as well as the complete set of survey questions is available in Appendix A (www.personal.psu.edu/acr181/SymbolStore_Appendix_A.pdf).

4.2.1. Task one: upload symbols

Participants were asked to upload an Esri *.style file containing eight point symbols and to create metadata for those symbols by tagging them with keywords, categories, and other relevant source information. When asked to rate how easy it was to upload a .style file, the average of user
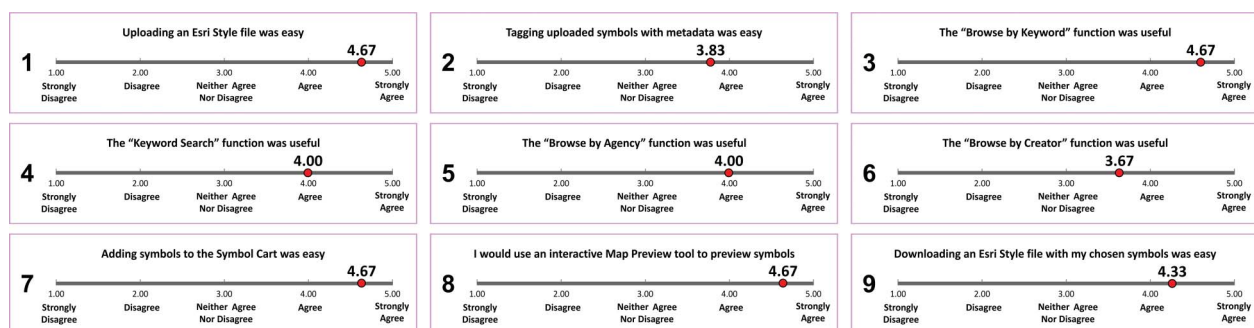
Figure 5. User rating results for key aspects of Symbol Store usability and utility.

ratings was 4.67 (Q1 in Figure 5) on the five-point agreement scale described above, indicating that users found this task easy to accomplish.

While participants generally agreed that the process of adding metadata to their symbols was easy (3.83 rating – Q2 in Figure 5), many suggestions on how to improve this part of the interface were offered. The names assigned to each symbol were not visible during the tagging process which led to much confusion over what each symbol was intended to be used for. Also, the text fields intended to contain the keywords and categories were not large enough. Other comments mentioned the need to apply changes in metadata to multiple symbols at one time (when assigning category names, for example).

4.2.2.   Task two: search for symbols

In the second task, participants were asked to search for the symbols to which they had just added metadata via both keyword search and basic browsing methods. In addition to typing search terms into the keyword search box, participants used the "Browse By" option to select from a list of categories and keywords, and they could also search by clicking on keywords associated with a symbol in the symbol list itself. Additionally, during this task, participants were given the option to update or change metadata they had previously entered for any given symbol. Our users found the "Browse By" option to explore existing keywords most useful (4.67 rating), followed by traditional keyword search (4.0 rating). Browsing by agency (4.0 rating) and creator (3.67 rating) metadata was also judged to be useful (Q3–Q6 in Figure 5).

Qualitative feedback from this task revealed that the most common concern with the search and browse options was there was no obvious way to clear a search and return to a previous view. Additionally, there was no way to show an overview of all symbols. Some study participants also found a bug that caused keywords and categories in the list to appear without associated symbols. A small interface design issue regarding the size of the symbol name was also highlighted by several users.

4.2.3.   Task three: select symbols to preview

In the third task, participants were asked to select approximately a dozen symbols and add them to their Symbol Cart to preview prior to download. This was a very simple task that required clicking icons on symbols of interest to add it to a "shopping cart" style interface. This feature's usability received strong support (4.67 rating, Q7 in Figure 5) and none of the participants provided comments or suggestions to improve this part of the interface.

4.2.4.   Task four: preview symbols

Our intention for the fourth task was to have participants evaluate the map preview feature to self-assess a set of

symbols added to the symbol cart. The development work to enable this feature was not completed in time for the user study, and instead our participants were given a mockup preview screen (shown in Figure 4) and asked to comment on the likelihood that they would use such a tool (4.67 rating, Q8 in Figure 5), and to rate the relative importance of each proposed map preview function.

Participants were unanimously in favor of all of the proposed functions suggested by the mockup concept. One specific additional suggestion was offered during the focus group discussion, which is to add the option to preview their symbols over a United States Geological Survey (USGS) Topographic base map.

4.2.5.   Task five: download symbols

Once study participants had chosen their symbols they were asked to download them. The evaluated prototype supported downloads using an Esri *.style file, which is one of several common means for formalizing the look of a digital map product. This task was very straightforward and widely viewed as easy (4.33 rating, Q9 in Figure 5) and the only suggestion from participants was to implement a way to combine the "Prepare" and "Download" steps to make downloading possible with a single click.

4.2.6.   Task six: test downloaded *.style file

The final task involved launching a new Esri ArcGIS project, adding the downloaded Esri *.style file to the map document and viewing the point symbols in the ArcGIS Style Manager or assigning those symbols to point features in the map document.

One study participant had trouble downloading the file on the first try (their subsequent attempt was successful), and another study participant found two of their chosen symbols did not download properly. The latter error is a known issue associated with having all of the necessary fonts installed locally where the *.style file is downloaded. A large proportion of symbols in Esri ArcGIS are drawn directly from custom fonts that use symbols in place of alphanumeric characters. The *.style file then connects to these fonts and draws them in specific ways based on chosen design attributes. In this case, the study participant was missing the font necessary to view two of their downloaded symbols. All other downloaded symbols worked properly for all of the study participants.

An improvement that can be made to the downloaded *. style file was suggested to incorporate the keywords associated with a symbol in the Symbol Store into the "tag" field associated with the *.style file so that those imported symbols can then be searched for using the ArcGIS symbol browsing tool. The evaluated Symbol Store prototype downloaded *.style file only contained the symbol name and Esri-software assigned default design

attributes present when the symbol was originally uploaded to the Symbol Store. Study participants expressed concern that they spent a good deal of time adding keywords and metadata to symbols and then were not able to use that information once the *.style file was downloaded to their local machine.

4.3. Evaluation summary

Our evaluation of the initial Symbol Store prototype with flood mapping users from the state of California yielded useful feedback on usability and utility. Approximately 20 small interface improvements were identified by the study participants, and a variety of major improvements were suggested:

- Implement the map preview as designed in the mockup
- Improve search behavior to retrieve additional relevant results beyond exact matching keywords
- Improve the interface look and feel to make functionality easier to find and visually distinct from search results
- Support a wider range of export formats to avoid problems with fonts
- Use tabbed pages to view symbol results incrementally
- Add a USGS topographic base map option to the map preview tool
- Develop a grid view to show a larger overview of available symbols when browsing
- Support import and retrieval of line and polygon symbols
- Implement user accounts so that draft symbol standards can be shared among a working group before being released to a wider audience
- Import/export more metadata when contributing or downloading *.style files

Overall, our evaluation results with this small group of real-world users demonstrate the potential of a tool like the Symbol Store to support usable and useful symbol discovery, retrieval, and contribution. User feedback after the survey indicated that our participants were eager to see future revisions to the prototype and for the tool to be robust enough to support regular, widespread use in their agency.

5. SymbolStore.org

Following the results of our initial evaluation with flood mapping users in California, we implemented many of their suggested changes and began transitioning the prototype Symbol Store to a public-facing site for widespread use. The Symbol Store is now available at www.symbol-store.org, and currently hosts over 2400 symbols that can be easily discovered, previewed on realistic maps, and downloaded in a variety of useful formats. Some of the major improvements made from the evaluated prototype include; a fully functioning map preview tool, a redesigned user interface to improve clarity and offer a standardized look and feel, pagination of search results to improve usability, and two new export formats (PNG and SVG) to avoid problems with sharing fonts and improve interoperability.

The improved primary interface is shown in Figure 6. In addition to the major improvements already listed, we fixed the bug that caused keywords and categories to appear that did not link to search results, and we have improved connections to Lucene to support more flexible search behavior to retrieve more results with single keywords. For example, a search for "fire" will now return anything that includes the stem of that term, so "firing range" will appear as a result rather than only exact matches for "fire."

The map preview tool (Figure 7) now allows users to interactively assess the symbols in their symbol cart using a range of common map design controls to change the base map design, alter the size/density of symbols on the map, and explore the symbols when used at three common scales. The map preview tool leverages ArcGIS Server to generate and manipulate real-time map previews on the web client.

A major step toward supporting wider symbol interoperability is the inclusion of new symbol formats with every Symbol Store download. Users now can retrieve a single .zip file archive which includes PNG symbols at a range of useful icon sizes, an Esri *.style file, and SVG vector graphics for use in graphic design software. SVG symbol export is made possible by tracing PNG images of symbols using an automated back-end routine that leverages the Inkscape open source graphic design software (www.inkscape.org).

Symbol contributions are supported in the public www.symbolstore.org site through a new, simplified interface shown in Figure 8. Currently, we do not allow public contributions to be processed and appear automatically to prevent potential abuse, and we are exploring ways to support metadata creation and editing for public users while ensuring that contribution quality will remain high.

6. Conclusions and future work

Following our initial evaluation and refinement effort, we will focus on collecting feedback on the second-generation prototype available at www.symbolstore.org from DHS mapmakers through a series of planned practitioner workshops. In these workshops, DHS users will complete realistic symbol-related tasks using the Symbol Store and we
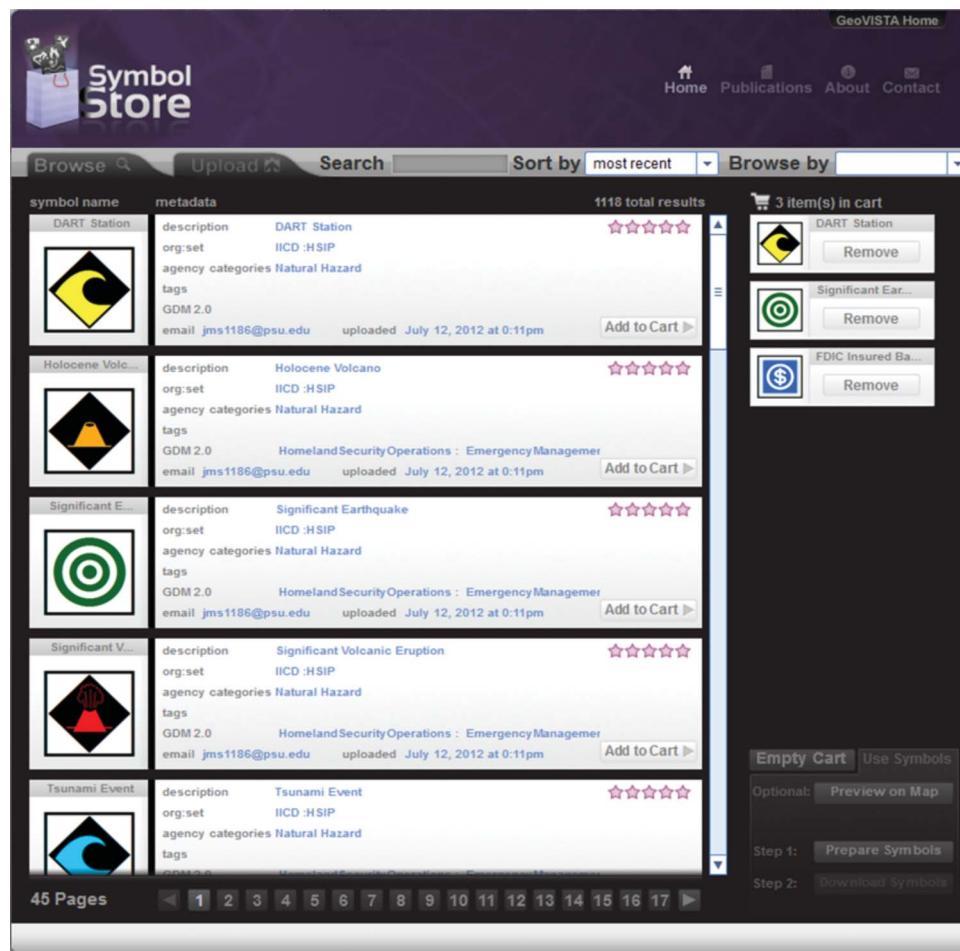
Figure 6.   The redesigned Symbol Store interface is now available for public use at www.symbolstore.org.

will engage these users in group discussions to identify next steps for Symbol Store development and integration into existing DHS work practices. We anticipate that the Symbol Store will be useful for a wide range of emergency management related tasks, including pre-emergency planning, post-disaster situational assessment mapping, and long-term recovery efforts. In terms of its immediate utility during the response phase of an emergency, we would anticipate that the Symbol Store would be a quicker way to discover, download, and use symbols than their current affordances, which rely on manual web searches and informal relationships with other mapmakers (Robinson, Roth, and MacEachren, 2011) allow.

A key focus for Symbol Store development going forward will be on integrating what we have learned from symbol standard development through the standardization process we developed and the e-Symbology Portal tools we used to conduct process tests with Customs and Border Protection (CBP), Federal Emergency Management Agency (FEMA), and Infrastructure Information Collection Division

(IICD) (Robinson et al. 2012). We will develop methods and techniques to move components from our standardization process into the Symbol Store interface to combine the two efforts in an elegant and effective unified environment. A key goal is to provide access to more sophisticated metadata and category standardization tools and procedures for small groups of motivated users and symbol set curators. These higher-level functions will require accounts and log-in permissions so as not to interfere with basic use by members of the public and mapmakers who simply want to quickly find and retrieve symbols.

Other challenges for future development include new methods to expand searches to return relevant results. One strategy we are currently implementing is to leverage WordNet (Miller 1995) measures of similarity between words in the English language to find relevant terms beyond an initial keyword and retrieve a wider set of relevant symbols to the user. Searching for symbols by visual similarity also remains an important, but difficult to achieve goal. Ideally, cartographers should be able to find
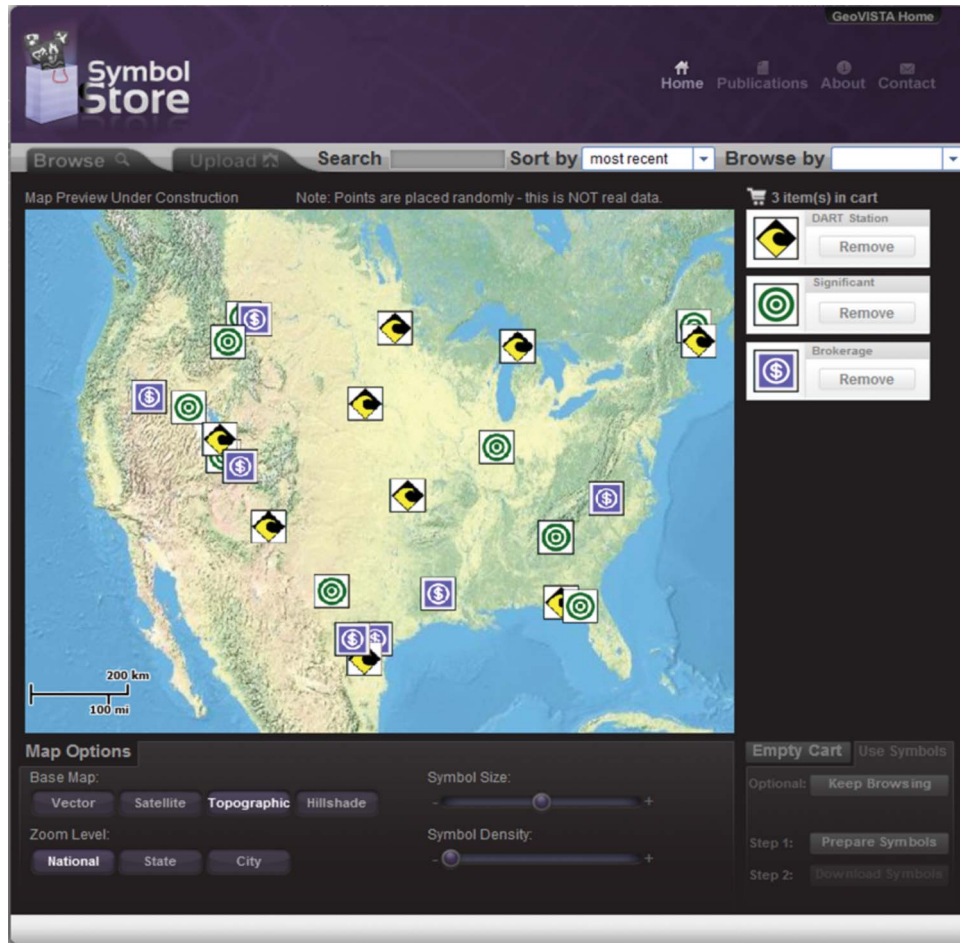
Figure 7.   The map preview function of the Symbol Store allows users to visually evaluate symbols prior to download.
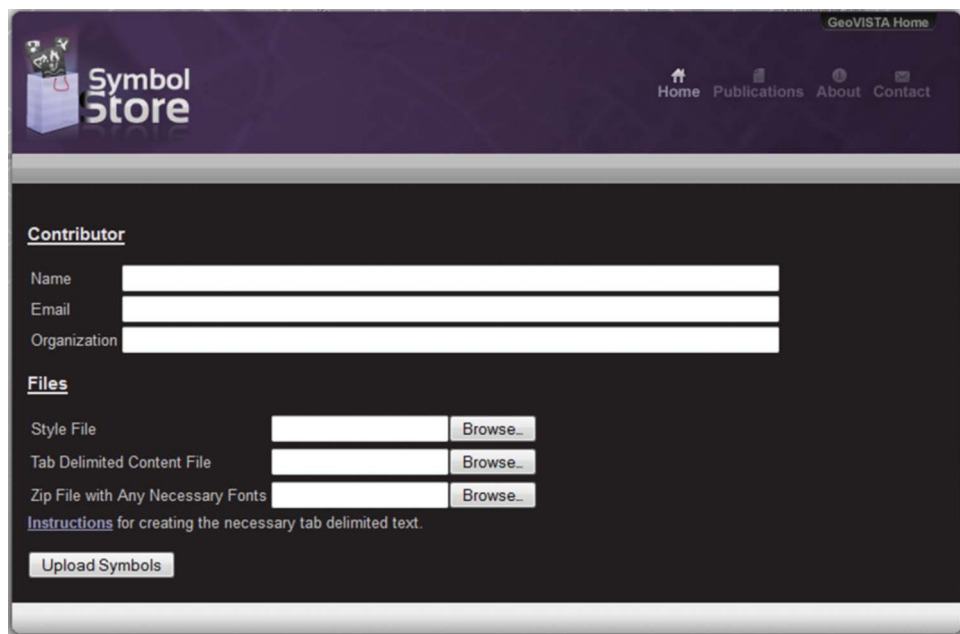
Figure 8.   Public contributions of symbols to the Symbol Store can make use of the new, simplified uploader shown here.

"more like this" when viewing a particular symbol. Along those lines, it may be possible for us to develop a mechanism to crawl the web to automatically collect symbols that already exist in a wide range of formal and informal repositories.

Additionally, many issues exist when it comes to how users define, assign, and interpret categories of symbols, and definitions associated with specific symbols. Multiple examples of the same graphical symbol carrying different meanings have been noted in our prior work with DHS mapmakers (Robinson , Roth, and MacEachren, 2011), and we remain in need of better strategies for highlighting such differences in search results in the Symbol Store to make users aware of different interpretations. Card-sorting and other methods for developing and evaluating categories associated with symbols (Roth et al. 2011) need to be adapted for individuals to use in web-based tools like the Symbol Store.

While supporting map interoperability involves challenges that extend well beyond the common and consistent representation of features using symbols, we believe our work to design and develop the Symbol Store contributes a novel web-based approach that has the potential to significantly help cartographers discover, retrieve, and share symbols beyond the means afforded by current GIS software and informal personal symbol collections.

Acknowledgments

References

Akella, M. K. 2009. "First Responders and Crisis Map Symbols: Clarifying Communication." *Cartography and Geographic Information Science* 36 (1): 19–28.

ANSI. 2006. *ANSI INCITS-415 2006 Homeland Security Mapping Standard – Point Symbology for Emergency Management*. Washington, DC: American National Standard for Information Technology.

Buttenfield, B. 1999. "Usability Evaluation of Digital Libraries." *Science & Technology Libraries* 17 (3): 39–59.

Cutter, S. L. 2003. "GI Science, Disasters, and Emergency Management". *Transactions in GIS* 7 (4): 439–446.

Department of Defense, USA. 2008. *Common Warfighting Symbology: MIL-STD-2525C*. Arlington, TX: Department of Defense.

Dymon, U. J. 2003. "An Analysis of Emergency Map Symbology." *International Journal of Emergency Management* 1 (3): 227–237.

Dymon, U. J., and E. K. Mbobi. 2005. "Preparing an ANSI Standard for Emergency and Hazard Mapping Symbology." In *International Cartographic Conference*, edited by Rodolfo Núñez de las Cuevas, July 9–16, A Coruña, Spain.

Harrower, M. and C. A. Brewer. 2003. "ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps." *Cartographic Journal* 40: 27–37.

Kostelnick, J. C., J. E. Dobson, S. L. Egbert, and M. D. Dunbar. (2008). "Cartographic symbols for humanitarian demining." *The Cartographic Journal* 45 (1): 18–31.

Miller, G. A. 1995. "WordNet: A Lexical Database for English." *Communications of the ACM* 38 (11): 39–41.

Robinson, A. C., R. E. Roth, J. Blanford, S. Pezanowski, and A. M. MacEachren. 2012. "Developing Map Symbol Standards through an Iterative Collaboration Process." *Environment and Planning B: Planning and Design* 39 (6): 1034–1048.

Robinson, A. C., R. E. Roth, and A. M. MacEachren. 2010. "Challenges for Map Symbol Standardization in Crisis Management." In *International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, edited by S. French, B. Tomaszewski, and C. Zobel, May 2–5, Seattle, WA.

Robinson, A. C., R. E. Roth, and A. M. MacEachren. 2011. "Understanding User Needs for Map Symbol Standards in Emergency Management." *Journal of Homeland Security and Emergency Management* 8 (1): 1–16.

Robinson, A. H. 1973. "An International Standard Symbolism for Thematic Maps: Approaches and Problems." *International Yearbook of Cartography* 13: 19–26.

Roth, R. E., B. G. Finch, J. I. Blanford, A. Klippel, A. C. Robinson, and A. M. MacEachren. 2011. "Card Sorting for Cartographic Research and Practice." *Cartography and Geographic Information Science* 38 (2): 89–99.

Saligoe-Simmel, J. 2010. *Ortelius' Community-Based Symbol Collection*. St. Petersburg, FL: North American Cartographic Information Society.

Sheesley, B. 2007. *TypeBrewer: Design and Evaluation of a Help Tool for Selecting Map Typography*. Madison, WI: Department of Geography, University of Wisconsin-Madison.

Sondheim, M., D. Charmley, and G. Leeming. 2010. *Emergency Management Symbology*, Version 1.0, 71. Victoria, BC: Refractions Research.

Journal of Maps

The Basic Ordnance Observational Management System: geovisual exploration and analysis of improvised explosive device incidents

Matthieu J. Murdock [a], Robert E. Roth [b] & Nicholas V. Maziekas [a]

[a] GeoVISTA Center, Department of Geography, The Pennsylvania State University, 302 Walker Building, University Park, PA, 16802

[b] Department of Geography, University of Wisconsin-Madison, 550 N. Park Street, Madison, WI, 53706
Published online: 15 Mar 2012.

PLEASE SCROLL DOWN FOR ARTICLE

SOCIAL SCIENCE

The Basic Ordnance Observational Management System: geovisual exploration and analysis of improvised explosive device incidents

Matthieu J. Murdock[a]*, Robert E. Roth[b] and Nicholas V. Maziekas[a]

[a]*GeoVISTA Center, Department of Geography, The Pennsylvania State University, 302 Walker Building, University Park, PA 16802;* [b]*Department of Geography, University of Wisconsin-Madison, 550 N. Park Street, Madison, WI 53706*

This paper introduces the Basic Ordnance Observation Management System, a prototype application supporting geovisual exploration and analysis of improvised explosive device (IED) incidents. Use of IEDs by terrorist cells has increased in geographic scale, frequency, and sophistication due to the relative cheap cost of acquiring the materials and the ease in keeping such weaponry covert. The Basic Ordnance Observational Management System is designed to facilitate spatiotemporal sensemaking of the IED incident dataset maintained by the National Counter Terrorism Center, Worldwide Incident Tracking System. The application expands upon existing geovisual analytics tools for understanding patterns, trends, and anomalies in IED activity through provision of a suite of capabilities that include: flexible spatial and temporal aggregates of the IED dataset; linked spatial, temporal, and attribute views of the incident information; temporal re-expression, particularly the generation of linear and composite sequences of the IED dataset; and temporal animation across IED activity, with 'VCR' controls and an interactive temporal legend to control the animation as well as visual benchmarks to assist with interpretation of the animation.

Keywords: geovisual analytics; sensemaking; spatiotemporal visualization; cartographic animation; cartographic interaction; visual benchmarks; intelligence analysis; improvised explosive devices

1. Introduction

Sensemaking describes the process of acquiring and synthesizing information relevant to a given problem, exploring this collection of information to extract insights into patterns, trends, and anomalies characterizing the problem, and presenting these key insights to make informed decisions about how best to respond to the problem (Pirolli & Card, 2005). The sensemaking process is particularly important in the context of intelligence analysis, as the information collected is often disparate in source, heterogeneous in format, voluminous in size, and poor in certainty. The field of *visual analytics* has been proposed as an area of research and development to investigate how the sensemaking process – particularly the exploration and analysis of large collections of information – can be supported through visual interfaces to computation methods (Thomas et al., 2005); the term *geovisual analytics* is used to describe the intersection between this research thrust and existing GIScience research (Andrienko et al., 2007).

Here, we introduce the Basic Ordnance Observational Management System (Figure 1), a map-based prototype application developed at the Penn State GeoVISTA Center that leverages established geovisual exploration and analysis techniques to support sensemaking of improvised explosive device (IED) incidents against civilian targets. The proof-of-concept application is publicly available at: http://www.geovista.psu.edu/BOOMsys/. The paper continues with three additional sections. In the following section, we introduce the sensemaking context supported by the Basic Ordnance Observational Management System: Geovisual exploration and analysis of IED incidents. This section provides a logical argument regarding the need for the visual exploration and analysis of IED incidents. In the third section, we introduce the design influences and key functionality of the Basic Ordnance Observational Management System. This description contributes to geovisual analytics by providing an

---

*Corresponding author. Email: mmurdock@psu.edu
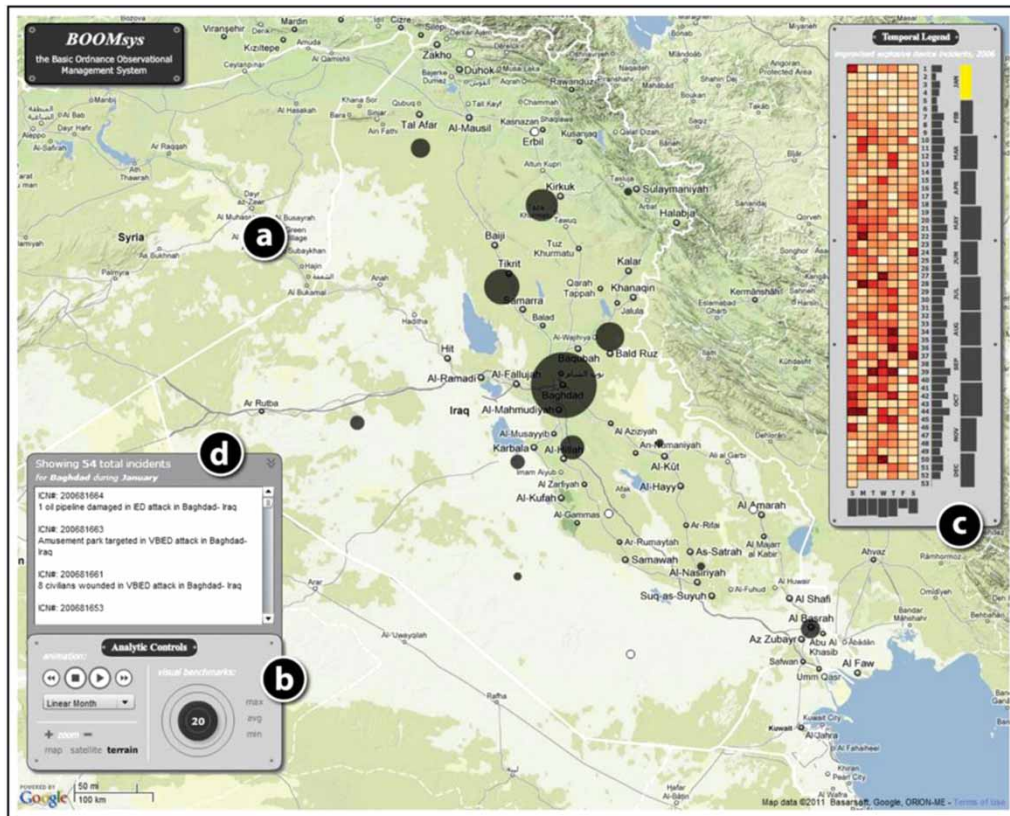
Figure 1. The Basic Ordnance Observational Management System: (http://www.geovista.psu.edu/BOOMsys/). (a) The central map view, showing a set of proportional symbols indicating the frequency of IED incidents for the selected time period; (b) the suite of spatiotemporal analytic controls, including functions for manipulating the map (flexible map browsing and basemap toggling), for controlling the animation (i.e., VCR controls), for changing the temporal binning, and for adding visual benchmarks to the display; (c) an interactive temporal legend based the reruns visualization; and (d) a records list providing metadata and links to each IED incident in the WITS database.

example of how existing and novel visual tools and technologies can be combined in support of spatiotemporal sensemaking. We offer concluding remarks and an outlook on future work in the fourth and final section.

2. Background and problem context

The specific sensemaking context supported by the Basic Ordnance Observational Management System is geovisual exploration and analysis of IED incidents in space and time. Use of IEDs by terrorist cells has increased in geographic scale, frequency, and sophistication due to the relatively cheap cost of acquiring the materials (when compared to military-issued weaponry) and the ease in keeping such weaponry covert. IED attacks accounted for approximately 23,000 civilian and military casualties worldwide in 2009 (NCTC, 2010) and have become the most consistent and substantial threat to multi-national forces deployed in active operational areas (Castro, 2007). A recent report from the United States Government Accountability Office (2009) identified counter-IED shortcomings in several areas, including IED data dissemination, IED spatial analysis, and IED threat analysis. The Basic Ordnance Observational Management System prototype examines the potential for ameliorating these shortcomings through provision of geovisual analytics tools and techniques.

The Basic Ordnance Observational Management System plots a publicly available IED incident database maintained online by the National Counterterrorism Center (NCTC) World Wide Incident Tracking System (WITS) (http://www.nctc.gov/site/other/wits.html), aggregated by country and city. For each incident in the dataset, information is provided about the event type (e.g., armed attack, bombing, kidnapping), the weapon type (e.g., IED, firearm, rocket), the location (e.g., city, country, geocode anchor), the date and time the incident occurred, the number of victims wounded and killed, and an incident summary report. The IED dataset is cleaned and loaded

into the application dynamically at run-time. For demonstrative purposes, the complete dataset has been constrained to only those incidents occurring in Iraq during 2006, producing a reduced subset of 2367 records.

The Basic Ordnance Observational Management System combines coordinated, multiple view visualization and high levels of user interaction to allow the analyst to produce a variety of unique representations of the WITS IED dataset, which in turn facilitates the generation and enrichment of insights (Kraak & MacEachren, 1999; MacEachren, 1994). The primary goal of the Basic Ordnance Observational Management System is to reveal previously unknown and actionable insights about IED incidents, which can be used to inform operational and policy decisions. In particular, an emphasis is placed on sensemaking in both space and time, a key research area within geovisual analytics (e.g., Andrienko & Andrienko, 2005). The Basic Ordnance Observational Management System was developed in ActionScript 3.0 using Flash CS4. The project is available as an open source code repository through the Penn State GeoVISTA Center (http://www.geovista.psu.edu/).

3.   Interactive map design

Navigation of the Basic Ordnance Observational Management System was designed to follow Shneiderman's (1996: 337) mantra of 'Overview first, zoom and filter, then details on demand'. In keeping with this mantra, overview information graphics are first presented in both space and time. Analysts can interact with these overviews to identify patterns, trends, and anomalies of interest. Each incident record then can be examined in detail, assisting the analyst in generating a set of competing hypotheses about the identified pattern, trend, or anomaly. The Basic Ordnance Observational Management System includes four coordinated panels in order to support this structured, top-down navigation: (a) a central map view, (b) a suite of space-time analytic controls, (c) an interactive temporal legend, and (d) a records list; each is considered in the following feature description. Figure 1 shows the four coordinated panels of the Basic Ordnance Observational Management System upon entry to the application.

The central map view (Figure 1a) plots overview spatial aggregates of the IED dataset atop tiles from the Google Maps web mapping service using proportional symbols. The map view provides the cartographic interaction operators typical of a 'slippy' web mapping service (i.e., *pan*, *zoom*, *overlay*, *retrieve*) through the direct manipulation interface style (Roth, 2011). The analyst can manipulate the basemap in accordance with Google Maps mashup conventions, which include flexible map browsing (i.e., panning and zooming; Roth & Harrower, 2008) and toggling among multiple basemap depictions (e.g., overlay of a vector map, satellite image, shaded relief with land cover types). Selection of an individual proportional symbol plotted in the map view activates a popup window containing summary information (i.e., information retrieval), as well as updates the records list.

The analytic controls panel (Figure 1b) includes a suite of interfaces for manipulating the temporal sequencing of the IED dataset as well as an information graphic providing overview temporal aggregates for the entirety of 2006. The analyst is able to toggle among three linear sequences varying by the binning unit, or temporal interval used for the flexible aggregation: (1) linear aggregation by day, producing 365 bins; (2) linear aggregation by week, producing 53 bins due to a partial week at the end of 2006; and (3) linear aggregation by month, producing 12 bins. The analyst also can focus upon cyclical patterns by generating a *composite* sequence of average day by week, resulting in seven bins (Moellering, 1976). The central map only shows incidents occurring during a single interval of time (e.g., a single bin within the complete sequence). To understand trends over time, the analyst can animate the map across all bins using the provided 'VCR' controls (Harrower & Fabrikant, 2008). The temporal binning and cartographic animation functionality leverages the event animation code library described in Roth and Ross (2009). Finally, the analyst also can add static *visual benchmark* rings (Figure 2) around the dynamic proportional symbols that depict the frequency in the minimum bin (i.e., the trough) and maximum bins (i.e., the peak), as well as the average frequency across all bins (Harrower, 2002). The use of visual benchmarks facilitates interpretation of spatiotemporal patterns by overcoming the problem of *disappearance* common to cartographic animation, or the problem of missing intricate or complex patterns due to constant display change, (Harrower, 2003).

To assist in the selection of an appropriate binning unit and sequencing method, an overview temporal graphic (Figure 1c) is provided similar to Weaver et al.'s (2007) *reruns* visualization. This temporal graphic shows the aggregated frequencies by day within the center of the graphic and aggregated frequencies by week (across the vertical axis) and month (across the horizontal axis) along the sides of the graphic. The temporal visualization also acts as a temporal legend for the animation, providing simultaneous overviews for each of the four provided binning sequences and highlighting the current time interval that is plotted to the map. As with the proportional symbols, interaction with the temporal graphic updates the records list; such interaction also advances the animation to the associated time interval, which in turn updates the map view.
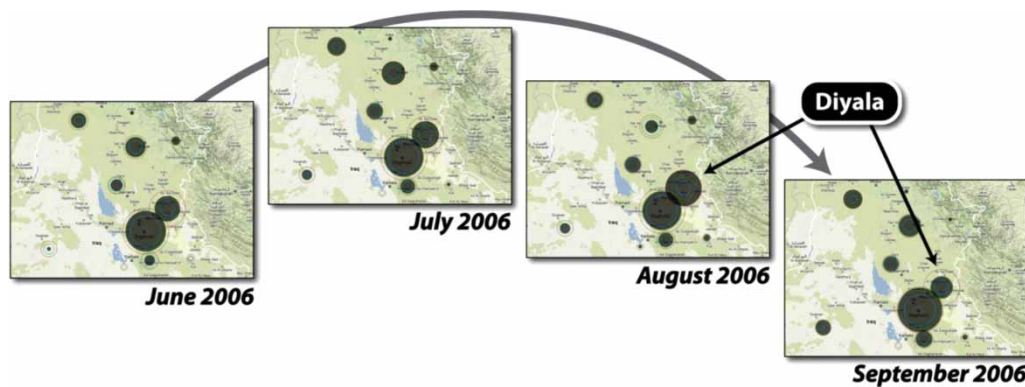
Figure 2.   Cartographic animation and visual benchmarks for space-time Analysis. Using cartographic animation and visual benchmarks in combination reveals that most places peak in September 2006, excepting Diyala, which peaks one month earlier instead (August 2006). Such an anomaly may provide important insight into the way in which IED attacks are planned, which ultimately can be leveraged to determine a course of action that is customized in both space and time.

The records list (Figure 1d) provides analysts with details about each IED incident, completing Shneiderman's (1996: 337) mantra of 'Overview first, zoom and filter, then details on demand'. The records list is populated once the analyst submits a spatial or temporal query (by selecting a proportional symbol or a bin in the temporal legend respectively). The analyst then can follow-up with the identified subset of records through the World Wide Incident Tracking System (WITS) website (http://www.nctc.gov/site/other/wits.html) to learn more about the individual incidents comprising the pattern, trend, or anomaly of interest.

4.   Conclusion and outlook

The Basic Ordnance Observational Management System is a geovisual analytics prototype to support sensemaking of IED incidents through provision of geovisual analytics tools and techniques. It enables analysts to extract temporal and spatial patterns, trends, and anomalies in past IED activity for use in strategic decision making. Future work includes increasing the analytical capabilities of the prototype application as well as scaling the system to a larger geographic extent (i.e., beyond Iraq in order to investigate global patterns of IED attacks on civilians) and temporal extent (i.e., beyond a single year in order to investigate seasonal patterns). In particular, we want to expand the functionality available for filtering the map by additional attributes contained within the IED incident dataset, allowing for exploration and analysis of more nuanced patterns, trends, and anomalies. We also expect to develop use case scenarios and to perform user studies with domain experts in order to refine and extend the system to better support the sensemaking needs of the targeted end users.

References

Andrienko, N., & Andrienko, G. (2005). *Exploratory analysis of spatial and temporal data: A systematic approach*. New York, NY: Springer-Verlag.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M.-J., MacEachren, A., et al. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*(8), 839–857.

Castro, R. (2007). *Improvised explosive device defeat*. Washington, DC: Department of the Army.

Harrower, M. (2003). Tips for designing effective animated maps. *Cartographic Perspectives*, *Winter*(44), 63–65.

Harrower, M., & Fabrikant, S. (2008). The role of map animation in geographic visualization. In M. Dodge, M. McDerby, & M. Turner (Eds.), *Geographic visualization: Concepts, tools, and applications* (pp. 49–65). West Sussex, England: John Wiley & Sons.

Harrower, M.A. (2002). *Visual benchmarks: Representing geographic change with map animation*, Geography, Penn State, University Park, PA.

Kraak, M.-J., & MacEachren, A. (1999). Visualization for exploration of spatial data. *International Journal of Geographical Information Science*, *13*(4), 285–287.

MacEachren, A.M. (1994). Visualization in modern cartography: Setting the agenda. In A.M. MacEachren & D. R. F. Taylor (Eds.), *Visualization in modern cartography* (pp. 1–12). Oxford, England: Pergamon.

Moellering, H. (1976). The potential uses of a computer animated film in the analysis of geographical patterns of traffic crashes. *Accident Analysis and Prevention*, *8,* 215–227.

(NCTC), N. C. C (2010). 2009 Report on Terrorism. Washington, DC: Office of the Director of National Intelligence.

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis, Paper read at International Conference on Intelligence Analysis, at McLean, Va.

Roth, R.E. (2011). Interacting with Maps: The science and practice of cartographic interaction, Geography, Pennsylvania State University, University Park.

Roth, R.E., & Harrower, M. (2008). Addressing map interface usability: Learning from the Lakeshore Nature Preserve Interactive Map. *Cartographic Perspectives*, *60*(Spring), 46–66.

Roth, R.E., & Ross, K.S. (2009). Extending the Google Maps API for event animation mashups. *Cartographic Perspectives*, *64,* 21–40.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualization, Paper read at IEEE Conference on Visual Languages, at Boulder, CO.

Thomas, J.J., Cook, K.A., Bartoletti, A., Card, S., Carr, D., Dill, J., & Earnshaw, R. (2005). *Illuminating the path: The research and development agenda for visual analytics*. Los Alametos, CA: IEEE CS Press.

Weaver, C., Fyfe, D., Robinson, A., Holdsworth, D., Peuquet, D., & MacEachren, A.M. (2007). Visual exploration and analysis of historic hotel visits. *Information Visualization*, *6,* 89–103.

Understanding the Utility of Geospatial Information in Social Media

Anthony C. Robinson
GeoVISTA Center
Department of Geography
Penn State University
arobinson@psu.edu

Alexander Savelyev
GeoVISTA Center
Department of Geography
Penn State University
savelyev@psu.edu

Scott Pezanowski
GeoVISTA Center
Department of Geography
Penn State University
spezanowski@psu.edu

Alan M. MacEachren
GeoVISTA Center
Department of Geography
Penn State University
maceachren@psu.edu

ABSTRACT

Crisis situations generate tens of millions of social media reports, many of which contain references to geographic features and locations. Contemporary systems are now capable of mining and visualizing these location references in social media reports, but we have yet to develop a deep understanding of what end-users will expect to do with this information when attempting to achieve situational awareness. To explore this problem, we have conducted a utility and usability analysis of SensePlace2, a geovisual analytics tool designed to explore geospatial information found in Tweets. Eight users completed a task analysis and survey study using SensePlace2. Our findings reveal user expectations and key paths for solving usability and utility issues to inform the design of future visual analytics systems that incorporate geographic information from social media.

Keywords

Geospatial information, evaluation, social media, visual analytics.

INTRODUCTION

Disasters of all kinds are now responsible for creating vast amounts of social media reports, a large proportion of which contain references to place names and other geographic features (Vieweg, Hughes, Starbird and Palen, 2010). To date there have been few examples of visual analytics evaluations that explore the utility and usability of geospatial information found in social media reports. We do not yet know enough about how users will expect to interact with this information and the nature of the analytical questions that they will wish to explore.

To remedy this deficiency, and to better understand our progress toward supporting geovisual analytics with social media, we constructed a user evaluation of SensePlace2, a geovisual analytics toolkit designed to explore location information found in Tweets in crisis situations (MacEachren, Jaiswal, Robinson, Pezanowski, Savelyev, Mitra, Zhang and Blanford, 2011). SensePlace2 is a browser-based social media mapping application that continuously monitors and mines Tweets that include crisis-related keywords and geocodes the content in each Tweet to assign one or more locations if placenames are present. To extract locations and other named entities from Tweets, SensePlace2 uses a modified version of the ANNIE framework (Cunningham, Maynard, Bontcheva and Tablan, 2002). Extracted locations are then geocoded using the Geonames.org web service. The detailed architecture of SensePlace2 is presented in MacEachren et al. (2011). SensePlace2 is one of the first tools to explore the geographic information found in Tweet contents, as opposed to locations provided by devices or user profiles as is common in many other systems, such as (Field and O'Brien, 2010; Marcus, Bernstein, Badar, Karger, Madden and Miller, 2011; Thom, Bosch, Koch, Worner and Ertl, 2012) . SensePlace2 lets users search crisis-related Tweets using keywords. Query results can then be filtered interactively to narrow by time and geography to explore patterns of location references in the messages themselves as well as the reporting locations (when included). A video overview of SensePlace2 is available at http://www.youtube.com/watch?v=fC7-yGwxhX4 which highlights its key features.

Proceedings of the 10th International ISCRAM Conference – Baden-Baden, Germany, May 2013
T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T.Müller, eds.

918

In this paper we describe a user evaluation to characterize the utility and usability of SensePlace2. Our methodology employs a task analysis component along with a rating survey to elicit qualitative and quantitative feedback. Our research results contribute valuable lessons learned from end-users regarding the utility and usability of a visual analytics environment designed to expose geospatial components of social media reports. Our users' stated analytical questions and feedback on SensePlace2 usability and utility can be used to improve SensePlace2 and to serve as input to the design of new systems that use geospatial social media in crises.

METHODOLOGY

Our aim in this research was to evaluate the prototype SensePlace2 environment to gauge its support for key tasks related to spatio-temporal analysis of qualitative data derived from social media sources. Results from this evaluation can then in turn lead to specific interface improvements and the further development of refined analytical methods. To satisfy these evaluation goals we developed a multi-part user study featuring task analysis (Hackos and Redish, 1998) and survey components to elicit qualitative and quantitative feedback on a range of related areas of concern. Eight participants (3 female, 5 male, all between the ages of 20-29) were recruited for our study from a graduate seminar course focusing on geographical analysis of social media. All participants are currently pursuing a graduate degree (6 in Geography, 1 in Criminal Justice, 1 in Information Science and Technology). We asked participants to rate their expertise in several broad areas, and they indicated their expertise was primarily in Geographic Information Systems, Information Science, and the Social Sciences.

The study procedure includes three parts. First, participants were given a tutorial document providing an overview of the key functions of SensePlace2, along with sample tasks to complete. Second, participants completed three representative tasks using SensePlace2. Finally, after completing these tasks, users completed a usability and utility survey to rate SensePlace2 against a wide range of metrics. The tutorial document, full task descriptions, and survey questions can be found at http://tiny.cc/rluyqw. The study procedure took place as a self-paced, distributed activity, building on prior experience with distributed evaluation methodologies (Bhowmick, Robinson, Gruver, A.M. and Lengerich, 2008). Participants were given instructions on how to access the tutorial and survey website and were instructed to complete the activities at the time of their choosing within a two-week period.

Our goals for evaluation were to take stock of the utility as well as the usability of the SensePlace2 system. To accomplish these goals, we designed several tasks to elicit feedback on the ability for SensePlace2 to support visual exploration of the locations associated with Tweets, to directly evaluate the ability for SensePlace2 to support comparison of different types of locations in Tweets, and to explore the utility of error correction tools we had designed to fix geocoding errors. In the second major component of our evaluation methodology, we employed a survey using 5-point agreement ratings to elicit feedback on common usability metrics as well as a set of metrics intended to reveal SensePlace2's support for situational awareness and analytical reasoning tasks. Our tasks and survey elements were designed based on results from earlier work to survey crisis managers about their current and intended use of geospatial social media, which we describe in MacEachren et al. (2011).
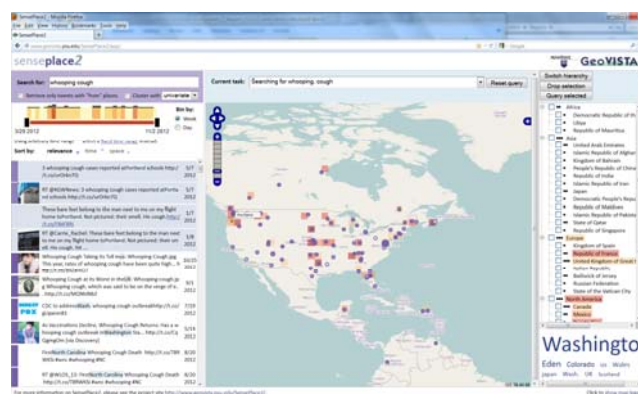
Figure 1: Screenshot showing results for "Whooping Cough" in SensePlace2.

TASK ANALYSIS RESULTS

The first portion of our user study asked participants to complete three tasks and provide qualitative feedback on their analytical findings and experience. These three tasks included basic search and exploration, comparing Tweets with different types of location information, and identifying/fixing geocoding errors.

Proceedings of the 10th International ISCRAM Conference – Baden-Baden, Germany, May 2013
T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T.Müller, eds.

919

Task 1: Basic Search / Exploration Task

Task 1 required each participant to search for Tweets about Whooping Cough (Figure 1) and to identify which places and time periods appeared to be most noteworthy. We asked participants to respond to four questions in this task. Each question and representative answers are highlighted below.

Q1: What geographic patterns do you see?

A1: "Tweets about whooping cough seem to be appearing predominantly in two places: the continental United States and countries bordering the English Channel. There are two spots in Australia and one in New Zealand that are also registering ""Whooping Cough"". This suggests, to me, some possible linkage from whooping cough to countries and people of Anglo descent. When examining locations, Tweets from locations in America tend to link to other locations in America. One exception to this is a person geolocated in Nice, France who is talking about how their mother caught whooping cough when they visited Melbourne, Australia."

Q2: What temporal patterns do you see?

A1: "Binned by week: From the timeline it's very evident that the majority of Tweets on the subject happened early on, in roughly the second week of May. This was part of an initial burst of development in the month of May, followed by a lull through most of the summer, and then resurgence in later August and September.

Q3: What did you learn from the content of the Tweets themselves?

A1: "I learned how the government coped with whooping cough in the way of recommending vaccines to infants and adults. I could guess the peaks from Tweets mentioning "the highest level in U.S." or "HIT 10-year High" and also specific regions where whooping cough broke out."

Q4: Provide at least two questions that you would ask another analyst to explore after seeing these patterns.

A1: "Why are there more outbreaks occurring in the fall? What factors are influencing the outbreak in the northeast?"

Task 2: Comparing Tweets About Places to Tweets From Places

Task 2 required participants to search for mentions of earthquakes and to switch modes in the interface to highlight Tweets that included a "from" location (Coordinates assigned by the device). We asked participants to respond to four questions in this task. Each question and representative answers are highlighted below.

Q1: What geographic patterns do you see?

A1: ""From" Tweets are more concentrated in Japan, Alaska and California, where earthquakes were happening, while "about" Tweets are more distributed."

Q2: What temporal patterns do you see?

A1: "Given that we are investigating earthquakes, I was trying to see whether FROM PLACE Tweets happened before other Tweets regarding a specific earthquake or not. But I had a hard time finding an appropriate one, as many of the Tweets with FROM location are Tweets from NEWS MEDIA or SEISMIC monitoring centers, and not from ordinary users. It seems that the time range was not long enough to find an interesting case."

Q3: What did you learn from the content of the Tweets themselves?

A1: "About Tweets are typically very general, "WTF happened" kind of statements or exaggerations, emotional, little geographical detail, and from individual Tweeters. From Tweets are very specific providing only details, little sentiment or emotion, and are often from "official" sources like news agencies or "earthquake watch" kinds of accounts."

Q4: Provide at least two questions that you would ask another analyst to explore after seeing these patterns.

A1: "It would be interesting to overlay fault lines on the map. How does information temporally and spatially spread from the earthquake epicenter in digital space (how are ideas spreading)? How does population affect the findings of this distribution?"

Task 3: Fixing Geocoding Errors

Task 3 asked participants to search for mentions of fires and to suggest changes to at least ten geocoding errors

Proceedings of the 10th International ISCRAM Conference – Baden-Baden, Germany, May 2013
T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T.Müller, eds.

920

in those results using the geocoding correction interface in SensePlace2. We asked participants to respond to three questions for this task. Each question and representative answers are highlighted. The third question asked for a multiple-choice response and we provide a verbal summary of those responses.

Q1: SensePlace2 allows you to make corrections for a range of geocoding errors. Are there other error types that should be fixable that are not currently supported?

A1: "I don't know how it is treated in the background, but the "misplaced" one is too general. The difference between Washington State and Washington DC is not the same as two completely unrelated locations and they are yet categorized together, that makes a systematized approach a bit difficult."

Q2: Would you add (or take away) from the SensePlace2 interface for handling geocoding errors in Tweets?

A1: "I think it would be nice to able to fix errors in a batch by excluding Tweets that meet certain criteria. For example, when you explore the Tweets about an earthquake in Kobe, Japan, you might want to exclude Tweets that contain keyword "Kobe" from Staples Center, Los Angeles."

Q3: In your opinion, what is an acceptable proportion of results having location accuracy or precision problems when working with social media in a tool like SensePlace2?

Participants indicated that they would be comfortable with a location error rate between 10-20%.

USABILITY AND UTILITY EVALUATION RESULTS

Users rated their experience with SensePlace2 along common usability metrics as well as specific utility metrics that we developed to assess SensePlace2's capabilities to support space-time analysis, situational awareness, analytical reasoning, and geocoding error remediation.

The highest usability ratings were for SensePlace2's overall integration, which participants generally agreed was well-conceived. The lowest ratings concerned SensePlace2's ease of use and the likelihood that most people would be able to learn how to use SensePlace2 quickly. In terms of its basic usability, our participants generally gave average to below-average support when asked to rate SensePlace2 along a range of common usability metrics (Nielsen, 1993) concerning appeal, learnability, intuitiveness, and ease of use (Q1 – Q10 in Figure 2).
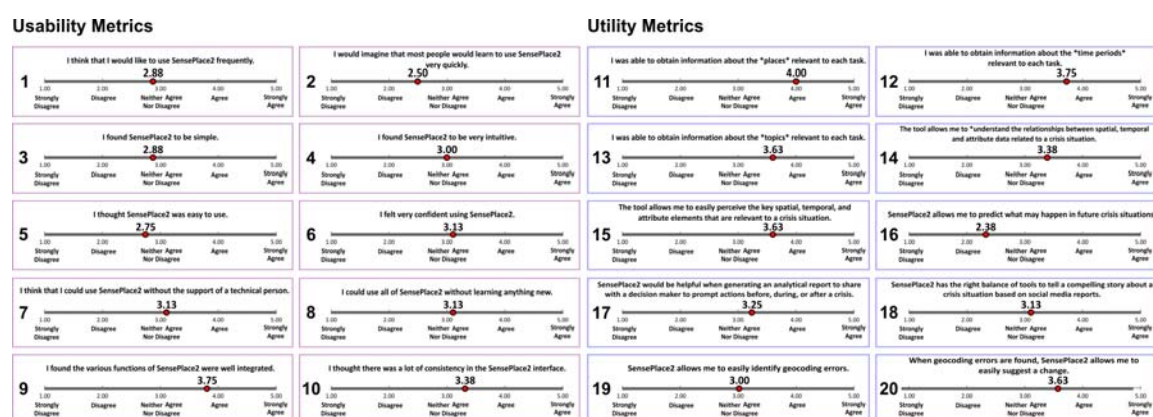
Figure 2: Usability and Utility survey results from the SensePlace2 user evaluation.

In terms of its utility for supporting space-time analysis (Q11-Q13 in Figure 2), participants agreed that SensePlace2 is capable of revealing spatial, temporal, and topical aspects of social media information. Strongest support however was shown for its spatial capabilities, with slightly weaker support for temporal analysis, and the lowest rating given to its ability to reveal topic information.

SensePlace2's support for situational awareness, as defined by Endsley (1995), (Q14-Q16 in Figure 2) was rated positively (above the mid-point) when it comes to perceiving key components of and understanding relationships between space, time, and attribute information. Support for the third component of situational awareness, which concerns prediction, garnered weak agreement from participants (Q16).

Participants did not reach consensus on whether or not SensePlace2 would be helpful for generating a report during a crisis situation or if it would help someone tell a compelling story about a crisis situation. Ratings for

Proceedings of the 10th International ISCRAM Conference – Baden-Baden, Germany, May 2013
T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T.Müller, eds.

921

both questions (Q17-Q18 in Figure 2) yielded average scores around 3.0 which neither agrees nor disagrees with the statement that SensePlace2 supports either design objective.

Finally, participant ratings yielded an average score of 3.0 in terms of its support for easily identifying geocoding errors (Q19 in Figure 2). This score signals no consensus in support for the assertion that SensePlace2 can help easily identify those errors. In contrast, once an error has been discovered, participants generally agreed that SensePlace2 allows one to easily suggest a change (Q20 in Figure 2).

CONCLUSIONS

Our evaluation results reveal that participants view SensePlace2 as having the capability to integrate and analyze geospatial dimensions of social media, but that the execution of the interface has many limitations related to ease of use and support for efficient analysis. Some aspects of situational awareness are well-supported, while others, such as predictive capabilities, have yet to be achieved.

Qualitative feedback from our tasks shows that users were able to generate good answers to our task prompts in most instances. However, users frequently mention that their answers were difficult to generate and that they were uncertain about the quality of those answers. This further supports the overall finding that they key mechanisms may exist to support solid analysis, but that the means for interacting with these mechanisms require significant further refinement. Our preliminary review of qualitative feedback to identify major bugs and ideas for new features provides us with goals for further refinement. Nine major bugs were identified by our participants, and thirteen new features were proposed for enhancing its utility. Following future SensePlace2 development, we will re-evaluate the system to determine if we are able to improve on our previous efforts.

ACKNOWLEDGMENTS

REFERENCES

1. Vieweg, S., Hughes, A.L., Starbird, K. and Palen, L. (2010) Microblogging during two natural hazards events: what twitter may contribute to situational awareness *SIGCHI Conference on Human Factors in Computing Systems* Atlanta, GA

2. MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., and Blanford, J. (2011) SensePlace2: Geotwitter Analytics Support for Situation Awareness *IEEE Conference on Visual Analytics Science and Technology*, Providence, RI.

3. Cunningham, D.H., Maynard, D.D., Bontcheva, D.K., and Tablan, M.V. (2002) GATE: A framework and graphical development environment for robust NLP tools and applications *Association for Computational Linguistics*, Philadelphia, PA.

4. Field, K. and O'Brien, J. (2010) Cartoblography: experiments in using and organising the spatial context of micro-blogging *Transactions in GIS* 14, 1, 5-23.

5. Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S., and Miller, R.C. (2011) TwitInfo: Aggregating and visualizing microblogs for event exploration *SIGCHI Conference on Human Factors in Computing Systems* Vancouver, BC.

6. Thom, D., Bosch, H., Koch, S., Worner, M., and Ertl, T. (2012) Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages *IEEE Pacific Visualization Symposium* Songdo, Korea.

7. Hackos, J.T. and Redish, J.C. (1998) User and task analysis for interface design, John Wiley & Sons.

8. Bhowmick, T., Robinson, A.C., Gruver, A., MacEachren A.M., and Lengerich, E.J.(2008) Distributed usability evaluation of the Pennsylvania Cancer Atlas, *International Journal of Health Geographics* 7

9. Nielsen, J. (1993) Usability Engineering, Academic Press.

10. Endsley, M.R. (1995) Toward a theory of situation awareness in dynamic systems *Human Factors*, 37, 4, 32-64.

Proceedings of the 10th International ISCRAM Conference – Baden-Baden, Germany, May 2013
T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T.Müller, eds.

922

Visual Semiotics & Uncertainty Visualization: An Empirical Study

Alan M. MacEachren, *Member, IEEE*, Robert E. Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan

Abstract—This paper presents two linked empirical studies focused on uncertainty visualization. The experiments are framed from two conceptual perspectives. First, a typology of uncertainty is used to delineate kinds of uncertainty matched with space, time, and attribute components of data. Second, concepts from visual semiotics are applied to characterize the kind of visual signification that is appropriate for representing those different categories of uncertainty. This framework guided the two experiments reported here. The first addresses representation *intuitiveness*, considering both visual variables and iconicity of representation. The second addresses relative performance of the most intuitive abstract and iconic representations of uncertainty on a map reading task. Combined results suggest initial guidelines for representing uncertainty and discussion focuses on practical applicability of results.

Index Terms—Uncertainty visualization, uncertainty categories, visual variables, semiotics.

✦

---

1 INTRODUCTION

Uncertainty is a fact of information; all information contains uncertainty, usually of multiple kinds. While there have been many calls for research about uncertainty visualization as a method to help information users understand and cope with uncertainty [e.g., 1, 2] and a large number of potential strategies and tools for representing uncertainty visually have been developed (see the Background section below), empirical research to assess uncertainty visualization methods has been relatively limited [exceptions include 3, 4, 5, 6, 7, 8, 9]. As a result, our understanding of when and why one uncertainty visualization strategy should be used over others remains incomplete.

Here, we address this gap by reporting on two experiments that provide insights on how to signify different categories of uncertainty. We focus on discrete symbols that could be used to signify uncertainty of individual items within information graphics, maps, or even tables or reports. The experimental design integrates theory from Visual Semiotics, Cartography, Information Visualization, and Visual Perception. Specifically, the experiments examine relative effectiveness of a set of uncertainty representation solutions—differing in the visual variable leveraged and level of symbol iconicity—when used to represent three types of uncertainty (due to accuracy, precision, and trustworthiness) matched to three components of information (space, time, and attribute). The paper is organized in four sections: Background, Experiment #1, Experiment #2, and Conclusion/Discussion.

2 BACKGROUND

Uncertainty representation and visualization has been addressed by a wide range of authors from many disciplinary perspectives. Research on uncertainty visualization has a long history [e.g., 10, 11, 12, 13, 14] and remains an active research topic within both Information Visualization and Cartography [9, 15, 16, 17, 18, 19, 20, 21]. There are multiple contemporary reviews of extant techniques for visualizing uncertainty, including MacEachren et al. [1], Zuk [7], and Bostrom [22]. Rather than summarize or repeat these reviews, we confine background to three topics that underpin the experiments

- *Alan M. MacEachren, Penn State University, e-mail:maceachren@psu.edu.*
- *Robert E. Roth, University of Wisconsin-Madison, e-mail: reroth@wisc.edu*
- *James O-Brien, Risk Frontiers, Macquarie University, e-mail: james.obrien@mq.edu.au*
- *Bonan Li, ZillionInfo, e-mail: bonan.li@zillioninfo.com*
- *Derek Swingley, Penn State University, e-mail: swingley@gmail.com*
- *Mark Gahegan: University of Auckland, e-mail: m.gahegan@auckland.ac.nz*

reported. First, we discuss conceptualizations / taxonomies of uncertainty that link components of information (space, time, and attribute) with the types of uncertainties that may be present in these components. Then, we summarize two visual semiotic frameworks used to inform the uncertainty visualizations examined in the experiments. We first review the *visual variables*, or basic building blocks of a graphic representation, and summarize extant visual variable typologies. Next we describe the difference between iconic and abstract symbols, or the degree to which the sign-vehicle mimics its referent. The background reviews on each of these three topics were used to structure the design of the pair of experiments.

2.1 Conceptualizing Uncertainty

Uncertainty has long been recognized as a multifaceted concept [23]. A typology of uncertainty initially proposed by Thomson et al. [24], and subsequently extended by MacEachren et al. [1], underpins the research presented here. To provide context, we review the core components of the extended typology. It is organized around two primary axes: components of information and types of uncertainty (Table 1). A fundamental distinction typically is made among three components of geographic information: (1) space, (2) time, and (3) attribute; this distinction underlies most efforts to develop efficient and effective information structures for spatiotemporal information and is basic to the human understanding of the world [25].

MacEachren, et al [1] match nine types of uncertainty to these three components of information: (1) accuracy/error, (2) precision, (3) completeness, (4) consistency, (5) lineage, (6) currency, (7) credibility, (8) subjectivity, and (9) interrelatedness. This results in 27 unique conditions of information uncertainty (Table 1). In a case study focusing on spatial uncertainty visualization to support decision making within the domain of floodplain mapping, Roth [26] found accuracy/error to be the most influential of the nine types of uncertainty on decision making, with precision and currency having a secondary influence. Additional empirical investigation across uncertainty conditions has been limited.

2.2 Visual Semiotics

Visual semiotics offers a theoretical framework to conceptualize the mechanisms through which graphic representations can signify both information and its associated uncertainty. In its simplest definition, semiotics is the study of sign systems; the core goal is to understand how a symbol (the sign-vehicle) becomes imbued with meaning (the interpretant) to represent a thing or concept (the referent) [27]. Semiotics provides a framework for understanding both why graphic representations work and how to revise graphic representations for optimal signification. Important to a semiotic theory of information visualization is the identification and articulation of the basic visual variables that can be manipulated to encode information (uncertainty or otherwise).

Table 1. Conditions of Information Uncertainty. 3 components of information (space, time, and attribute) paired with 9 uncertainty types (accuracy/error, precision, completeness, consistency, lineage, currency/timing, credibility, subjectivity, and interrelatedness). Table updated from MacEachren et al. [1]

Category	Space	Time	Attributes
Accuracy/ error	coordinates., buildings	+/- 1 day	counts, magnitudes
Precision	1 degree	once per day	nearest 1000
Completeness	20% cloud cover	5 samples for 100	75% reporting
Consistency	from / for a place	5 say M; 2 say T	multiple classifiers
Lineage	# of input sources	# of steps	transforma-tions
Currency/ timing	age of maps	C = Tpresent - Tinfo	census data
Credibility	knowledge of place	reliability of model	U.S. analyst vs. informant
Subjectivity	local ←→ outsider	expert ←→ trainee	fact ←→ guess
Interrelatedness	source proximity	time proximity	same author

The concept of visual variables was originally outlined by Bertin (under the label of "retinal variables") in 1967 and made available in an English translation in 1983 [28]. Bertin's contention—one that is still generally accepted in Information Visualization and Cartography—was that there are a set of fundamental visual variables, or manipulable primitives of graphic sign vehicles, from which any information graphic can be built. Bertin identified seven visual variables: (1) location, (2) size, (3) color hue, (4) color value, (5) grain, (6) orientation, and (7) shape. Morrison [29] suggested the addition of two more visual variables: (8) color saturation and (9) arrangement. Subsequently, MacEachren [11, 27] proposed adding three variables made practical by advances in graphics technology: (10) clarity (fuzziness) of sign vehicle components, (11) resolution (of boundaries and images), and (12) transparency (each is potentially relevant for signification of uncertainty).

Bertin and others used the concept of visual variables to develop a syntactics of graphic sign vehicles. Syntactics often are described as the 'grammatical rules' of a sign system, detailing how and when the primitive elements of a sign-system should be used for signification. Bertin based his graphical syntactics upon the level of measurement of the signified dataset, giving a rating of acceptable or unacceptable to each visual variable for numerical, ordinal, and categorical data. MacEachren [27] describes the syntactics for the above twelve visual variables, giving a three-step rating of good, marginal, and poor for use with numerical, ordinal, and categorical data. The usefulness of such syntactics of visual variables was demonstrated in Mackinlay's [30] early implementation of an expert system for automating the design of graphical presentations.

The syntactic relations of eleven of the twelve visual variables for representing uncertainty were examined in the first series of each experiment. Figure 1 provides examples of variation in the eleven tested visual variables. Resolution, as presented by MacEachren [27], is omitted because it is applicable to line symbols and images only, while the experiments reported here focus on point symbols only.

2.3 Symbolic Iconicity

Based on accepted information visualization and cartographic principles, we can predict that symbols with a dominant perceptual order will be more effective in tasks that take advantage of pre-attentive visual processes (e.g., visual search tasks, symbol comparison tasks, visual aggregation and region comparison tasks) [27]. Thus, highly abstract symbols that vary only a single visual

THE VISUAL VARIABLES
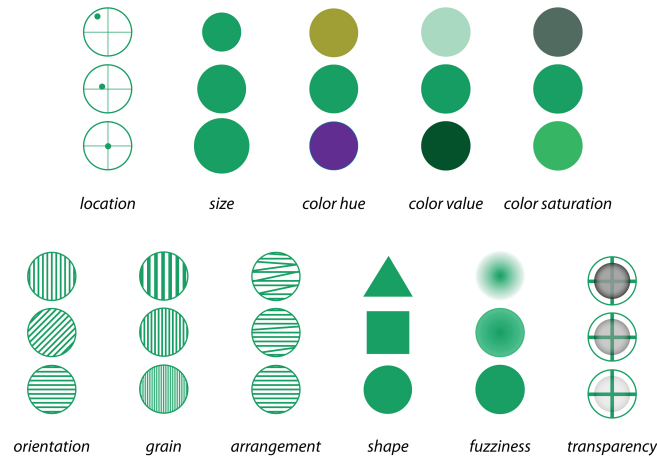
Fig. 1. Visual variables applied to point symbol sets.

variable should be effective at these tasks. In contrast, we also can predict that sign vehicles prompting appropriate metaphors will be easier to match correctly with qualitatively different aspects of information, such as different categories of uncertainty. To prompt metaphors, the variation in symbols needs to incorporate a high degree of iconicity (thus be associative or pictorial rather than geometric; see Figure 2). The characteristics of sign-vehicles that make them iconic, however, often interfere with pre-attentive processing because they are more visually complex.

Ideal symbols, then, are likely to be ones that are easily understood (i.e., that are logically associated with the concept they represent) while also being effective for map reading tasks that require visual aggregation or visual search (i.e., that support pre-attentive processing). These symbol goals represent a fundamental trade-off between abstract sign vehicles, which rely on a single visual variable to communicate differences in the information, and iconic sign vehicles, which are designed to prompt particular interpretants through commonly understood metaphors.

The experiments described below addressed aspects of these two criteria separately; Experiment #1 addressed symbol intuitiveness (i.e., extent to which symbols are directly apprehended or readily understood) while Experiment #2 addressed task performance in situations in which multiple symbols appear on a display.

3 EXPERIMENT #1: ASSESSING INTUITIVENESS

Experiment #1 required participants to judge suitability of symbol sets for representing variation in a given category of uncertainty. Experimental design was informed by the framework of uncertainty conditions introduced in Table 1 and the principles of visual semiotics relating to the visual variables and symbol iconicity. To make the experiment practical, we narrowed the nine-part
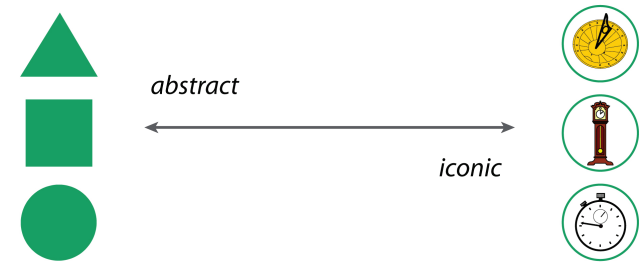

Fig. 2. Symbol Iconicity. Abstract symbols (those that are geometric, varying only a single visual variable) are good for tasks that take advantage of pre-attentive processing. However, iconic symbols (those that are associative or pictorial, prompting metaphors) are potentially easier to match correctly with qualitatively different aspects of data, such as uncertainty conditions.

uncertainty typology detailed by MacEachren et al. [1] to three high-level types: (1) accuracy, defined as correctness or freedom from mistakes, conformity to truth or to a standard or model, (2) precision, defined as exactness or degree of refinement with which a measurement is stated or an operation is performed, and (3) trustworthiness, defined as source dependability or the confidence the user has in the information, thus a broad category that includes aspects of the final seven categories in Table 1. This leaves nine conditions of uncertainty for examination in the experiment (space + accuracy, space + precision, space + trustworthiness, time + accuracy, time + precision, time + trustworthiness, attribute + accuracy, attribute + precision, and attribute + trustworthiness).

Below we describe: (1) design of the symbol sets used in both experiments and (2) design, analysis, and results of Experiment #1.

3.1 Symbol Set Design

Each symbol set contained three symbols matched to a range from high to low certainty; the 3-step scale matched the typology cited above. Symbol sets designed were either iconic (resembling or having similarity with the referent) or abstract (having an arbitrary link with referent, here varying only a single visual variable). The individual symbol sets were grouped into 10 series: one for the general representation of uncertainty and one for each of the nine categories of uncertainty described above. The general series included only abstract symbol sets based upon variation in visual variables. The remaining nine series included both abstract and iconic symbol sets, allowing for comparison between two levels of iconicity. The iconic symbol sets were designed to prompt metaphors specific to the condition of uncertainty represented by the series. For the remainder of the manuscript, we use the term *symbol set* to mean a group of three symbols that could be used to depict three ordinal levels of uncertainty and the term *series* to refer to a group of symbol sets that are compared for a specific condition of uncertainty.

The Series #1 symbol sets conveyed variation in uncertainty by manipulating only a single visual variable; see Figure 1. In Experiment #1, this series of eleven symbol sets were presented in two different directions, with opposite ends 'up' in each variant, resulting in 22 symbol sets. We adopted two design constraints when designing the Series #1 symbol sets. First, color attributes (hue, value, saturation, transparency) were controlled, except when they were the visual variable under consideration. For example, all symbols used the same green hue, except the symbol set relying on color hue to convey information. The use of transparency differed from the others because it is not possible to recognize transparency unless there is an additional feature under the symbol that can be seen through it [31]. Second, all symbol sets, excepting the one using shape, had a circular outline that, excepting the symbol set using size, was the same size. Results from Series #1 provided input to decisions not only about symbolization of uncertainty on its own. Results are relevant to application of each visual variable to redundant signification (e.g., to enhance contrast of iconic symbols that might be logical but not easily located on a map) and to multivariate signification (signification of information plus its uncertainty and/or multiple aspects of uncertainty).

Design of the symbol sets for Series #2-10 focused on determining an appropriate metaphor for each of the nine uncertainty categories. We constructed 10 symbol sets for each of the nine conditions of uncertainty (90 total, subsequently narrowed to 60, see below). We adopted three design constraints. First, within the 10 total symbol sets for a category, five were abstract and five were iconic. Second, the abstract symbol sets, due to their generic design, were included in multiple series to provide a basis for comparability; this approach was not possible for the iconic symbols due to the pictorial customization for each condition of uncertainty. Of the five abstract symbol sets for each series, one abstract symbol set (the color saturation set from Series #1; see Figure 1) was included for all nine conditions of uncertainty. This decision was based on multiple suggestions in the literature that color saturation provides an intuitive method to signify uncertainty [11, 32]. Of the remaining four
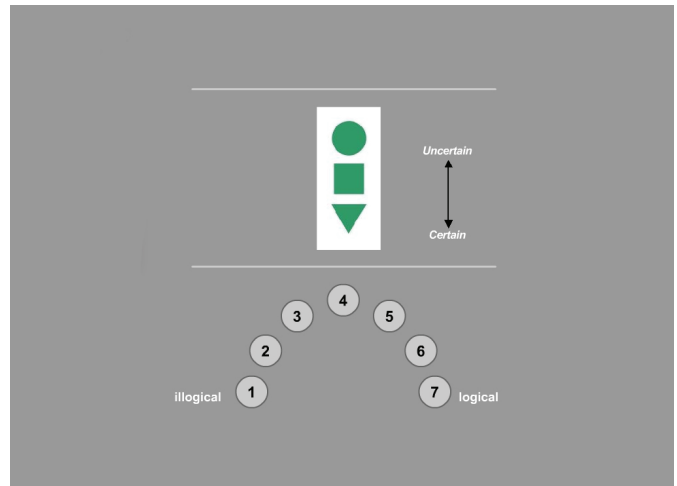
Fig. 3. The Experiment #1 trial interface.

abstract symbol sets, two were common to each component of information (space, time, and attribute) and two were common to each type of uncertainty (accuracy, precision, and trustworthiness). Finally, while the logic behind the design of iconic symbol sets for each series was much more difficult to formalize, it can be noted that each series included iconic symbols emphasizing both confidence ranges and ambiguity. The final 76 symbol set designs are illustrated in Figure 4 along with descriptive statistics from Experiment #1.

3.2 Experimental Design

Experiment #1 focused on assessing symbol set *intuitiveness* (logic) for each uncertainty category and for uncertainty generally. Because inclusion of 102 symbol sets (22 in Series #1 and 90 in Series #2-10) would make the experiment prohibitively lengthy, a pilot study was run with 31 undergraduate students from Penn State University. Participants were asked to rate on a scale of 1-7 the intuitiveness of a symbol set to represent an explicitly defined category of uncertainty from Series #2-10. The top three rated abstract and iconic symbol sets in each series were selected for inclusion in Experiment #1, narrowing each series from 10 symbol sets to 6. The number of symbol sets in Series #1 was left unaltered so that syntactic relations for uncertainty visualization could be formalized for the full set. Thus, the number of tested symbols sets for Experiment #1 was reduced to 76 (22 in Series #1 and 54 in Series #2-10).

Due to inclusion of map-like displays in Experiment #2 (which drew on Experiment #1 results to determine the included symbol sets), participants were purposefully sampled to ensure they had some knowledge of maps and mapping. Therefore, undergraduate students with a GIScience major, graduate students researching a GIScience topic, and professionals working in GIScience and related fields were recruited for participation in Experiment #1. Seventy-two (n=72) participants completed timed suitability ranking tasks with the 76 symbol sets.

An experimental apparatus was created that presented instructions and tasks consistently and to record answers and *response time (RT)*. Participants in Experiment #1 worked in a computer lab with an experiment proctor present, but all instructions were embedded in the experiment application. Each session began with a descriptive overview of the experiment purpose. This was followed by a practice question to introduce the experimental interface. The experiment then progressed through the 10 series of symbol sets described above (thus 76 trials), in each case starting with the Series #1 symbol sets representing uncertainty generally. Between Series #1 and the rest, the components of information (space, time, and attribute) and types of uncertainty (accuracy, precision, and trustworthiness) were introduced in separate screens. Then, prior to beginning a new series, a preview screen containing all symbol sets to be tested in that series appeared for 10 seconds to familiarize participants with the range of symbols in the series. Order

of Series #2-10 as well as order of tasks within all series was randomized to prevent order effects.

After each preview screen, the trial interface was loaded (Figure 3). The interface had two primary components: (1) a symbol set and (2) a set of intuitiveness ranking responses. For each symbol set, the top symbol was labeled as uncertain and the bottom as certain.

Participants specified the intuitiveness of the symbol set by selecting one of the seven interactive ranking buttons. Intuitiveness ranking responses were presented as a discrete visual analog scale (DVAS) from 1 (illogical) to 7 (logical). A DVAS is similar to the more commonly known Likert scale in that they both rely upon evenly-spaced integers to provide quantifiable metrics of participant assessment or preference [33]. However, a Likert scale is presented as a diverging scheme with a central middle point representing the neutral state, with each step in either direction explicitly labeled. The more generic DVAS is presented as a sequential scheme with no neutral middle-point, requiring the labeling of only the poles of the continuum. The DVAS ranking buttons were presented in a half circle, rather than the more traditional horizontal alignment, so that all buttons are an equal distance from this repositioned cursor location. Intuitiveness rankings and RTs were collected for each trial.

Following selection of a intuitiveness ranking, an update screen appeared. The update screen served four purposes: (1) notify about number of trials left in the series and number of series left in the experiment, (2) remind the user about the uncertainty condition for which they are rating each symbol set in the current series, (3) afford a mental break between trials, and (4) ensure that the mouse cursor was at a neutral location prior to every trial.

3.3 Data Analysis

Inferential statistical analysis was applied to the Experiment #1 results in two stages. In the first stage of analysis, differences in intuitiveness and RT were examined within and across series. This stage of analysis was designed to identify the most intuitive symbol set for each condition of uncertainty; this was done for abstract symbols, iconic symbols, and symbols overall. In addition, results of the first stage of analysis provided input for the delineation of syntactic relations among the visual variables for representation of ordinal levels of uncertainty.

In the second stage of analysis, differences between the abstract and iconic symbol sets were examined within and across series. This round of inferential hypothesis testing was designed as a first step to determine the relative merits of abstract versus iconic symbolization for visualizing uncertainty. Series #1 was excluded from the second stage of analysis because of its focus on abstract symbolization only.

For both stages of analysis, nonparametric statistics were applied to intuitiveness rankings, as the recorded random variable is non-continuous when using a DVAS, and parametric testing was applied to the RTs, which were continuous [34]. For the first stage of analysis, the Kruskal-Wallis test (nonparametric) was applied to the intuitiveness rankings and the ANOVA test (parametric) was applied to the RTs; both tests examine statistical difference across three or more groupings. For the second stage of analysis, the Mann-Whitney test (nonparametric) was applied to the intuitiveness rankings and the independent two-group t-test with Welsh df modification (parametric) was applied to the RTs; the Mann-Whitey and t-test are nonparametric and parametric equivalents for examining statistical difference between two unmatched groups. All statistical analysis, descriptive and inferential, was performed using R.

3.4 Results

Results for the first stage of analysis are summarized in Supplement-Table A. Differences in intuitiveness rankings for the Series #1 symbol sets were found to be significant at alpha='0.01'. This confirmed expectation that not all visual variables are intuitive for visualizing ordinal uncertainty information. There was no significant difference in RT, suggesting that participants found the task of judging intuitiveness to be similarly easy/difficult.

Further patterns were identified within Series #1 by looking at descriptive statistics (see Figure 4). Three symbol sets (fuzziness, location, and value) received a mean intuitiveness ranking over '5.0', with fuzziness and location both having a mode of '7' (the highest value on the DVAS). Based on this evidence, we find fuzziness, location, and value to be good for visualizing discrete entity uncertainty reported at the ordinal level. Three symbol sets (arrangement, size, and transparency) received a mean intuitiveness ranking between '4.0' and '5.0' and a modal intuitiveness ranking of '5.0' or higher (with means and medians at the scale midpoint or better), suggesting that they were deemed by participants as somewhat logical for the visualization of uncertainty. Therefore, we find arrangement, size, and transparency to be acceptable for visualizing discrete entity uncertainty reported at the ordinal level. The remaining symbol sets (saturation, hue, orientation, and shape) had mean, median, and modal intuitiveness rankings below '4.0' and were therefore deemed as unacceptable for visualizing discrete entity uncertainty reported at the ordinal level. This is particularly interesting for saturation, which is a commonly cited variable thought to be intuitively related to uncertainty [1].

It is important to note that the presented directionality of both good and marginal symbol sets mattered in their intuitiveness for visualizing uncertainty, as only one direction was deemed intuitive by participants (fuzziness: more fuzzy=less certain; location: further from center=less certain; value: lighter=less certain; arrangement: poorer arrangement=less certain; size: smaller=less certain; transparency: more obscured=less certain).

Returning to Supplement-Table A, a significant difference at alpha='0.01' was found in the intuitiveness ratings across Series #2-10. There are two possible explanations for this finding. The first is that it was more difficult for participants to conceptualize one or several of the uncertainty conditions compared to the rest (e.g., they understood how uncertainty is present in the space and attribute components, but not the time component, or, they understood the accuracy and precision categories of uncertainty, but not the trustworthiness category). The participants may miss the metaphor prompted by a given symbol set if they have a poor conceptualization of the associated condition of uncertainty. The second possible explanation is a difference in logic of symbol sets by series, thus participants may have understood the concepts to be represented, but they did not find the symbol sets in some categories to be logically matched with those concepts. Because a significant difference in RT was not found across Series #2-10—showing that participants did not need to spend more time interpreting some series compared to others—the second explanation is more likely.

Six of the nine conditions of uncertainty (space + accuracy, space + precision, space + trustworthiness, time + trustworthiness, attribute + precision, and attribute + trustworthiness) reported a significant difference in intuitiveness ratings at alpha='0.05' for the symbol sets within the given series (four of these are significant at alpha='0.05'). Thus, all space conditions and all trustworthiness conditions exhibit differences in symbol set intuitiveness ratings. In only one case (see below) is the difference attributable to differences between iconic and abstract symbol sets generally. When examining the descriptive statistics for individual symbol sets within each series (Figure 4), the difference in intuitiveness rankings for the three trustworthiness series is caused by one symbol set receiving distinctly higher ratings while the difference in intuitiveness rankings for space series (aside from space + trustworthiness) is caused by one symbol set receiving distinctly lower ratings.

Only three of nine conditions of uncertainty (space + precision, space + trustworthiness, and attribute + precision) reported a significant difference in RTs at alpha='0.05' (time + trustworthiness is significant at alpha='0.10'). However, all series exhibiting a significant difference in RT also exhibited a significant difference in intuitiveness ranking. This relationship is to be expected, as symbol sets that are not logical or do not invoke proper metaphors will likely take longer to interpret, and therefore longer to rate for intuitiveness. Because this match between differences in ratings and RT was not
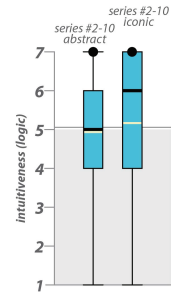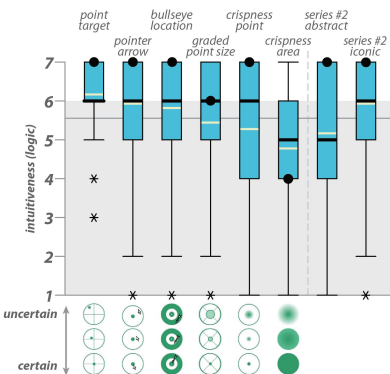
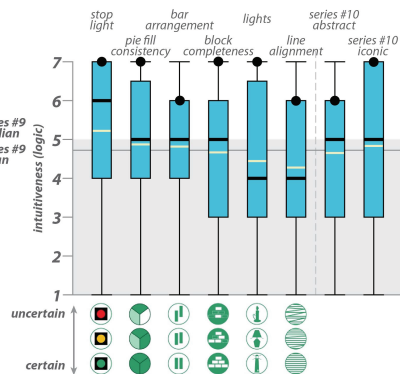# EXPERIMENT #1: INTUITIVENESS (LOGIC)
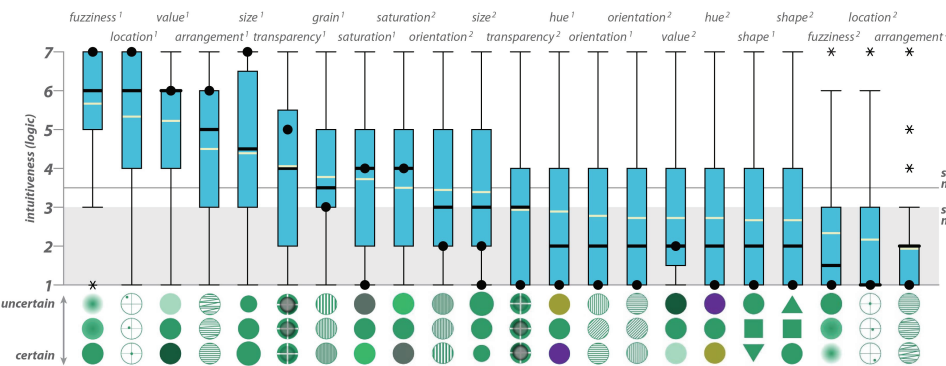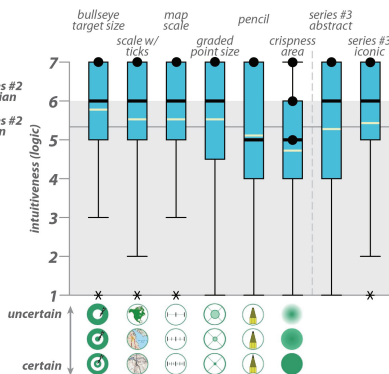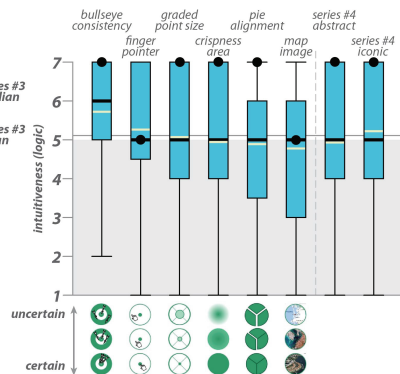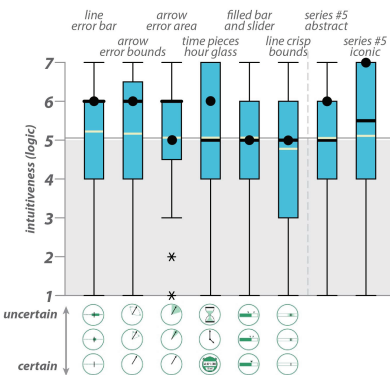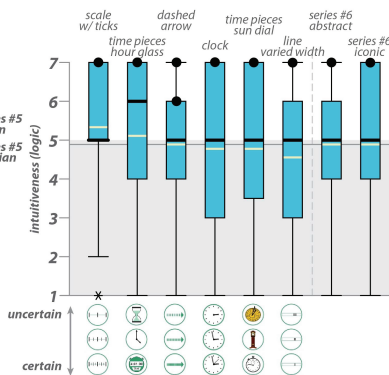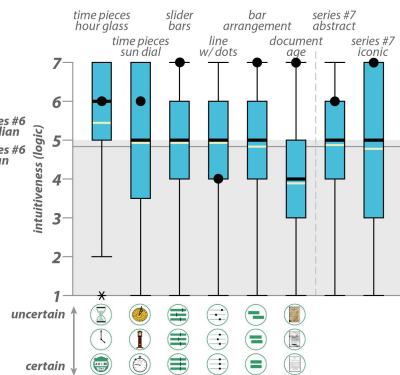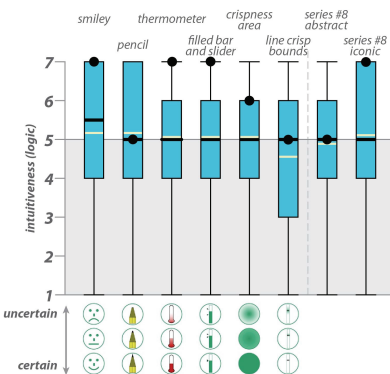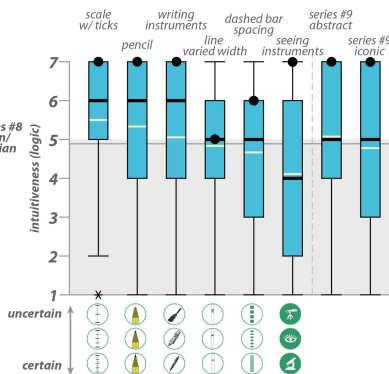
Fig. 4. Descriptive statistics by series and symbol set with results for abstract symbols based on visual variables (Series 1) at the top followed by Series 2-10. On box-plots mean is shown as a black line, median as a gray line, and mode as a black dot.

exhibited in Series #1, in which each symbol set isolated a single visual variable, the relationship between intuitiveness and RT is perhaps only apparent with an increase in symbol iconicity.

As shown in Figure 4 descriptive statistics, the average intuitiveness scores for the Series #2-10 symbol sets were generally higher than those from Series #1, with a large majority of symbol sets receiving a score over '5.0' (the threshold used for Series #1 for marking a particular symbol set as good for visualizing uncertainty). This finding was expected, as the Series #1 set of symbols were designed without any particular category of uncertainty in mind and participants were asked to judge their intuitiveness for general uncertainty signification.

Using the descriptive statistics in Figure 4 with the inferential statistics in Supplement-Table A, it is possible to recommend an abstract and iconic symbol set 'intuitiveness winner' for each condition of uncertainty and to determine if this 'win' is significant (i.e., if the lowest ranking symbol sets can be discredited or if it remains viable). Identifying a intuitiveness winner is useful for the actual application of the symbol sets, but was also essential for administration of Experiment #2. Table 2 summarizes the winning abstract and iconic symbol set for each condition of uncertainty, identifying the symbol sets by the names given in Figure 4, and an asterisk if the win was significant.

The second stage of analysis examined the difference between abstract and iconic symbolization in Series #2-10. The results of this round of analysis are provided in Table 3. Looking at Series #2-10 pooled together, there was a significant difference in intuitiveness rankings at alpha='0.01' between abstract and iconic symbol sets. The descriptive statistics for abstract and iconic symbol sets provided in Figure 4 reveal that iconic symbol sets received a slightly higher mean intuitiveness ranking overall ('5.13') than their abstract counterparts ('4.98'). However, the difference was significant at alpha='0.01' for only one of the individual series, space + accuracy is, with attribute + accuracy significant at alpha='0.10'. This mismatch may be caused by the added statistical power provided when pooling Series #2-10 symbol sets, allowing for the detection of smaller differences between groups with the same level of statistical significance. For three of the nine series (time + precision, time + trustworthiness, and attribute + precision), the abstract symbol sets scored slightly higher than the iconic symbol sets. Thus, it is not possible to state that the iconic symbolization is consistently more intuitive regardless of uncertainty condition.

There was a significant difference in RT between abstract and iconic symbolization at alpha='0.01' when Series #2-10 were grouped. Unlike intuitiveness rankings, however, this relationship also was present when looking at the difference between abstract and iconic symbol sets within a majority of individual series. Five of the nine series (space + precision, space + trustworthiness, time + accuracy, time + precision, and time + trustworthiness) showed a significant difference between RTs at alpha='0.05'; an additional two series (attribute + precision and attribute + trustworthiness) were significant at alpha='0.10'. For all but one of the series (attribute + accuracy), participants required more time to determine the intuitiveness of iconic symbol sets than their abstract counterparts.

The overall result that iconic symbol sets are rated slightly higher on intuitiveness for uncertainty representation but require slightly longer to rate matches theoretically-grounded expectations. Abstract symbol sets should be fast to judge since the process of interpreting order and directionality (i.e., which end means more and which means less) is largely a perceptual task. Iconic symbol sets will require more cognitive processing to identify the intended metaphorical relationship with the uncertainty condition signified. But, since the iconic symbol sets have been designed explicitly to prompt a metaphorical relationship with the uncertainty condition signified, when the design is successful, the rating of intuitiveness should be higher. The fact that iconic symbol sets were not overwhelmingly rated as more intuitive suggests that: (a) the uncertainty conditions are hard for users to conceptualize, thus the match with any metaphor will be weak, (b) differences among

Table 2. Abstract and iconic intuitiveness choices for each series. Groupings with significant differences in intuitiveness ranking (from Table 2) at alpha='0.05' are marked (*)

Series #	Abstract Winner	Iconic Winner
Series #2. Space + Accuracy	graded point size*	point target
Series #3. Space + Precision	scale w/ ticks*	bullseye target size
Series #4. Space + Trustworthiness	crispness area	consistency bullseye*
Series #5. Time + Accuracy	line error bar	arrow error bounds
Series #6. Time + Precision	scale w/ ticks*	time pieces hour glass
Series #7. Time + Trustworthiness	line w/ dots	time pieces sun dial*
Series #8. Attribute + Accuracy	filled bar and slider	smiley
Series #9. Attribute + Precision	scale w/ ticks*	pencil*
Series #10. Attribute + Trustworthiness	pie fill consistency	stop light

uncertainty conditions, while understood by users, do not have obvious visual analogs, or (c) we were simply not successful in designing symbol sets that prompt a metaphor that fits the conceptualization of different uncertainty conditions. The latter was a factor in the *seeing instruments* and *lights* symbol sets for attribute precision and trustworthiness, respectively and in the *document age* set for temporal trustworthiness (Figure 4).

Table 3. Results for the second stage of analysis, assessing statistical differences between abstract and iconic symbolization. The Mann-Whitney test was applied to the intuitiveness rankings and the independent two-group t-test with Welsh df modification was applied to the RTs. Significant results at alpha='0.10', alpha='0.05', and alpha='0.01' are marked in increasing shades of red.

Series #	Intuitiveness Ratings		Response Times		
	W	p-value	t	df	p-value
Series #2. Space + Accuracy	16370.0	*0.0000*	1.4303	341.947	0.1535
Series #3. Space + Precision	21939.0	0.2706	-2.7179	394.349	*0.0069*
Series #4. Space + Trustworthiness	21530.5	0.1574	-3.3146	421.223	*0.0010*
Series #5. Time + Accuracy	22087.5	0.3293	-2.0233	317.988	*0.0439*
Series #6. Time + Precision	23085.5	0.8493	-2.8751	354.435	*0.0043*
Series #7. Time + Trustworthiness	23702.5	0.7696	-2.4773	373.571	*0.0137*
Series #8. Attribute + Accuracy	21150.0	*0.0873*	1.4040	356.348	0.1612
Series #9. Attribute + Precision	24016.0	0.5896	-1.7144	405.775	*0.0872*
Series #10. Attribute + Trustworthiness	22070.5	0.3254	-1.8319	351.631	*0.0678*
Across Series #2-10	1763637.0	*0.0002*	-4.4664	3731.04	*0.0000*

4 EXPERIMENT #2: SYMBOL SETS IN MAP DISPLAYS

4.1 Experimental Design

Experiment #2 complements the focus on symbol intuitiveness from Experiment #1 with a focus on symbol effectiveness for a typical map use task: assessing and comparing the aggregate uncertainty in two map regions. Thirty participants completed the assessment of aggregate uncertainty tasks in a computer lab with a proctor present. As with Experiment #1, undergraduate students with a GIScience major, graduate students researching a GIScience topic, and professionals working in GIScience and related fields were purposefully recruited for participation in Experiment #2 to ensure they had some knowledge of maps and mapping.

The assessment of aggregate uncertainty tasks was completed using the most intuitive abstract and iconic symbol sets identified in Series #2-10 of Experiment #1 (Table 3). In two cases (Series #4 abstract and Series #7 iconic), we used the 2nd highest scoring symbol set for Experiment #2 because the winner already was selected for a different uncertainty type. We included two additional abstract symbol sets from Series #1 of Experiment #1 (fuzziness and color value) that were not identified as the winner for any condition of uncertainty (i.e., in Series #2-10 of Experiment #1), giving us a total of 20 symbol sets for examination, two per series from Experiment #1. Each of the 20 symbol sets was tested in 12 different map region configurations (details below), producing 240 total trials.

Like Experiment #1, Experiment #2 began with a descriptive overview of the experiment, followed by a practice question using the experimental interface. The experiment then progressed in 10 series of 24 trials each. Each trial included two screens shown individually in sequence: (1) a legend showing the three symbols in the tested symbol set with an indication of their order from uncertain to certain (Figure 5) and (2) the map region trial itself (Figure 6). The symbol set legend screen served the secondary purpose as an update screen (as described above for Experiment #1) that offered a mental break and repositioned the mouse cursor to a neutral location.

As shown in Figure 6, the second screen of the Experiment #2 trial interface presented the participants with a map-like display containing nine locations in each of three regions for which uncertainty was indicated by one of the symbols in the trial set. These were presented to the geographically knowledgeable participants as "maps" with two "regions," but the maps were abstract enough to represent information displays more generally. The participant's task was to select the region of the pair for which information is least certain overall. Thus, participants had to conceptually combine nine symbols in each region into an assessment of aggregate uncertainty. Participants submitted their choices by clicking directly on the chosen map region.
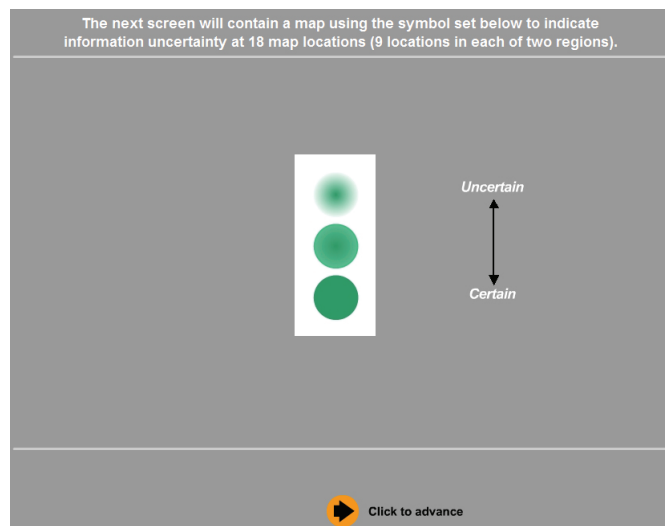
Fig. 5. Example screen #1 of an Experiment #2 trial.
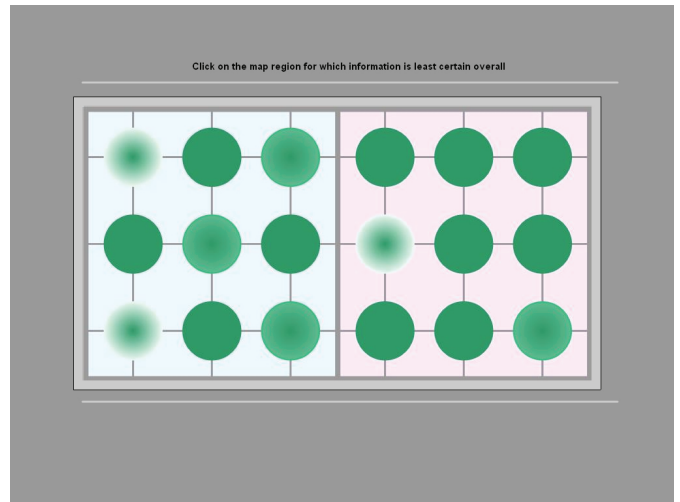
Fig. 6. Example screen #2 of an Experiment #2 trial. The trial interface presents two map regions to the participant, each with uncertainty signified for nine locations. The participant must conceptually aggregate the uncertainty of each region and select the region that is least certain by directly clicking on the map.

The spatial configuration of the uncertainty symbols is likely to influence aggregate judgments. This was controlled for by devising a spatial configuration strategy that prevented participants from being able to memorize the configuration of uncertainty (which might influence their accuracy and speed in responding to the map region comparison task), yet kept the task functionally equivalent from one series to the next (so that the overall level of difficulty for each series of trials was the same). We designed the symbol configurations so that each map region fell into one of four degrees of aggregate uncertainty selected to generate tasks covering a range of difficulty: (1) Highly Uncertain: 7-H + 1-M + 1-C (where H = most uncertain symbol, M = middle symbol, and C = most certain symbol in symbol set); (2) Moderately Uncertain: 4-H + 3-M + 2-C; (3) Moderately Certain: 2-H + 3-M + 4-C; (4) Highly Certain: 1-H + 1-M + 7-C. There are 12 non-equivalent configuration pairings when each individual map region is allowed to fall into one of four degrees of aggregate uncertainty (see Figure 7). We removed configurations where both map regions have equal amounts of uncertainty (i.e., the 1-1, 2-2, 3-3, and 4-4 pairings) so that each trial had a 'correct' answer. All 12 configurations were tested in an individual trial for each of the included symbol sets (20 symbols sets for 12 map region configurations produced 240 total trials).

As in Experiment #1, the first series of trials focused on general uncertainty, without a particular uncertainty condition mentioned. Series #1 included all map region configurations for the crispness (12 trials) and color value (12 trials) symbol sets. After participants completed Series #1, background information on the nine conditions of uncertainty (as reviewed above) was provided to the participants (the same background information as used define the conditions in Experiment #1). The order of the remaining nine series of trials was randomized. Each of the subsequent series included the map region configurations for the abstract (12 trials) and iconic (12 trials) symbol set winners from Experiment #1 for the associated condition of uncertainty; the series numbering is the same in both Experiment #1 and Experiment #2 in the following analysis and reporting. The viewing order of individual trials was randomized within each of the remaining 10 series, as with Series #1. Suitability rankings and RTs were collected for each trial.

4.2 Data Analysis

As with Experiment #1, inferential statistical analysis was applied to the Experiment #2 results in two stages. In the first stage of analysis, differences in accuracy and RT were examined across Series #2-10. This analysis provided insight into the nature of

MAP REGION CONFIGURATIONS

Configuration #1. Highly Uncertain (7-H + 1-M + 1-C)
Configuration #2. Moderately Uncertain (4-H + 3-M + 2-C)
Configuration #3. Moderately Certain (2-H + 3-M + 4-C)
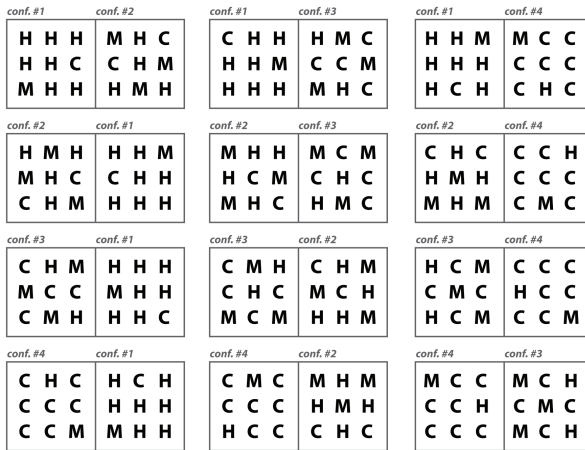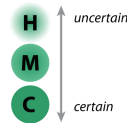**Configuration #4.** Highly Certain (1-H + 3-M + 4-C)

Fig. 7. The 12 map region configurations. Each individual map region was allowed to fall into one of four degrees of aggregate uncertainty, producing twelve possible map region configurations.

geospatial uncertainty and the relative difficulties exhibited when performing map reading tasks under different uncertainty conditions. The inferential statistical analysis considered all Series #2-10 symbol sets together, as well as the abstract and iconic symbol sets individually; Series #1 was not included in this analysis, as this pair of symbol sets was designed for general uncertainty. Descriptive summary statistics were used to identify the symbol sets and uncertainty conditions that garnered the best and worst performance.

In the second stage of analysis, differences between the abstract and iconic symbol sets were examined within and across series. This stage included examining differences between the two Series #1 symbol sets to determine if either supported more accurate or faster assessment of aggregate uncertainty generally. It also included analysis of differences between abstract and iconic symbols in Series #2-10 symbol sets, both pooled together and within each series individually. This step was designed to determine the relative merits of abstract versus iconic symbolization for visualizing uncertainty. The inferential statistical analysis in both stages provided performance measures to complement the intuitiveness measure provided in Experiment #1.

For both analysis stages, nonparametric statistics were applied to assessment accuracy, as the recorded random variable was binary and therefore non-continuous, and parametric testing was used for RTs, which were continuous [34]. For the first stage, the Pearson's chi-square test with Yates' continuity correction (nonparametric) was applied to the accuracy recordings and the ANOVA test was applied to the RTs. For the second stage of analysis, the Pearson's chi-square test with Yates' continuity correction (non-parametric) was applied to the accuracy recordings and the independent two-group t-test with Welsh df modification (parametric) was applied to the RTs. As with Experiment #1, all analysis for Experiment #2 was performed using the statistical software package R.

4.3   Results

The results from the first stage of analysis for Experiment #2 provided the most clear and consistent set of results from either experiment. As shown in Table 4, significant differences in both assessment accuracy and RT were reported at alpha='0.01' across the nine series. The same level of significance was found when examining abstract or iconic symbol sets in isolation or when

Table 4. Statistical results for stage 1 analysis, Experiment #2, differences across uncertainty condition. Pearson's chi-square test with Yates' continuity correction was applied to accuracy recordings and ANOVA was applied to RTs. Significant results at alpha='0.01' are marked in increasing shades of red

Table 5. Results for stage 2, Experiment #2, analyzing differences within and across symbol sets. Pearson's chi-square with Yates' continuity correction is applied to accuracy recordings and the independent two-group t-test with Welsh df modification is applied to RTs. Significant results at alpha='0.10', alpha='0.05', and alpha='0.01' marked in increasing shades of red

pooling all symbol sets together. This finding suggests that participants were not equally comfortable making assessments of aggregate uncertainty for all uncertainty conditions.

Subset	Assessment Accuracy			Response Time		
	x^2	df	p-value	F	df	p-value
Series #2-10 all	31.4829	8	0.000	36.271	8,6471	0.0000
Series #2-10 all abstract	35.2147	8	0.000	24.182	8,3231	0.0000
Series #2-10 all iconic	25.7732	8	0.001	34.838	8,3231	0.0000

Results for the second analysis stage for Experiment #2 are summarized in Table 5 and Figure 8. Pooled data for Series #2-10 exhibited no significant difference in assessment accuracy between abstract and iconic symbol sets. Participants were more accurate using iconic symbols for five of the nine series (space + precision, space + trustworthiness, time + accuracy, time + precision, time + trustworthiness), but only two of these had significant differences (space + trustworthiness, alpha='0.01'; and time + precision, alpha = '0.05'). One series resulted in the abstract symbol set being significantly more accurate (space + accuracy, alpha='0.05'). Overall, the level of iconicity did not have a consistent influence on accuracy of aggregate uncertainty assessment.

Series #	Accuracy			Response Time		
	x^2	df	p-value	t	df	p-value
Series #1. General	0.9976	1	0.318	-0.4745	717.68	0.6353
Series #2-10	0.0549	1	0.459	-5.3275	6231.70	0.0000
Series #2. Space + Accuracy	4.8774	1	0.027	-5.8958	680.60	0.0000
Series #3. Space + Precision	0	1	1.000	-1.511	701.27	0.1312
Series #4. Space + Trustworthiness	11.9707	1	0.001	-8.5933	426.44	0.0000
Series #5. Time + Accuracy	0.2009	1	0.654	-1.5461	717.967	0.1225
Series #6. Time + Precision	6.3712	1	0.0116	2.9178	717.99	0.0036
Series #7. Time + Trustworthiness	1.6911	1	0.194	7.7868	679.033	0.0000
Series #8. Attribute + Accuracy	0.25	1	0.617	-1.2987	710.974	0.1945
Series #9. Attribute + Precision	2.1879	1	0.139	-6.4604	641.259	0.0000
Series #10. Attribute + Trustworthiness	2.6585	1	0.103	1.9579	618.503	0.0507

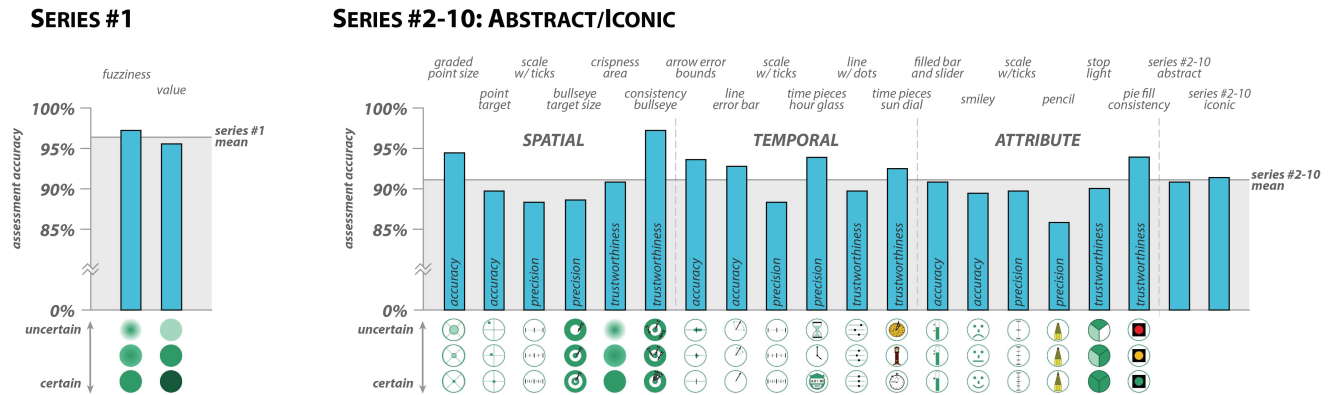# EXPERIMENT #2: ASSESSMENT ACCURACY

Fig. 8. Experiment #2 descriptive statistics by series and symbol set.

As with Experiment #1, RT is related to degree of iconicity in Experiment #2 overall. Pooled results for Series #2-10 exhibited a significant RT difference between abstract and iconic symbol sets at alpha='0.01'. Within series, five of the nine series (space + accuracy, space + trustworthiness, time + precision, time + trustworthiness, and attribute + precision) also reported significant responses time differences at alpha='0.01', with a sixth (attribute + trustworthiness) significant at alpha='0.10'. Like Experiment #1, it generally took longer for participants to compare regions of iconic symbols (mean RT = 3800.41 milliseconds) than regions of abstract symbols (average RT = 3147.81 milliseconds). However, this was not consistent across all series, as significantly more time was taken to respond to abstract symbols for three of the nine series (time + precision, time + trustworthiness, and attribute + trustworthiness). Finally, no significant difference in accuracy or RT was found for the two tested Series #1 symbol sets.

5   CONCLUSIONS & DISCUSSION

Like any controlled experiments, this pair necessarily constrained the problem of uncertainty visualization in multiple ways to enable valid analysis. Thus, applicability of results needs to be considered in relation to the constraints. Within these constraints, the research produced several potentially generalizable conclusions. One is that there is a clear difference in intuitiveness for representing uncertainty among abstract sign-vehicles based upon individual visual variables. Fuzziness and location work particularly well; value and arrangement are also rated highly and both size and transparency are potentially usable. As noted above, saturation, often cited as intuitively related to uncertainty, was ranked quite low. These results, since they relate to fundamental visual variables, may prove to be applicable well beyond the kinds of displays tested here.

Another generalization is that, while iconic sign-vehicles can be more intuitive and more accurately judged when aggregated (than are abstract sign-vehicles), the abstract sign-vehicles can lead to quicker judgments. Plus, and not surprisingly, iconic sign vehicles only work well if users understand both the aspect of uncertainty being signified and the metaphor upon which the sign-vehicles are based (this conclusion is our intuition about how to explain the evidence, but needs further research to assess in depth). Finally, while Experiment #2 focused on "maps", these maps were generic enough that results should generalize to other information displays with multiple points-per-region (e.g., displays depicting cluster results for documents). More importantly, the combined experiments allowed for key principles of sign-vehicle design to be assessed and provide input into guidelines for methods to represent various kinds of uncertainty (individually or in combination) in a range of contexts.

As with any empirical research, many things were not tested, thus results can be considered only a step toward comprehensive understanding of the important parameters for effective uncertainty visualization. Multiple questions remain unanswered. Building on the conceptual framework outlined plus empirical results, the following questions are ones that we feel are particularly relevant to address:

- What symbolization methods work best if there is a need to integrate both data and data uncertainty representation into the same sign-vehicles?
- How scalable are the point symbols (sign-vehicles) tested here? Will they work if reduced in size for use on mobile devices?
- How much impact does the background display have on speed and accuracy of sign-vehicle interpretation?
- How does the spatial distribution of symbolized information impact interpretation?
- Do insights about visual signification of uncertainty at discrete locations (as tested here) extend to linear or area (field) data?

More broadly, the experiments reported here were limited to very simplistic display that was non-interactive for tasks that were simple judgments of suitability or information retrieval. These limitations highlight two important additional next steps in research. First, attention needs to be directed to signification of uncertainty in interactive environments in which users have the ability to control factors such as when uncertainty signification is visible and the relative visual balance between data and data uncertainty in displays showing both at once. Second, once design guidelines are developed to specify the best strategy to signify (or interact with) information uncertainty, an equally important question to answer is how the visualization of uncertainty influences reasoning and decision making in problem context for which uncertainty matters. In spite of experimental limitations and open questions, we believe that the approach to considering uncertainty presented here is a general one that can serve as a framework for deeper understanding of visual signification of uncertainty.

REFERENCES

[1] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler, "Visualizing geospatial information uncertainty: What we know and what we need to know," Cartography and Geographic Information Science, vol. 32, no. 3, pp. 139-160, 2005.

[2] J. J. Thomas, and K. A. Cook, "A visual analytics agenda," Computer Graphics and Applications, vol. January/February, pp. 10-13, 2006.

[3] B. Evans, "Dynamic display of spatial data-reliability: Does it benefit the map user?," Computers & Geosciences, vol. 23, no. 4, pp. 409-422, 1997.

[4] A. M. MacEachren, C. A. Brewer, and L. W. Pickle, "Visualizing georeferenced data: Representing reliability of health statistics," Environment and Planning A, vol. 30, pp. 1547-1561, 1998.

[5] M. Leitner, and B. P. Buttenfield, "Guidelines for the display of attribute certainty," Cartography and Geographic Information Science, vol. 27, no. 1, pp. 3-14, 2000.

[6] J. C. Aerts, K. C. Clarke, and A. D. Keuper, "Testing popular visualization techniques for representing model uncertainty," Cartography and Geographic Information Science, vol. 30, no. 3, pp. 249-261, 2003.

[7] T. D. Zuk, "Visualizing Uncertainty," Dissertation, Department of Computer Science, University of Calgary, Calgary, Canada, 2008.

[8] R. E. Roth, "The impact of user expertise on geographic risk assessment under uncertain conditions," Cartographic and Geographic Information Science, vol. 36, no. 1, pp. 29-43, 2009.

[9] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. J. Moorhead, "A user study to compare four uncertainty visualization methods for 1d and 2d datasets," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 1209-1218, 2009.

[10] H. Ibrekk, and M. G. Morgan, "Graphical communication of uncertain quantities to non-technical people," Risk Analysis, vol. 7, no. 4, pp. 519-529, 1987.

[11] A. M. MacEachren, "Visualizing uncertain information," Cartographic Perspectives, vol. 13, fall, pp. 10-19, 1992.

[12] N. D. Gershon, "Visualization of fuzzy data using generalized animation," in Visualization, Boston, Massachusetts, 1992, pp. 268-273.

[13] B. Buttenfield, "Representing data quality," Cartographica, vol. 30, no. 3, pp. 1- 7, 1993.

[14] P. F. Fisher, "Visualizing uncertainty in soil maps by animation," Cartographica, vol. 30, no. 3, pp. 20-27, 1993.

[15] T. Zuk, and S. Carpendale, "Theoretical anaylsis of uncertainty visualization," in Proc. of SPIE, Visualization and Data Analysis 2006, San Jose, CA, USA, 2006, pp. 66-79.

[16] M. Riveiro, "Evaluation of uncertainty visualization techniques for information fusion," in Proceedings of The 10th International Conference on Information Fusion (FUSION 2007), Quebec, Canada, 2007.

[17] C. Collins, S. Carpendale, and G. Penn, "Visualization of uncertainty in lattices to support decision-making," in Eurographics/IEEE VGTC Symposium on Visualization, Norrköping, Sweden, 2007, pp. 51-58.

[18] A. Streit, "Encapsulation and abstraction for modeling and visualizing information uncertainty," Dissertation, Information Technology, Queensland University of Technology, Brisbane, Australia, 2008.

[19] M. Skeels, B. Lee, G. Smith, and G. G. Robertson, "Revealing uncertainty for information visualization," Information Visualization, vol. 9, no. 1, pp. 70-81, 2009.

[20] D. Lloyd, and J. Dykes, "Exploring Uncertainty in Geodemographics with Interactive Graphics," IEEE Transaction on Visualization & Computer Graphics, vol. 17, no. 12, pp. 2498-4507, 2011.

[21] D. Spiegelhalter, M. Pearson, and I. Short, "Visualizing Uncertainty About the Future," Science, vol. 333, no. 6048, pp. 1393-1400, 2011.

[22] A. Bostrom, L. Anselin, and J. Farris, "Visualizing seismic risk and uncertainty: A review of related research," Annals of the New York Academy of Sciences, vol. 1128, pp. 29-40, 2008.

[23] D. Sinton, "The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study," Harvard papers in GIS #7, D. Dutton, ed., Cambridge, MA: Harvard University, 1978.

[24] J. Thomson, B. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel, "Typology for visualizing uncertainty," in IS&T/SPIE Symposium on Electronic Imaging, Conference on Visualization and Data Analysis, San Jose, CA, 2005, pp. 146-157.

[25] D. J. Peuquet, "It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems," Annals of the Association of American Geographers, vol. 84, no. 3, pp. 441-461, 1994.

[26] R. E. Roth, "A qualitative approach to understanding the role of geographic information uncertainty during decision making," Cartography and Geographic Information Science, vol. 36, no. 4, pp. 315-330, 2009.

[27] A. M. MacEachren, How maps work, New York, NY, USA: The Guilford Press, 1995.

[28] J. Bertin, Semiology of graphics: Diagrams, networks, maps, Madison, WI: University of Wisconsin Press, 1967|1983.

[29] J. L. Morrison, "A theoretical framework for cartographic generalization with the emphasis on the process of symbolization," International Yearbook of Cartography, vol. 14, pp. 115-127, 1974.

[30] J. Mackinlay, "Automating the design of graphical presentations of relational information," ACM Transactions on Graphics, vol. 5, no. 2, pp. 110-141, 1986.

[31] R. E. Roth, A. W. Woodruff, and Z. F. Johnson, "Value-by-alpha Maps: An alternative technique to the cartogram," The Cartographic Journal, vol. 47, no. 2, pp. 130-140, 2010.

[32] D. M. Schweizer, and M. F. Goodchild, "Data quality and choropleth maps: An experiment with the use of color," in GIS/LIS, San Jose, CA, 1992, pp. 686-699.

[33] R. F. DeVellis, Scale development: Theory and applications, 2nd edition, Newbury Park, CA: Sage Publications, 2003.

[34] J. E. Burt, G. M. Barber, and D. L. Rigby, Elementary statistics for geographers, 3rd Edition, New York, NY: Guilford Press, 2009.

Purdue

University

Detection of Symmetric Shapes on a Mobile Device with Applications to Automatic Sign Interpretation

Andrew W. Haddad, Shanshan Huang, Mireille Boutin, Edward J. Delp*

*Video and Image Processing Lab *(VIPER)*
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

ABSTRACT

We present a light-weight method for automatically detecting shapes that have an approximate rotational symmetry (e.g., a square or equilateral triangle) on discrete-space images. Our motivation is the problem of automatically detecting and recognizing hazardous material placards on a mobile platform (e.g., a mobile telephone) equipped with a camera. The proposed method is well-suited for mobile device applications, which are characterized by limited memory, processing power and battery life. It is based on comparing the magnitude of the coefficients of the Fourier series of the centralized moments of the Radon transform of the image after segmentation. However, in our approach, the computation of the Radon transform is bypassed as we obtain these coefficients directly from the rows of the *Pascal Triangle* of the segmented image. The Pascal Triangle of an image is composed of complex moments arranged in a pyramidal fashion similar to the binomial coefficients. These complex moments are obtained from a coarse segmentation of the shape represented by a gray-scale image. In particular, the contours of the object do not need to be precisely defined, and the shape needs not be connected. Moreover, our approach is invariant under translation, rotation, and scaling. We tested our method on images from the MPEG-7 shape database as well as images from our own database of hazardous material placards.

Keywords: Invariant Moments, Radon Transform, Symmetry Detection, Shape Detection, Image Recognition, Rotational Symmetry

1. INTRODUCTION

Automatic recognition of shapes is a common and important task in image interpretation. Shape can be used in the detection, identification and recognition of objects in a scene as well as content-based image retrieval, object tracking, and many other applications. In this paper we present a light-weight method for automatically detecting shapes that have a K-fold rotational symmetry, such as a square or equilateral triangle. A K-Fold Symmetry is defined as a rotation by an angle $\frac{2\pi}{K}$, with respect to the center of mass of the image, which maps the image onto itself. We are dealing with discrete-space (digital) images, thus the symmetries we are trying to detect are only approximate ones, as few rotations preserve the grid structure of the image pixels.

Our motivating application is the problem of recognizing and interpreting hazardous material (HAZMAT) placards - like those found on semi-trailers or oil-tankers on the road. We are interested in developing a system that serves first-responders and civilians by providing information from the Emergency Response Guidebook (ERG)[1] - a guidebook for the initial phase of a dangerous goods/hazardous materials transportation incident - in a faster manner than by searching through the guidebook manually. The challenges encountered in image processing on a mobile device include limited processing power, RAM and battery life. In our application scenario, there is the added challenge of dealing with a time sensitive emergency due to the hazardous nature

of the material involved. However, the four-fold rotational symmetry of the diamond shape characteristic of hazardous material placards, along with their distinguished colors, can be leveraged to detect these placards more efficiently.

The first step of our proposed method is to obtain a coarse segmentation of the object of interest. This can either be a deterministic (binary) segmentation (1 if pixel is in the object, 0 if it is not), or a statistical segmentation where the probability of each pixel belonging to the object of interest is proportional to its gray scale intensity in the segmented image. We then compute the entries of certain rows of the Pascal Triangle of this gray scale image.[2] As we explain in the next section, each row n of the Pascal Triangle contains the coefficients of the Fourier series of the nth order moment of the Radon transform of the segmented image. Thus, after mapping the center of mass of the image to the origin, (approximate) K-Fold Rotational Symmetries correspond to the (near) vanishing of specific entries of the Pascal Triangle. Rotation of the image with respect to the center of mass merely changes the phase of the entries of the Pascal triangle, and rescaling the image multiplies every elements of a row n by the same scaling factor. Thus, we can test for the presence of K-Fold Rotational Symmetries in a rotation and scaling invariant fashion by comparing the magnitude of the coefficients of certain rows of the Pascal Triangle.

We tested our symmetry detection method on a number of shapes from the MPEG-7 shape database. We also implemented and tested our method on a Nexus One Android Mobile Telephone with 512MB RAM (limited to 5MB/process), 1 GHz Qualcomm QSD 8250 processor, 5MP max rear-camera with 2X digital zoom. We used HAZMAT pictures acquired from both the device camera and digital cameras with more sophisticated sensors at distances up to 250 feet, using a number of camera resolutions, with and without analog zoom. After roughly segmenting the image into regions based on the Hue, Chroma, and Lightness (YCH) color space, we checked whether each of the segmented region had a 4 fold symmetry by comparing the Fourier series coefficients of the 6th order centralized moment (i.e., up to a constant multiple, the entries of the 6th row of the Pascal Triangle). Our results indicate that HAZMAT sign can be successfully detected in this fashion by using an appropriately large threshold on the ratio of the magnitude of the relevant coefficients, so that the symbols and numbers contained in the diamond shape are considered as noise.

2. MATHEMATICAL BACKGROUND

The use of invariant moments in pattern recognition, image analysis and interpretation has been used for many years. Hu is often quoted as a pioneer in this area by introducing moments and algebraic invariants to pattern recognition in 1962.[3] More recently,[2] a connection has been established between the geometric features (e.g., elongation, skewness, symmetry) of a shape on an image and the complex moments of that image through the Radon transform. Our work is based on that connection, which we summarize briefly below.

Consider a gray scale image defined on a finite discrete grid $\{(x_k, y_k)\}_{k=1}^N$. More specifically, we have a mapping $\rho : \{(x_k, y_k)\}_{k=1}^N \to \mathbb{R}_{\geq 0}$, where $\rho(x_k, y_k)$ represents the intensity of pixel (x_k, y_k). To simplify our calculations, we use complex coordinates: $z_k = x_k + iy_k$.

DEFINITION 2.1. *We define the Pascal Triangle $T(\rho)$ of the discrete image ρ as the following (infinite) pyramid:*

$$\mu_{0,0}$$
$$\mu_{0,1} \quad \mu_{1,0}$$
$$\mu_{0,2} \quad 2\mu_{1,1} \quad \mu_{2,0}$$
$$\mu_{0,3} \quad 3\mu_{1,2} \quad 3\mu_{2,1} \quad \mu_{3,0}$$
$$\mu_{0,4} \quad 4\mu_{1,3} \quad 6\mu_{2,2} \quad 4\mu_{3,1} \quad \mu_{4,0}$$
$$\vdots$$

$$\mu_{0,n} \quad \binom{n}{1}\mu_{1,n-1} \quad \cdots \quad \binom{n}{l}\mu_{l,n-l} \quad \cdots \quad \binom{n}{n-1}\mu_{n-1,1} \quad \mu_{n,0},$$

$$\vdots$$

where $\mu_{j,l} = \sum_{k=1}^{N} z_k^j \bar{z}_k^l \rho(x_k, y_k)$, $j, l \in \mathbb{Z}_{\geq 0}$, is the complex moment of order (j,l) of the discrete image ρ.

Observe that the top-most moment $\mu_{0,0}$ of the Pascal triangle is normalized to $\mu_{0,0} = 1$ when the image intensity map is a distribution, i.e. when $\sum_{k=1}^{N} \rho(x_k, y_k) = 1$. Observe also the conjugate symmetry property of $T(\rho)$, which is a corollary of the fact that $\mu_{j,l} = \mu_{l,j}^*$. In particular, the central column contains only real-valued entries μ_{jj}.

Note that a continuous analogue for the complex moments $\mu_{j,l}$ have previously been used as features for shape recognition.[4] However, the arrangement of the complex moments into such a pyramid allows us to gain more insight into their geometric meaning. Indeed, if we label the rows of $T(\rho)$ starting from zero, then row n of the Pascal Triangle of an image, which contains the $\mu_{j,l}$ with $j+l = n$, corresponds to the nth order moment of the Radon transform. More specifically, the nth order moment $m_n(\theta)$ of the Radon transform $f_\theta(r)$ of a discrete image ρ is a periodic function (with period 2π) of θ whose Fourier series coefficients can actually be expressed as a finite sum. The coefficients of that series are the entries of for n of the Pascal Triangle, divided by 2^n, as stated in the following lemma.[2]

LEMMA 2.2. *The nth order moment $m_n(\theta)$ of the Radon transform $f_\theta(r)$ is given by the following Fourier series:*

$$m_n(\theta) = \frac{1}{2^n} \sum_{l=0}^{n} \binom{n}{l} \mu_{l,n-l} e^{i(n-2l)\theta}. \tag{1}$$

The Radon transform has been used in the past for image analysis and pattern recognition.[5678] While the connection to the complex moments provided by Lemma 2.2 is interesting, it also allows us to obtain the moments $m_n(\theta)$ without computing the Radon transform.

Translation invariance: Observe that the coefficients of $m_n(\theta)$ are not invariant under translation. However, translation invariance can be obtained by a simple change of coordinates consisting of mapping the center of mass of the image to the origin. This is equivalent to replacing the complex moments $\mu_{j,l}$ by their centralized version:

$$\tilde{\mu}_{j,l} = \sum_{k=1}^{N} (z_k - z_0)^j (\bar{z}_k - \bar{z}_0)^l \rho(x_k, y_k), \quad j, k \in \mathbb{Z}_{\geq 0},$$

where $z_0 = x_0 + iy_0 = \sum_{k=1}^{N} z_k \rho(x_k, y_k)$ is the center of mass of the image ρ. This change of coordinates transforms the moments of the Radon transform into their central counterpart:

$$\tilde{m}_n(\theta) = \sum_{l} (r_l - r_0(\theta))^n f_\theta(r_l) = \sum_{k=1}^{N} \left(r_k(\theta) - r_0(\theta)\right)^n \rho(x_k, y_k),$$

where $r_0(\theta) = x_0 \cos\theta + y_0 \sin\theta$. The relationship between the centralized moments $\tilde{m}_n(\theta)$ of the Radon transform and the centralized complex moments $\tilde{\mu}_{j,l}$ of the image is given by the rows of the centralized Pascal Triangle (i.e. the Pascal Triangle of the centralized image):

$$\tilde{m}_n(\theta) = \frac{1}{2^n} \sum_{l=0}^{n} \binom{n}{l} \tilde{\mu}_{l,n-l} e^{i(n-2l)\theta}.$$

Our method for detecting K-Fold Symmetries in images is based on the following lemma:

LEMMA 2.3.[2] *Let K be an (finite) positive integer. An image ρ has an K-fold rotational symmetry if and only if $\tilde{\mu}_{j,l} = 0$ for all j, l such that $\frac{j-l}{K}$ is not an integer.*

3

Note that the necessity of this vanishing condition for rotationally symmetric shapes was previously observed for continuous-space images.[4] Note also that it would be sufficient to check the condition $\tilde{\mu}_{j,l} = 0$ for all j,l such that $\frac{j-l}{K} \notin \mathbb{Z}$ only for all j,l such that $j + k \leq N$ in order to guarantee the existence of a K-Fold Symmetry.[2]

We illustrate the previous lemma with the following examples of triangles for two, three, and four-fold rotationally symmetric shapes.

Pascal Triangle for Two-Fold Symmetric Shape

$$
\begin{array}{ccccccccccccccccc}
 & & & & & & & & \tilde{\mu}_{0,0} & & & & & & & & \\
 & & & & & & & 0 & & 0 & & & & & & & \\
 & & & & & & \tilde{\mu}_{0,2} & & 2\tilde{\mu}_{1,1} & & \tilde{\mu}_{2,0} & & & & & & \\
 & & & & & 0 & & 0 & & 0 & & 0 & & & & & \\
 & & & & \tilde{\mu}_{0,4} & & 4\tilde{\mu}_{1,3} & & 6\tilde{\mu}_{2,2} & & 4\tilde{\mu}_{3,1} & & \tilde{\mu}_{4,0} & & & & \\
 & & & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & & & \\
 & & \tilde{\mu}_{0,6} & & 6\tilde{\mu}_{1,5} & & 15\tilde{\mu}_{2,4} & & 20\tilde{\mu}_{3,3} & & 15\tilde{\mu}_{4,2} & & 6\tilde{\mu}_{5,1} & & \tilde{\mu}_{6,0} & & \\
 & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & \\
\tilde{\mu}_{0,8} & & 8\tilde{\mu}_{1,7} & & 28\tilde{\mu}_{2,6} & & 56\tilde{\mu}_{3,5} & & 70\tilde{\mu}_{4,4} & & 56\tilde{\mu}_{5,3} & & 28\tilde{\mu}_{6,2} & & 8\tilde{\mu}_{7,1} & & \tilde{\mu}_{8,0}
\end{array}
$$
$$\vdots$$

Pascal Triangle for three-fold symmetric shape

$$
\begin{array}{ccccccccccccccccc}
 & & & & & & & & \tilde{\mu}_{0,0} & & & & & & & & \\
 & & & & & & & 0 & & 0 & & & & & & & \\
 & & & & & & 0 & & 2\tilde{\mu}_{1,1} & & 0 & & & & & & \\
 & & & & & \tilde{\mu}_{0,3} & & 0 & & 0 & & \tilde{\mu}_{3,0} & & & & & \\
 & & & & 0 & & 0 & & 6\tilde{\mu}_{2,2} & & 0 & & 0 & & & & \\
 & & & 0 & & 5\tilde{\mu}_{1,4} & & 0 & & 0 & & 5\tilde{\mu}_{4,1} & & 0 & & & \\
 & & \tilde{\mu}_{0,6} & & 0 & & 0 & & 20\tilde{\mu}_{3,3} & & 0 & & 0 & & \tilde{\mu}_{6,0} & & \\
 & 0 & & 0 & & 21\tilde{\mu}_{2,5} & & 0 & & 0 & & 21\tilde{\mu}_{5,2} & & 0 & & 0 & \\
0 & & 8\tilde{\mu}_{1,7} & & 0 & & 0 & & 70\tilde{\mu}_{4,4} & & 0 & & 0 & & 8\tilde{\mu}_{7,1} & & 0
\end{array}
$$
$$\vdots$$

Pascal Triangle for Four-Fold Symmetric Shape

$$
\begin{array}{ccccccccccccccccc}
 & & & & & & & & \tilde{\mu}_{0,0} & & & & & & & & \\
 & & & & & & & 0 & & 0 & & & & & & & \\
 & & & & & & 0 & & 2\tilde{\mu}_{1,1} & & 0 & & & & & & \\
 & & & & & 0 & & 0 & & 0 & & 0 & & & & & \\
 & & & & \tilde{\mu}_{0,4} & & 0 & & 6\tilde{\mu}_{2,2} & & 0 & & \tilde{\mu}_{4,0} & & & & \\
 & & & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & & & \\
 & & 0 & & 6\tilde{\mu}_{1,5} & & 0 & & 20\tilde{\mu}_{3,3} & & 0 & & 6\tilde{\mu}_{5,1} & & 0 & & \\
 & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & \\
\tilde{\mu}_{0,8} & & 0 & & 28\tilde{\mu}_{2,6} & & 0 & & 70\tilde{\mu}_{4,4} & & 0 & & 28\tilde{\mu}_{6,2} & & 0 & & \tilde{\mu}_{8,0}
\end{array}
$$
$$\vdots$$

As one can see, a K-fold rotational symmetry, can be easily observed in the Pascal Triangle; It is characterized by sequences of $K - 1$ columns of zero entries separated by non-zero columns, corresponding to the different harmonics of the symmetry.

3. PROPOSED METHOD FOR K-FOLD ROTATIONAL SYMMETRY DETECTION

By Lemma 2.3, exact symmetries can be detected by the vanishing of certain columns of the Pascal Triangle. More specifically, to detect a K-fold rotational symmetry, one can compare the coefficients in the $(K+2)$th ,$(K+4)$th , $\cdots$, $(K+m)$th rows of the Pascal triangle of the image, such that $K+m \geq N-1$. Thus, approximate symmetries should correspond to large differences between the magnitude of certain coefficients of the Pascal Triangle. For example, if an image contains an approximate four-fold rotational symmetry, then the coeffients of the 6th row of $T(\rho)$ contains a large $\tilde{\mu}_{1,5}$ and small $\tilde{\mu}_{0,6}$ and $\tilde{\mu}_{2,4}$. Therefore, one method for detecting four-fold rotational symmetries is the following:

$$
\begin{aligned}
&\text{If } max(|\tilde{\mu}_{0,6}|, |\tilde{\mu}_{2,4}|)/|\tilde{\mu}_{1,5}| < T \\
&\Rightarrow \text{ Decide contains a four-fold rotational symmetry} \\
&\text{else} \\
&\Rightarrow \text{ Decide it does not contain a four-fold rotational symmetry}
\end{aligned}
\tag{2}
$$

And similarly for three-fold and five-fold rotational symmetries

$$
\begin{aligned}
&\text{If } max(|\tilde{\mu}_{0,7}|, |\tilde{\mu}_{2,5}|, |\tilde{\mu}_{3,4}|)/|\tilde{\mu}_{1,6}| < T \\
&\Rightarrow \text{ Decide contains a three-fold rotational symmetry} \\
&\text{else} \\
&\Rightarrow \text{ Decide it does not contain a three-fold rotational symmetry}
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
&\text{If } max(|\tilde{\mu}_{0,7}|, |\tilde{\mu}_{1,6}|, |\tilde{\mu}_{3,4}|)/|m_{2,5}| < T \\
&\Rightarrow \text{ Decide contains a five-fold rotational symmetry} \\
&\text{else} \\
&\Rightarrow \text{ Decide it does not contain a five-fold rotational symmetry}
\end{aligned}
\tag{4}
$$

where T is the symmetry detection threshold. Obviously, this detection method can potentially lead to false positives, as one would need to test more than a single row to verify the existence of symmetry. However, because we are searching for approximate symmetries, it is reasonable to believe that one row is sufficient. This claim will be investigated further in Section 5.

Rotation invariance: A rotation of the image clockwise by an angle θ_0 will result in each pixel location $z = x + iy$ being transformed into $ze^{-i\theta_0}$. Hence

$$
\tilde{\mu}_{j,l} = \sum_{k=1}^{N}(z_k - z_0)^j(\bar{z}_k - \bar{z}_0)^l \rho(x_k, y_k) \;\; \rightarrow \;\; \sum_{k=1}^{N}(z_k - z_0)^j e^{-ij\theta_0}(\bar{z}_k - \bar{z}_0)^l e^{il\theta_0}\rho(x_k, y_k) = e^{i(l-j)\theta_0}\tilde{\mu}_{j,l}.
$$

Thus the complex magnitude $|\tilde{\mu}_{j,l}|$ is unchanged under a rotation of the image. So our proposed symmetry detection is rotation invariant.

Scaling invariance: After rescaling the image $\rho(x_k, y_k) \rightarrow \rho(\frac{x_k}{\lambda}, \frac{y_k}{\lambda})$, we have

$$
\tilde{\mu}_{j,l} = \sum_{k=1}^{N}(z_k - z_0)^j(\bar{z}_k - \bar{z}_0)^l \rho(x_k, y_k) \;\; \rightarrow \;\; \sum_{k=1}^{N}\frac{(z_k - z_0)^j}{\lambda^j}\frac{(\bar{z}_k - \bar{z}_0)^l}{\lambda^l}\rho(x_k, y_k) = \frac{\tilde{\mu}_{j,l}}{\lambda^{j+l}}.
$$

Thus

$$
\tilde{\mathfrak{m}}_n(\theta) \;\; \rightarrow \;\; \frac{\tilde{\mathfrak{m}}_n(\theta)}{\lambda^n}.
$$

$\tilde{\mathfrak{m}}_n(\theta)$ is the notation of $m_n(\theta)$ after both the translation and rotation invariance. Therefore the ratio of the coefficients of any given row of the Pascal Triangle of the image is invariant under rescaling. Therefore the magnitude of the ratio between any two coefficients of the same row of the Pascal Triangle is invariant under both a rotation and a rescaling of the image. As a result, our method is invariant under a rescaling of the image.

4. HAZARDOUS MATERIAL PLACARD DETECTION

A real-world application of our symmetry detection method is the problem of recognizing hazardous material placards. Hazardous material placards are characterized by their diamond shape (four-fold rotational symmetry). Placards contain specific colors, each with their own meaning. Colors used in hazardous material placards are Red, Orange, Yellow, Green, Blue and White. In this analysis we consider only signs with high Chroma, so White placards, are not considered. Hazardous material placards often contain, letters, numbers and symbols used to differentiate between materials. For simplicity, we assume the hazardous material placard lies approximately in a plane which is parallel to the image, so we do not need to rectify the image apriori.

We perform a segmentation by transforming each pixel from the RGB color space to the YIQ space (in order to separate the luminance channel from the chrominance channels) and finally, we transform the YIQ space to the YCH color space (separating the hue and chroma channels from the chrominance channels of YIQ). The YCH color space is chosen to take advantage of the property of separation between Hue, Chroma, and Lightness channels.[9] The YCH color space lends itself to the higher order processes that help humans perceive and differentiate color.[10] Such higher order processes are evidenced by the common language used to describe color: Hue, Saturation (relative Chroma) and Lightness. As a matter of fact, as early as 1704, Newton used a representation for color, relating them on a wheel,[11] which is similar to that of Hue.

YIQ Color Space Transformation

The YIQ space was used by the NTSC for transmission of television broadcast signals. It separates the Luminance channel (Y) from the Chrominance channels (I and Q). The transformation of RGB to YIQ is a simple linear transformation described below.[12] The transformation to the Y component is equivalent to the Y component of the 601 Recommendation of the International Telecommunications Union.[13]

$$
\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.59716 & -0.274453 & -0.31263 \\ 0.211456 & -0.522591 & 0.311135 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{5}
$$

YCH Color Space Transformation

The I and Q components can be used to plot RGB as a plane, see Figure 1(a). What we see from this figure is the L_2 norm of the components is representative of the Chroma of the color and the Hue component can be represented as the argument of polar form, found by converting I and Q by use of the two dimensional arctangent (7). The luminance component (Y) of the YIQ space does not require any transformation into the YCH space. This gives us the following definitions for Chroma and Hue.

$$
\begin{aligned}
H &= atan2(I,Q) \\
C &= \sqrt{I^2 + Q^2} \\
Y &= Y
\end{aligned} \tag{6}
$$

While this provides the previously noted benefit of a color space for which the Luminance, Chroma, and Hue can be measured separately, the transformation is the source of a bottle neck in our procedure, four quadrant inverse tangent (7). Thus we use a faster estimate of the four-quadrant inverse tangent (8).

$$
atan2(I,Q) = \begin{cases}
atan(\frac{Q}{I}) & I > 0 \\
\pi + atan(\frac{Q}{I}) & Q \geq 0, I < 0 \\
-\pi + atan(\frac{Q}{I}) & Q < 0, I < 0 \\
\frac{\pi}{2} & Q > 0, I = 0 \\
\frac{-\pi}{2} & Q < 0, I = 0 \\
\text{undefined} & Q = 0, I = 0
\end{cases} \tag{7}
$$

(a) Plane from the YIQ space, varying I and Q from -1 to 1 and constant Y value of 0.5
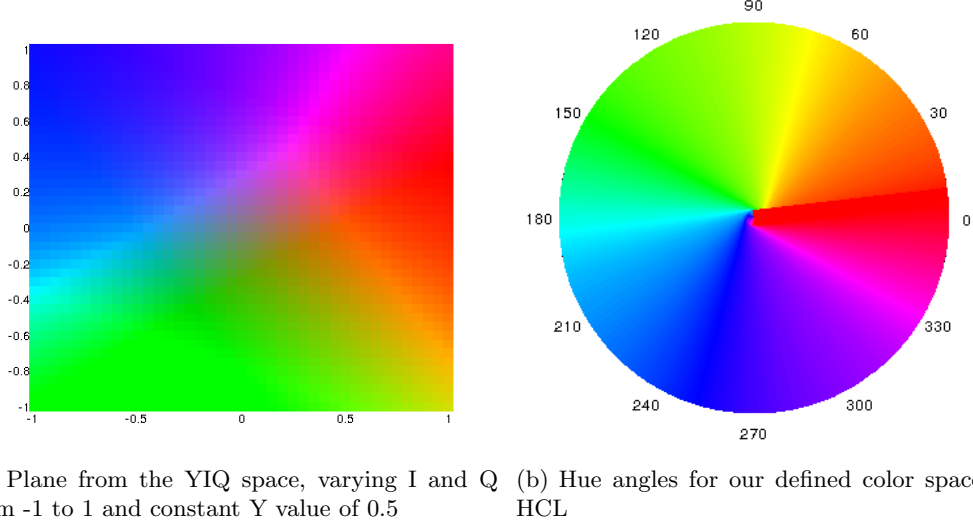
(b) Hue angles for our defined color space HCL

Figure 1. Plane from YIQ space and Hue wheel from HCL space

$$fast\_atan2(I, Q) = \begin{cases} \frac{\pi}{4} - \frac{3\pi}{4}\left(\frac{I-|Q|}{I+|Q|}\right) & I \geq 0, Q \geq 0 \\ \frac{\pi}{4} - \frac{3\pi}{4}\left(\frac{I+|Q|}{I-|Q|}\right) & I < 0, Q \geq 0 \\ -\left(\frac{\pi}{4} - \frac{3\pi}{4}\left(\frac{I-|Q|}{I+|Q|}\right)\right) & I \geq 0, Q < 0 \\ -\left(\frac{\pi}{4} - \frac{3\pi}{4}\left(\frac{I+|Q|}{I-|Q|}\right)\right) & I < 0, Q < 0 \end{cases}$$ (8)

The implementation of $fast\_atan2$ is based on an implementation by Shima.[14] We analyze the effects and gains of using the faster estimated $atan2$ function in the experimental results section of this paper.

Chroma and Lightness Thresholding

To determine if a particular pixel can be accurately segmented using we threshold based on the Chroma and Luminance. Indeed, low chroma pixels tell us little about the color of the region they are contained in and so we discard them as components. Similarly, because we have chosen to focus on hazardous material placards of the color Red, Orange, Yellow, Green or Blue, we can discard pixels that are either too dark or too light to have come from one of these placards. To discard a pixel, we set its value in the label-image, described in the following section, to 0.

Pixel-Labeling Based on Hue

We create label-images corresponding to each color, i.e. a segmentation where pixels are labeled 0 to 5. A pixel labeled with a 0 will not be analyzed because it was not saturated or was too dark/light to be contained within a hazardous material placard.

To label each pixel, we proceed as follows. If a pixel has passed through both of the aforementioned chroma and luminance thresholds, a label is chosen based on the Hue angle, θ, of that pixel. We determine the label of a particular pixel based on nearest neighbor - with regard to angular distance, δ_i (10) - from a set of predefined angles (see Table 1). Predefined Hue angles were chosen based on the Hue angle of signs from our test set. Our nearest neighbor decision method is desribed in Equations (9) and (10).

$$label = \underset{i \in \{1,2,3,4,5\}}{argmin} (\delta_i)$$ (9)

Before   After

Figure 2. Before and after applying the luminance and chroma thresholds. $(0.2 < Y < 0.8, c > 0.1)$

Color	Angle, θ_i, (Radians)	Label
red	1.0521	1
orange	1.8716	2
yellow	2.2928	3
green	3.6664	4
blue	5.5104	5

Table 1. Hues and their angles and labels used in this procedure

$$\delta_i = min(|\theta - \theta_i|, |(\theta - \theta_i) + 2 * pi|); \tag{10}$$

The label of the nearest neighbor is considered to be the current pixel's label. The label-images created from this procedure are then used as the basis for segmentation via a Connected Component analysis.[15] Each component found in this analysis is later tested using our method for four-fold rotational symmetry detection described in Section 3.

5. EXPERIMENTS AND RESULTS

5.1 Symmetry Detection Using the MPEG-7 Shape Database

In order to analyze the effectiveness of our symmetry detection method, we vary the the threshold T used to determine whether or not a symmetry exists, from 0 to 1 at increments of 0.05. We complete this analysis for three, four, and five-fold rotational symmetries. We chose 140 rotationally symmetric shapes from the MPEG-7 shape database.The shapes chosen contain at least three, four, or five-fold rotational symmetry but many contained higher order symmetries (6 or 8-Fold) as well. Figure 3 shows the precision, recall, and accuracy for each experiment.

These experiments indicate the need to vary thresholds for different rotational symmetries. Further analysis is required to detemine why certain symmetries are easier to detect than others. Our experiments indicate a reasonable threshold for 5-Rotational Symmetries to be between 0.3 and 0.35. This threshold produced a nearly perfect precision and recall. 3 and 4-Rotational Symmetries recalled fewer results - seemingly due to the inclusion of 6 and 8-fold rotational symmetry components included in our experiment.

The accuracy results for three-fold rotational symmetries indicates 0.7 to be a good threshold for detection. We find the experiments on four-fold rotational symmetries to indicate a proper threshold around 0.4. Figure 4 gives a sample of some of the shapes chosen for analysis in this experiment. Figure 4(a) shows two shapes considered to be 3 or 4-rotationally symmetric for which our method failed to recognize as symmetric, even with a large threshold.

8

(a) 3-FRS, Max Accuracy: 90.1%

(b) 4-FRS, Max Accuracy: 81.3%

(c) 5-FRS, Max Accuracy: 98.1%
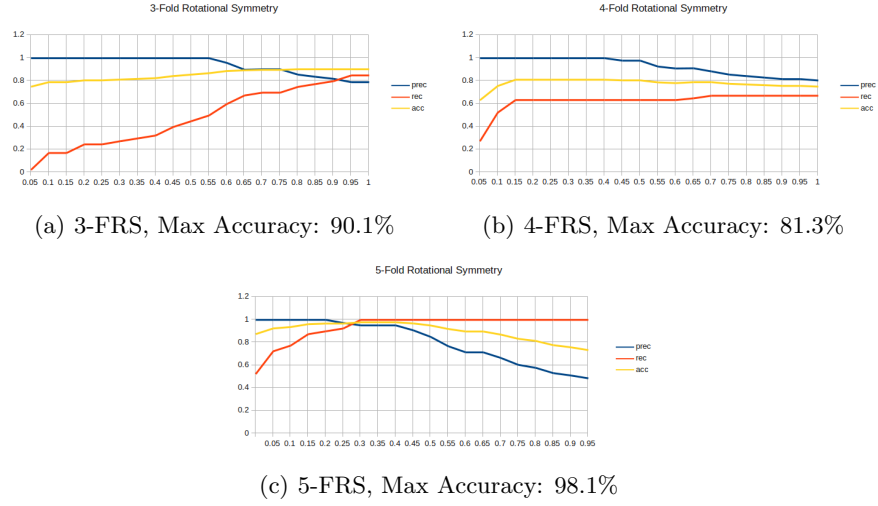
Figure 3. Precision, Recall and Accuracy for T from 0 to 1 at increments of 0.05

(a) Undetected symmetries

(b) Detected as three-fold rotational symmetry

(c) Detected as four-fold rotational symmetry

(d) Detected as five-fold rotational symmetry
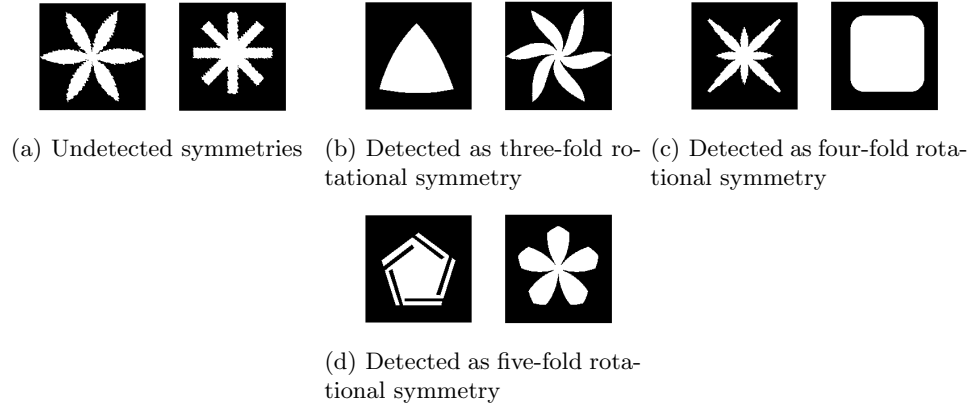
Figure 4. Small selection of symmetric shapes chosen from the MPEG-7 Shape Database
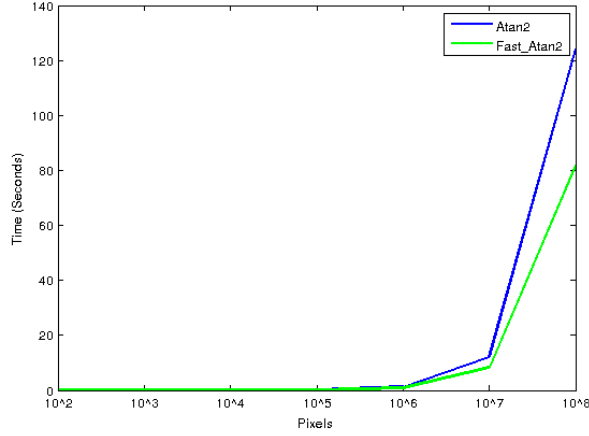
Figure 5. Exectution time for varying number of pixels for $atan2$ and $fast\_atan2$

5.2 Symmetry Detection Using Hazardous Material Placards

5.2.1 YCH Transformation Comparison

In Section 4 we discussed our technique for transformation to the YCH space. Our technique requires the use of the four-quadrant inverse tangent, $atan2$, function. One issue that arises from the use of this function is the high computational expense. For that reason we used an estimate of the $atan2$ function, called $fast\_atan2$. In order to determine suitability o $fast\_atan2$ as a replacement, we analyze the error between the two functions in the range of possible inputs from the YIQ space described in Section 4. In order to do this, we calculated the Hue value for every possible RGB triplet using both methods and computed the absolute differences between methods.

$$
\begin{aligned}
\text{Mean Absolute Difference} &= 0.0461 \text{ radians} \\
\text{Max Absolute Difference} &= 0.0711 \text{ radians}
\end{aligned}
\tag{11}
$$

These numbers show a small mean absolute difference between the results of the $atan2$ and $fast\_atan2$ functions. The maximum absolute difference is slightly larger than the mean and is a small trade-off for the decreased computational complexity. In order to compare the time required for both methods, we tested on an Android mobile device - in particular, a Nexus One Android Mobile Telephone with 512MB RAM (limited to 5MB/process), 1 GHz Qualcomm QSD 8250 processor. In Figure 5 we see a very small savings for a small number of calculations but the required computation for large images (5 megapixels and up) is significant.

5.2.2 Segmentation Results

As we are not analyzing the segmentation method chosen for the process of symmetry detection, we do not provide results regarding its effectiveness. However, we provide a number of results, see Figure 6, that we believe to show minimum effectiveness to analyze the usefulness of our symmetry detection method.

As can be seen in Figure 6(d), the segmentation algorithm suffers from over segmentation on small images of hazardous material placards. How this effects our result is discussed in greater detail in the following Section.

10

5.2.3 Symmetry Detection Thresholds

We have applied our symmetry detection method to find hazardous material placards in images. To determine the effectiveness on this real-world application, we analyze the effect of varying our threshold, T. Again, we vary our threshold between 0 and 1 at increments of 0.05. We leverage the four-fold rotational symmetry of the hazardous material placard and use the same method as explained in Section 3 and used in Section 5.1.

In Figure 7 we exhibit some results from our symmetry detection method on four-fold rotational symmetries and we see a number of components which were properly rejected (Figures 7(c) and 7(d)). Figures 7(a) and 7(b) show how increasing our threshold can increase the number of signs detected - this of course risks the false detection of non-signs, as can be seen by the curves in Figure 8. Notice, our first occurence of a false positive appears at a threshold of 0.55. After this quick drop in precision, it begins to rise again, along with our recall, making a threshold of 1 appear to be a good choice for our data. Notice of course that the recall never quite reaches one - even at high thresholds. This is due to oversegmentation of signs with a small number of pixels (Figure 6(d)).

6. CONCLUSION AND FUTURE WORK

We have presented a method for automatically detecting approximate K-fold rotationally symmetric shapes in discrete images. The method is based on comparing the magnitude of the Fourier series coefficients of the Radon transform of a coarse segmentation of the shape. To keep the computational cost low, we bypass the computation of the Radon transform and obtain the coefficients directly from the complex (centralized) moments of the segmented image; after arranging these coefficients into a triangle in a fashion similar as the binomial coefficients, we obtain the Pascal Triangle of the segmented image, which has been shown to be a lossless representation of the image.[2]

We tested our method on rotationally symmetric shapes from the MPEG-7 shape database. The shapes tested had three, four or five-fold rotational symmetries and we were successful in finding the appropriate symmetries for most shapes in our test set.

We also tested our method to detect hazardous material placards based on their four-fold rotational symmetry. We coarsely segmented our images using a nearest neighbor classification method in the YCH color space and a connected component procedure to find possible symmetric shapes in the images. Then, we use the 6-th central moment of the Radon transform (obtained from the Pascal triangle) of a discrete components in an image to identify approximate symmetries. We implemented and tested this method on a Nexus One Android Mobile Telephone with 512MB of RAM (limited to 5MB/process), 1 GHz Qualcomm QSD 8250 processor, 5MP max rear-facing camera with 2X digital zoom. We found more or less than equal success in this test as was found in our test on rotationally symmetric, pre-segmented components of the MPEG-7 database.

We can improve our method by making use of color calibration methods to realign the hue angles, determined by the imaging device used. This would improve the segmentation method used prior to our symmetry detection. For further improvements to the segmentation we could replace the hard Chroma and Luminance thresholds by probabilistic framework to determine inclusion in a particular segment.

REFERENCES

[1] [*Emergency Response Guidebook, 2008*], US Department of Transportation Pipeline and Hazardous Materials Safety Administration et. al. Office of Hazardous Materials Initiatives and Training, Washington D.C., USA (2008).

[2] Huang, S., "On the Pascal Triangle of a discrete image," (2012). Manuscript in Preparation.

[3] Hu, M.-K., "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on Information Theory* **8**, 179–187 (1962).

[4] Flusser, J., Zitova, B, and Suk, T., [*Moments and Moment Invariants in Pattern Recognition*], Wiley and Sons Ltd. (2009).
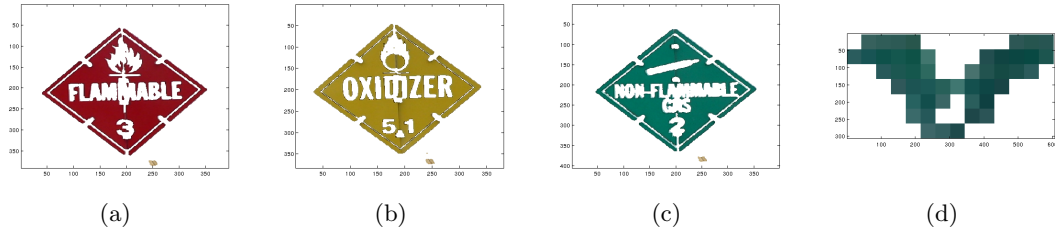
(a)   (b)   (c)   (d)

Figure 6. Segmentation results for Red, Yellow and Green hazardous material placards. 6(d) suffers from over segmentation

(a)   (b)
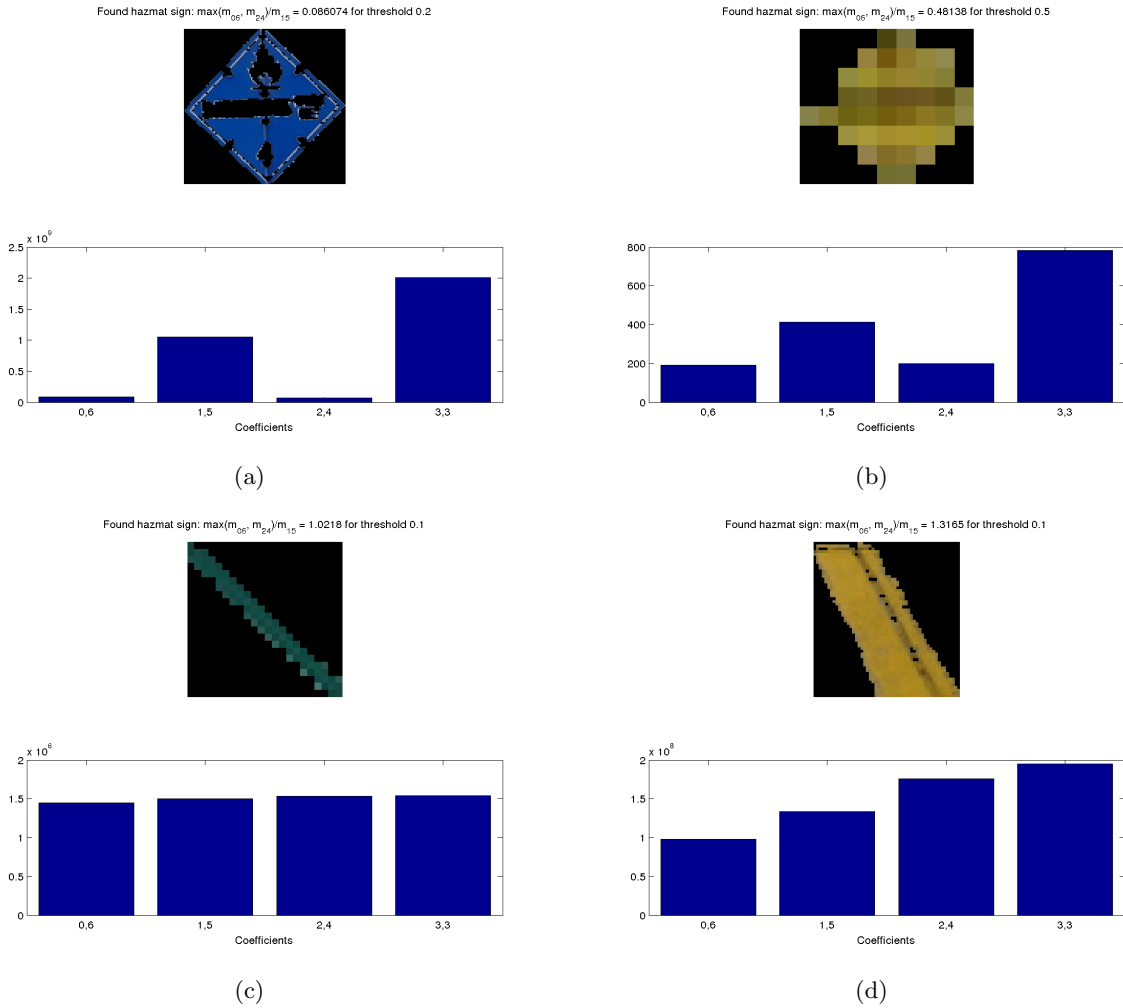

(c)   (d)

Figure 7. Component analysis from segmentation using 6-th central moment separates placard from non-placard components by comparing coefficients. Analysis completed using a 0.2 threshold
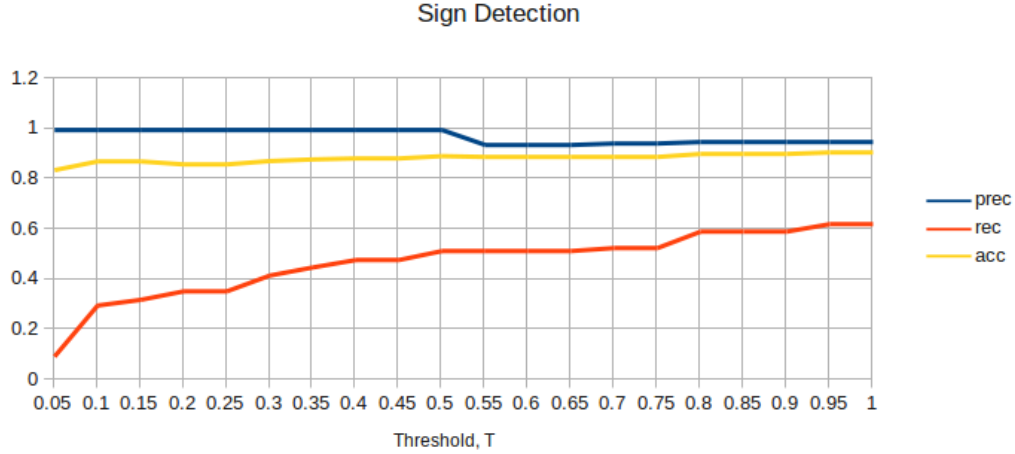
12

Figure 8. Precision, Recall and Accuracy for T from 0 to 1 at increments of 0.05, Max Accuracy: 90.8%

[5] Galigekere, R., Swamy, M., Fenster, A., "Moment Patterns in the Radon space," *Optical Engineering* **39**(4), 1088 − 1097 (2000).

[6] Galigekere, R., Swamy, M., Fenster, A., "Moment Patterns in the Radon space," *Optical Engineering* **40**(7), 1409 − 1411 (2001).

[7] Li, J., Pan, Q., Zhang, H. and Cui, P., "Image Recognition using Radon Transform," *Intelligent Transportation Systems* **1**, 741 − 744 (2003).

[8] Al-Shaykh, O., Doherty, J., "Invariant Image Analysis Based on Radon Transform and SVD," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing* **43**(2), 123 − 133 (1996).

[9] Hanbury, A., "A 3D-Polar Coordinate Colour Representation Well Adapted to Image Analysis," *Proceedings of the 13th Scandinavian Conference on Image Analysis* , 804–811 (2003).

[10] Lucchese, L. and Mitra, S., "Color Image Segmentation: A State-of-the-Art Survey," *Proceedings of the Indian National Science Academy (INSA-A)* **67 A**, 207–221 (2001).

[11] Newton, Sir Isaac , [*Opticks: or, A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*], Printed for William Innys at the West-End of St. Paul's, London (1704).

[12] Skarbek, W. and Koschan, A., "Colour Image Segmentation - A Survey," tech. rep., Technical University of Berlin, Department of Computer Science (1994).

[13] International Telecommunications Union, *Recommendation ITU-R BT.601, Encoding Parameters of Digital Television for Studios* (Geneva, Switzerland, 1992).

[14] Shima, J., "Fixed-pt. Atan2 with Self Normalization," (April 1999).

[15] Rosenfeld, A., "Connectivity in digital pictures," *J. ACM* **17**, 146–160 (January 1970).

13

HAZARDOUS MATERIAL SIGN DETECTION AND RECOGNITION

Albert Parra, Bin Zhao, Andrew Haddad, Mireille Boutin, Edward J. Delp

Video and Image Processing Lab (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

ABSTRACT

In this paper we describe two methods for hazardous material (hazmat) sign recognition. The first method is based on segment detection and grouping using geometric constraints. The second method is based on the use of a saliency map and convex quadrilateral detection. Our experimental results show a detection accuracy of 57.7% on a set of hazmat signs taken in the field under various lightning conditions, distances, and perspectives.

Index Terms— Sign detection, shape detection, saliency map, Hough Transform.

1. INTRODUCTION

Hazardous materials can react differently to environmental stimuli and cause problems in accidents and emergency situations and therefore makes these materials particularly dangerous to civilians and first responders. A federal law in the US requires vehicles transporting hazardous materials be marked with a standard sign (i.e., a "hazmat sign") identifying the type of material the vehicles is carrying [1]. These signs have identifying information described by the sign shape, color, symbols, and numbers. In this paper we describe two methods for hazmat sign detection and recognition. Each method detects (segments) the sign using shape information and then color information is used for sign identification. We first describe a sign recognition method based on segment detection and grouping using geometric constraints. Although this method is fast, it has several disadvantages. For example, low resolution images can cause missed straight edges at $\pm 45°$. In order to overcome these issues, we describe a second method that replaces the initial edge detection with a saliency map and combines contour detection and the Hough Transform. The second method is robust to rotation, perspective distortion, sign distance from the camera, distance between multiple signs, and blurred and low resolution images.

2. REVIEW OF EXISTING METHODS

Sign detection can be classified into three main categories: shape-based [2], color-based [3] and saliency-based [4]. Shape-based approaches first generate an edge map and then use shape information to find objects. Color-based approaches overcome the problems of shape variation, partial occlusion, and perspective distortion. However, colors are sensitive to lightning conditions and illumination

changes. Saliency-based approaches utilize selective visual attention models. A saliency-based visual attention (SBVA) model was presented in [4] using images features with a Gaussian pyramid. A graph-based visual saliency (GBVS) method was proposed in [5], to highlight conspicuous regions. A histogram-based contract (HC) method and a region-based contract (RC) method were introduced in [6] to construct saliency maps. HC-maps produce better performance over RC-maps but at the expense of increasing the computation time. A saliency map generation method was described in [7] using image signature (IS) to highlight sparse salient regions based on RGB or Lab color spaces.

Sign recognition methods can be classified into: geometric constraint methods, boosted cascades of features, and statistical moments[8, 9, 10]. Methods based on geometric constraints include the use of Hough-like methods [11, 12], contour fitting [13, 14], or radial symmetry detectors [15, 16]. Methods based on the boosted cascades of features commonly use the Viola-Jones framework [17, 18, 19]. These approaches often use object detectors with Haar-like wavelets of different shapes, and produce better results when the feature set is large. Methods based on statistical moments [20, 21, 22] use the central moments of the projections of the object to be detected. These methods are not robust to projective distortions or non-uniform lightning conditions.

3. HAZMAT SIGN DETECTION AND RECOGNITION

We developed two methods for hazmat sign detection and recognition. In the first method we detect and group segments using geometric constraints. In the second method we used saliency map to localize regions that potentially contain hazmat signs and then find the sign in these regions by checking for convex quadrilaterals. Both methods use color information for sign identification.

Segment Detection Using Geometric Constraints: We find edges in the image using the Canny edge detector. Since hazmat signs can be present at various distances, we use median auto-thresholding. To deal with non-uniform illumination changes in the scene, we also grayscale histogram equalize the image. We assume: 1) any sign in the image has to be approximately upright with its major axes aligned with the XY axis (Figure 2 illustrates the difference between an upright sign and a distorted sign); and 2) the projective distortion has to be small. (i.e., edges have to be approximately at $\pm 90°$ with respect to each other). Given these assumptions, we use morphological filters to eliminate edges not belonging to a hazmat sign. We create structuring elements at $\pm 45°$ and use them separately to erode the Canny edge map. The resulting edge map is the superposition of the two erosions. Signs that do not satisfy the two underlying assumptions will not preserve edges in

the resulting edge map. Once the edge map has been filtered, we find line segments using the probabilistic Hough Transform [23]. We set the minimum gap allowed between points on the same line to 5 pixels and the maximum gap to 5% of the maximum dimension of the original image (width or height). We next proceed to group the segments into candidates. Each candidate consists of a set of segments having one reference segment, at least one parallel segment, and two orthogonal segments (one to the left and one to the right of the reference segment). The reference segment is chosen at random from the list of segments that have not been grouped yet. Parallel segments need to have similar slope and length relative to the reference segment. The thresholds are set so that $|m_p - m_r| < 0.1$ and $|l_p - l_r| < 0.75e$, where m_p and m_r are the slopes of the parallel and reference segments respectively, l_p and l_r are the lengths of the parallel and reference segments respectively, and $e = max(l_p, l_r)$. The distance d between the reference and the parallel segments has to be in the range $0.5e < d < 2.5e$. This distance is defined between the middle points of the parallel and the reference segments. Also, the angle between the reference and the parallel segments has to be less than $20°$. This angle is defined by the normal of the parallel segment at its middle point and the vector joining the middle points of the parallel and the reference segments. Orthogonal segments need to have opposite slope and similar length to the reference segment, that is, $|m_p + 1/m_r| < 0.1$ and $|l_p - l_r| < 0.75e$. The distance d between the reference and the orthogonal segments has to be in the range $0.5e < d < 2.5e$. The angle between the reference and the orthogonal segments is defined as positive when the orthogonal segment is to the right of the reference segment, and defined as negative when the orthogonal segment is to the left of the reference segment. For each candidate set satisfying the geometric constraints we compute its minimal bounding box. We then discard any candidate with a bounding box aspect ratio smaller than 1.3. Finally, we check the remaining candidates and remove those that correspond to the same sign. This can be done by first dividing all bounding boxes that overlap more than 50% into groups, and then finding the optimal bounding box for each group. We consider the optimal bounding box to be the one with its nodes closest to its centroid (i.e., closest to a square). Figure 1 illustrates an example of the complete process. Once a hazmat sign is segmented, its color is set to the average hue inside the optimal bounding box and the color is used to identify the sign.

with a saliency map to detect regions potentially containing hazmat signs. The saliency map assigns higher saliency to more visually attractive regions. We used both the Lab and RGB color spaces as a combination of the IS-Lab and the IS-RGB methods in [7] to generate our saliency map. This method, which we call IS-RGB+Lab, gave us the best results in the experiments. We threshold this map to extract the most salient regions in the image. For each salient region found, we detect signs using specific color channels. This allows us to do both sign detection and color recognition at the same time, since we will assume that the color of any hazmat sign found in the region will correspond to the color channel associated to it. Currently the channels used are red, green, blue, and we also use the grayscale version to account for white or black signs. Note that the possible colors for hazmat signs also include yellow and orange, but these can be obtained by transforming the image from RGB to a hue-based color space and then segment the hue image. The grayscale and the color channels are then thresholded to account for highly chromatic areas using an empirically determined threshold. Each of the thresholded images is binarized, and morphologically opened to remove small objects containing less than 0.05% of the total number of pixels. We also use dilation to merge areas that may belong to the same object. We then retrieve the contours from the resulting binary image [24]. For each contour, we use the standard Hough Transform [25] to find straight lines that approximate the contour as a polygon. The intersections of these lines give us the corners of the polygon, which can be used to discard non-quadrilateral shapes. If the contour is approximated by four vertices, we find its convex hull [26]. If the convex hull still has four vertices, we check the angles formed by the intersection of its points. If each of these angles is in the range $90° \pm 1.5°$, and the ratio of the sides formed by the convex hull is in the range 1 ± 0.5, we can assume that we have found a convex quadrilateral. Finally, we use the same technique as in the first method to remove quadrilaterals that correspond to the same hazmat sign. Figure 4 illustrates a successful detection of two signs, one is affected by rotation and perspective distortion. Figure 5 illustrates a successful detection of one sign and also a false positive. In this particular case the issue could be addressed by using an optical character recognition to detect the text inside the sign candidate.

Fig. 1: First method (left to right): original image, segments at $\pm45°$, grouped segments, optimal bounding box.
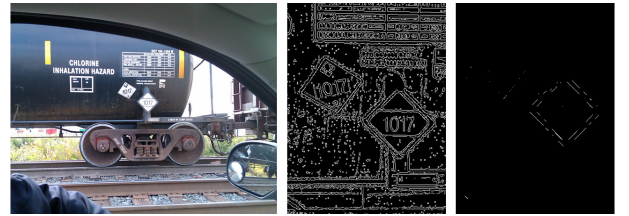
Fig. 2: Issue with first method: sign distortion.

Convex Quadrilateral Detection Based on Saliency Map: Our first method described above has some drawbacks. With low resolution images, the resulting edge map will not contain straight edges at $\pm45°$ and the erosion process will then delete most of them. The same happens with blurry images. Hazmat signs not satisfying the two assumptions of the first method will be removed during the erosion process, as shown in Figure 2. The gap threshold of the probabilistic Hough Transform may cause the segment grouping process to merge two segments from two close signs, as shown in Figure 3. Our second technique replaces the initial edge detection

Our second method offers multiple advantages. First, it is robust to rotation, since there is no erosion at $\pm45°$. Second, it is robust to perspective distortion, since convex quadrilaterals can be skewed. Third, it is able to detect signs close to each other, since there is no overlapping of line segments caused by the probabilistic Hough Transform. Fourth, it is more robust to blurred and low resolution images, since there is no edge detection is performed on the sign recognition step. Lastly, it is more robust to color recognition, since it detects signs already in specific color channels. The only disadvantage is its execution time.
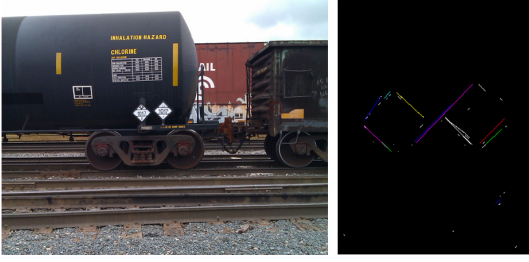
Fig. 3: Issue with first method: segment merging.

Fig. 4: Second method: true positives.

Fig. 6: Samples from the image dataset. From left to right: low resolution, blurred sign, shaded sign, geometrical distortion.

4. EXPERIMENTAL RESULTS

The first experiment consisted of images from a dataset and manually comparing the results with ground-truth information. The second experiment consisted of evaluating the saliency map methods. The tests were executed on a desktop computer with a 2.8GHz CPU and 2GB RAM. The ground-truth information included the sign distance from the camera, sign color, projective distortion of the sign, image resolution, possible shadow affecting the sign, and sign location on the image. Note that we only used the color and not the text of the sign for sign identification for these experiments. The image dataset consisted of 40 images each containing one or more hazmat signs (52 hazmat signs in total). The images were taken using three different cameras: a 8.2 Mpx Kodak Easyshare C813, a 16 Mpx Nikon Coolpix S800c, and a 5 Mpx camera on an HTC Wildfire mobile telephone. The images were acquired in the field, under various lightning conditions, distances, and perspectives. Among the 40 images: 17 were taken at 10-50 feet, 17 at 50-100 feet, and 6 at 100-200 feet; 3 had motion blur, 8 had geometrical distortion (i.e., perspective or rotation), 7 had shaded signs, and 6 had low resolution. Figure 6 illustrates some samples from the image dataset.

Table 1: Image Analysis Results.

Method	Sign	Accuracy	Color	Accuracy	Total
1	16	30.7%	6	11.5%	52
2	30	57.7%	22	42.3%	52

Table 1 shows the results of the first experiment using the two methods proposed in this paper. We determined how many signs were successfully detected (*Sign*) and how many were successfully identified (i.e., sign detected plus correct color (*Color*)). Note that the sign color recognition was done only if a sign was detected. Among the successfully detected signs we had a higher accuracy for color recognition. The first method recognized the correct color in 37.5% of the successfully detected signs, while the second method recognized the correct color in 73.3% of the successfully detected signs. The low accuracy of our first method is caused by multiple factors, including segment merging, edge detection failure on low resolution images, distortion and rotation of the sign, and multi-colored signs. However, note that multi-colored signs may also cause our second method to miss the detection, given that we detect signs at individual color channels. The first method had an average execution time of 2.3 seconds in total. The accuracy of the second method is influenced by the saliency map thresholding, the color recognition method based in specific color channels, and the morphological operations and Hough Transform on low resolution images. The second method had an average execution time of 5.1 seconds in total. Although the first method is faster, the second method doubles the sign detection accuracy, while still being fast enough to be used in real time applications.

Table 2 shows the results of the second experiment, including the average execution time and the score of the saliency map. The saliency map methods evaluated in the second experiment are: SBVA from [4], GBVS from [5], and IS-RGB and IS-Lab from [7]. Figures 7 and 8 illustrate examples of each method. Note that this process is only done as the first step of our second method. It can be

Fig. 5: Second method: True positive/False positive.
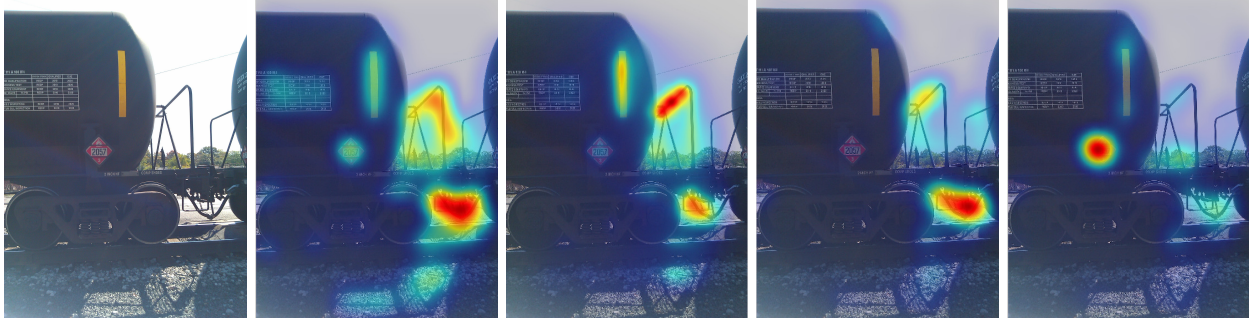
**Fig. 7**: Examples of saliency maps. From left to right: SBVA, GBVS, IS-RGB, IS-Lab.
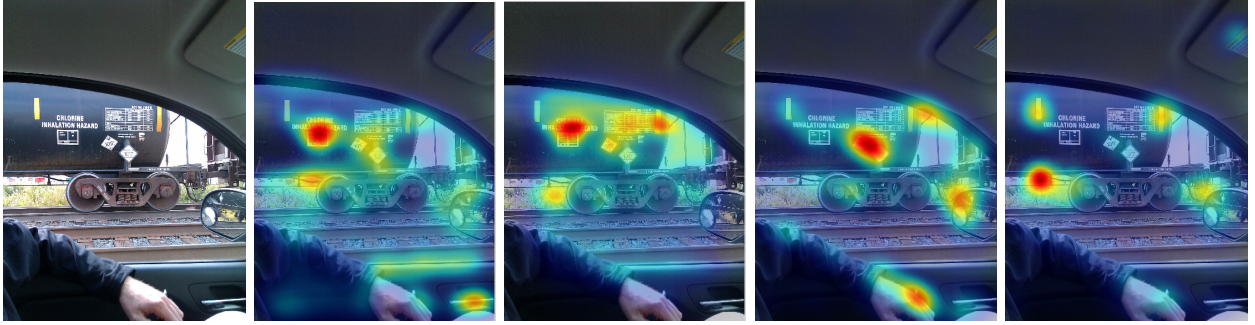
Fig. 8: Examples of saliency maps. From left to right: SBVA, GBVS, IS-RGB, IS-Lab.

Table 2: Average Execution Time, Distribution and Score of Each Saliency Map Method.

Method	Time	Good	Fair	Bad	Lost	Score
SBVA	1.92s	23	15	11	3	110
GBVS	3.10s	19	16	11	6	100
IS-RGB	0.40s	36	5	4	7	122
IS-Lab	0.41s	18	3	22	9	82
IS-RGB+Lab	0.81s	42	3	4	3	136

Fig. 9: Saliency map categories (top to bottom, left to right): original image, good, fair; original image, bad, lost.

seen in Figure 8 that although the IS-RGB saliency map covers two close signs in only one region, we can detect them separately (see Figure 4). We manually classified the saliency map results into four categories: good, fair, bad, and lost. Figure 9 illustrates examples of each case. For each of the 52 signs we assigned 3 points to a good map (sign was mostly contained in a high saliency-valued region), 2 points to a fair map (sign was mostly contained in a middle saliency-valued region), 1 point to a bad map (sign was mostly contained in a low saliency-valued region), and 0 point to a lost map (sign was not contained in any saliency-valued region). The score of each saliency map method is the sum of the points assigned to each image in the dataset, which ranges from 0 to 156 for each saliency map method. The IS method using the LAB or RGB color space has higher score and executes faster than the SBVA and the GBVS methods. The IS-RGB+Lab method based on two color spaces had the highest score, at the cost of increasing the execution time, with respect to the two IS methods using a single color space. However, the IS-RGB+Lab method still runs 2.37 times faster than the SBVA method and 3.83 times faster than the GBVS method.
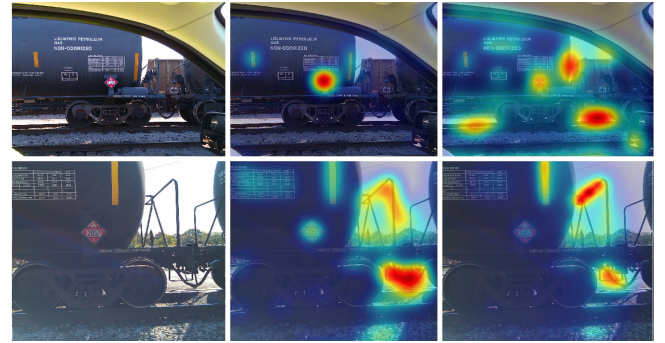
5. CONCLUSIONS AND FUTURE WORK

We described two methods for hazmat sign detection and recognition. The experimental results showed that our second method is more accurate for both sign detection and color recognition. We can further increase the accuracy by color calibrating the images to enhance the color recognition, developing a blur correction method to reduce the impact of blurred images, and developing an optical character recognition method to interpret the text inside the hazmat signs.

6. REFERENCES

[1] United States Department of Transportation, *Code of Federal Regulations, Title 49, DOT Hazmat*, Labelmaster, 2012 edition, October 2012.

[2] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1274–1286, October 2003.

[3] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, no. 3, pp. 453–464, March 1999.

[4] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, November 1998.

[5] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Proceedings of the Annual Conference on Neural Information Processing Systems*, pp. 545–552, December 2006, Vancouver, B.C., Canada.

[6] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 409–416, June 2011, Colorado Springs, CO.

[7] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, January 2012.

[8] R. Belaroussi, P. Foucher, J-P. Tarel, B. Soheilian, P. Charbonnier, and N. Paparoditis, "Road sign detection in images: A case study," *Proceedings of the International Conference on Pattern Recognition*, pp. 484–488, August 2010, Istambul, Turkey.

[9] A. Mogelmose, M.M. Trivedi, and T.B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, December 2012.

[10] K. L. Bouman, G. Abdollahian, M. Boutin, and E. J. Delp, "A low complexity sign detection and text localization method for mobile applications," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 922–934, October 2011.

[11] D.C.W. Pao, H.F. Li, and R. Jayakumar, "Shapes recognition using the straight line hough transform: theory and generalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1076–1089, November 1992.

[12] S. Houben, "A single target voting scheme for traffic sign detection," *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 124–129, June 2011, Baden-Baden, Germany.

[13] H. Fleyeh and Ping Zhao, "A contour-based separation of vertically attached traffic signs," *Proceedings of the Annual Conference of Industrial Electronics*, pp. 1811–1816, November 2008, Orlando, FL.

[14] L.-W. Tsai, J.-W. Hsieh, C.-H. Chuang, Y.-J. Tseng, K.-C. Fan, and C.-C. Lee, "Road sign detection using eigen colour," *IET Computer Vision*, vol. 2, no. 3, pp. 164–177, September 2008.

[15] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 959–973, August 2003.

[16] N. Barnes, A. Zelinsky, and L.S. Fletcher, "Real-time speed sign detection using the radial symmetry detector," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 322–332, June 2008.

[17] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[18] C.G. Keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. Baratoff, "Real-time recognition of U.S. speed signs," *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 518–523, June 2008, Eindhoven, Netherlands.

[19] X. Baro, S. Escalera, J. Vitria, O. Pujol, and P. Radeva, "Traffic sign recognition using evolutionary Adaboost detection and Forest-ECOC classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 113–126, March 2009.

[20] A.R. Rostampour and P.R. Madhvapathy, "Shape recognition using simple measures of projections," *Proceedings of the Annual International Phoenix Conference on Computers and Communications*, pp. 474–479, March 1988, Scottsdale, AR.

[21] P. Gil-Jimenez, S. Lafuente-Arroyo, H. Gomez-Moreno, F. Lopez-Ferreras, and S. Maldonado-Bascon, "Traffic sign shape classification evaluation. Part II. FFT applied to the signature of blobs," *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 607–612, June 2005, Las Vegas, NV.

[22] A. W. Haddad, S. Huang, M. Boutin, and E. J. Delp, "Detection of symmetric shapes on a mobile device with applications to automatic sign interpretation," *Proceedings of the IS&T/SPIE Electronic Imaging on Multimedia on Mobile Devices*, vol. 8304, January 2012, San Francisco, CA.

[23] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A probabilistic hough transform," *Pattern Recognition*, vol. 24, no. 4, pp. 303–316, February 1991.

[24] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, April 1985.

[25] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, January 1972.

[26] Jack Sklansky, "Finding the convex hull of a simple polygon," *Pattern Recognition Letters*, vol. 1, no. 2, pp. 79–83, December 1982.

Location-Aware Gang Graffiti Acquisition and Browsing on a Mobile Device

Albert Parra,[a] Mireille Boutin,[b] Edward J. Delp[a]

[a]Video and Image Processing Lab *(VIPER)*
[b]Computational Imaging Lab *(CIL)*
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

ABSTRACT

In this paper we describe a mobile-based system that allows first responders to identify and track gang graffiti by combining the use of image analysis and location-based-services. The gang graffiti image and metadata (geoposition, date and time) obtained automatically are transferred to a server and uploaded to a database of graffiti images. The database can then be queried with the matched results sent back to the mobile device where the user can then review the results and provide extra inputs to refine the information.

Keywords: gang graffiti,mobile telephone, geolocation, interactive map

1. INTRODUCTION

Gangs are a serious threat to public safety throughout the United States. They are responsible for an increasing percentage of crime and violence in many communities.[1] Street gang graffiti is their most common way gangs use to communicate messages, including challenges, warnings or intimidation to rival gangs. It is an excellent way to track gang affiliation and growth.

First responders have the potential for finding and documenting graffiti evidence in real time. However, the number of actions that can be taken while on the street are minimal. If there is an incident, or they need to compare information, they have to communicate with the local police department. For example, if a new graffiti is spotted by a first responder, the information that can be obtained in situ is very limited. In the best case scenario, the first responder may have expertise on gang graffiti interpretation and carries a camera. The only actions they can take are reduced to taking an image and noting some basic contextual information.

Our goal is to develop a mobile-based system capable of using location-based-services, combined with image analysis, to provide accurate and useful information to a first responder based on a database of gang graffiti images. We call this system Gang Graffiti Automatic Recognition and Interpretation or GARI.[2] The analysis includes using metadata (geoposition, date and time) obtained at the time a gang graffiti image is acquired and then using image analysis methods to extract information from the graffiti image for interpretation and indexing. The information is stored in a graffiti image database. The database can then be queried with the matched results sent back to the mobile device where the user can then review the results and provide extra inputs to refine the information.

In addition to being able to send and retrieve multimedia data to the database, the first responder can take advantage of the location-based-services that the mobile device provides. The graffiti database can be filtered to retrieve data in a specific radius from the current location of the user. The data includes not only the images, but information related to it, such as date and time, geoposition, gang, gang member, colors, or symbols. This can be used to keep track of gang activity in the area, such as showing the geolocation of the graffiti on a map

or using augmented reality to show the position of the graffiti from the viewpoint of the user. By providing first responders with this mobile-based capabilities, the process of identifying and tracking gang activity is made more efficient. This can lead to a faster intervention by law enforcement.

There are methods currently used to identify gang graffiti using feature matching, as well as to keep track of gang graffiti using large databases. Below we describe two of the current methods, indicating their advantages and disadvantages, and how they compare to GARI.

Graffiti-ID is an ongoing project at Michigan State University.[3] The project is focused on matching and retrieval of graffiti images. There is non-published work which extends the project to gang and moniker identification.[4] The goal of Graffiti-ID is to identify gang/moniker names related to a graffiti image based on visual and content similarities of graffiti images in a database. The system consists of two modules, one for populating the database (offline) and another for querying and obtaining results from the database (online). The offline module includes two processes. First, automatic feature extraction using the Scale Invariant Feature Transform (SIFT).[5] Second, manual annotation of graffiti images by letters and numbers. This is done on images taken from an external gallery of images with the information stored in a database. The online modules includes manual annotation of input images to filter the database and SIFT feature extraction to obtain keypoint matching.

The image database used is based on the Tracking Automated and Graffiti Reporting System (TAGRS) from the Orange County Sheriff Department in California. The database consists of 64,000 graffiti images with the main sources of the images are the Orange County Transportation Authority and crime reports. A subset of 9,367 images were used for evaluation. Each of these images contains up to four information parameters: moniker, gang, date and time, and address.

Graffiti Tracker is a web-based system that began in 2002.[6] It was designed to help first responders identify, track, prosecute and seek restitution from graffiti vandals. It is primarily used by law enforcement and public works agencies. The database contains more than 2 million manually analyzed graffiti images from 75 cities in two countries and nine states, mainly from the state of California. The web-based services include graffiti analysis, interactive map browsing, graffiti storing and organization, and graffiti report. Graffiti Tracker provides clients with GPS-enabled digital cameras to generate reports of graffiti activity. The images can then be uploaded through the web interface to the database, they are manually analyzed by trained analysts within 24 hours of submission. The GPS coordinates of each image are used to build an interactive map where the user can view activity from individual vandals or monikers to specific crews or gangs. Gang trends or migration can be identified if the volume of graffiti for the same gang or vandal is large. A part from the interactive map, the user can browse the stored graffiti by moniker, gang, type of incident, graffiti surface, or removal method. The information can be used to generate reports based on gang or moniker activity, such as total square feet of damage, locations of the incidents, or frequency of graffiti vandalism over a specific period of time.

Although our proposed system, GARI, shares some goals with both of the above systems, our methodology is somewhat different. Both Graffiti-ID and GARI have goals of identifying gangs and gang members based on the graffiti content. Graffiti-ID uses SIFT features. GARI currently uses color recognition techniques along with metadata information from an image to query the database.[2] Future goals of GARI include the use of SIFT features to detect if an image of a same graffiti was already acquired at a specific location, and also the use of shape techniques to detect graffiti components. Both Graffiti Tracker and GARI keep track of gang activity based on GPS tags from the images and the graffiti content. However, Graffiti Tracker image analysis is done manually.

In Graffiti Tracker, image analysis is performed manually by trained analysts with the results obtained within 24 hours of submission. The goal of GARI is to perform the analysis in the field, automatically and in real-time, either on the device or on the server. Graffiti-ID uses SIFT features to match images on the server automatically, but the analysis of the content of the graffiti is done manually, by labeling the image. Moreover, it allows for the labels to be numbers (0-9) or letters (a-z), not symbols or other features such as color.

Graffiti-ID does not provide any type of gang activity tracking, while both Graffiti Tracker and GARI provide interactive maps that allows first responders to browse the database and keep track of specific gangs or individuals. The advantage of GARI is that it also provides additional methods for tracking gang activity, including browsing the database by radius from specific locations, or by graffiti color/s. One advantage of Graffiti Tracker is that

its database is currently dramatically larger than the GARI database. Therefore, the results retrieved from the Graffiti Tracker database can indicate more accurate gang activity.

In summary, our system combines features from both Graffiti-ID and Graffiti Tracker and adds more services and functionality. The advantages our system has over Graffiti-ID and Graffiti Tracker are the following. We provide a mobile application that lets first responders act in the field, where the graffiti is located, and upload and browse the database of graffiti in situ. The image acquisition in our system is device independent; any image from any type of camera can be uploaded using either of our supported platforms: an Android-based mobile telephone or through our web-based interface.

2. SYSTEM IMPLEMENTATION

We implemented a prototype of the GARI system as an application for Android devices and as a web-based interface accessible from any web browser. Figure 1 illustrates the GARI system, which is divided in two groups:

1. **Client-side**: Performs operations on the Android device and communicate with the database of gang graffiti through either the WiFi or 3G networks.

2. **Server-side**: Performs operations on the database of gang graffiti and communicate with the client.

The client-side includes the device and methods available to the users, either to operate without the use of a network connection (offline services) or to make queries to the database (online services). The offline services are only available from Android devices. The online services are available from both Android devices or any web browser. This includes desktop and laptop computers as well as iPhone and Blackberry smartphones. The server-side includes all operations performed on the server including image analysis and queries to the database from both the Android application and the web-based interface. The database comprises gang graffiti images and metadata information for each entry, such as EXIF data, image geolocation and the results of the image analysis on each image whether it was performed on the server or client.
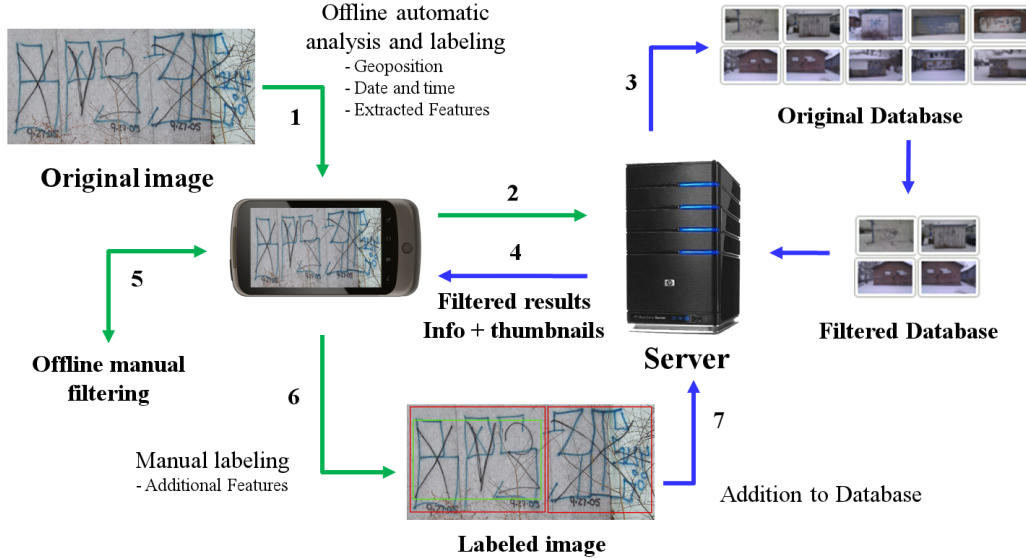
Figure 1: Overview of the GARI System - Client-Side Components (green) and Server-Side Components (blue).

2.1 Database of Gang Graffiti

In this section we describe how the image database is organized. We will first describe the database schema and then show by an example how the information GARI acquires is added to the database. The database of gang graffiti was deployed for three reasons:

1. To collect and organize graffiti images acquired by first responders. This includes the images, metadata, and any interpretation or other information provided by the first responder.

2. To store the results of the image analysis.

3. To manage first responders' credentials, allowing them to access the services available through the Android application and the web based interface.

Our database is implemented in PostgreSQL[7] on a Linux server. It consists of eight tables structured as shown in Figure 3. Note that the schema does not show all the fields in all the tables but just the relevant fields to indicate the association between the tables. A more complete list is described in the reference.[2] The various IDs mentioned below (e.g., image ID) will be discussed in more detail after the tables are described in the following list.

1. **images**: Stores EXIF data from the images along with image location and general image information and the results from the image analysis. This data is distributed along a total of 51 fields.

2. **imageColors**: Stores all color IDs related to each image ID. This table is especially useful when more than one color is found in the same graffiti image.

3. **colors**: Stores the association between color IDs and color names.

4. **imageBlobs**: Stores the number of blobs in each graffiti, the ID of each component for each blob, and the color ID of each component. This also stores special attributes of components. These attributes may include a specific component being crossed-out, upside-down, etc.

5. **blobComponents**: Stores the association between component IDs and component names, as well as the type ID for each component. Each component belongs to any of the following types: symbol, character, number, acronym, nickname, string.

6. **componentTypes**: Stores the association between type IDs and type names.

7. **gangComponents**: Stores the association between gang IDs and gang names, as well as the component ID (or multiple component IDs) associated with each gang. This table is especially useful when more that one component is associated with the same gang name.

8. **users**: Stores users' credentials to access to the system services as well as information concerning administrative privileges, email addresses, and registration and login status.

2.2 Adding Images to the Database

The following example illustrates the process of adding a graffiti image to the database. Figure 2 shows the example image that has been manually labeled to facilitate the explanation. Each labeled circle represents a blob and each blob contains a distinguishable component of the graffiti. The blob labeling of the image corresponds with the field *blobID* from table *imageBlobs* in the database.

First, we fill table *imageColors* with the colors found in the graffiti. This is, black, green, and blue. Second, we analyze the blobs separately. Blob 1 contains a black X3; blob 2 contains a green SPV; blob 3 contains a blue X3; blob 4 contains a blue LK crossed-out in green; blob 5 contains a blue ES crossed-out in green.

Note that the meaning of the acronyms and the type of the components is not addressed here. This information is assumed to already exist in the database.

Once the image analysis is complete the image, along with the blob information, is added to the database. Figure 3 shows the database fields filled with the information obtained from the graffiti in Figure 2. First, the user ID of the first responder who captured the image and the image ID are added to the *images* table. The image ID is a unique identifier of the graffiti image and it is automatically updated every time an image is uploaded to the server. Although it is not shown in Figure 3, some additional image information (i.e., EXIF data, GPS coordinates) is extracted from the uploaded image and added to the *images* table. Second, the color IDs for the three colors found in the graffiti, which are obtained by checking the color description field, (labeled *colorName* in Figure 3), are added to the *imageColors* table, and linked to the graffiti ID. At the same time, the five blobs are added to the *imageBlobs* table. Each blob has a corresponding component ID, which is obtained by checking the component description field, (labeled *compName* in Figure 3), of the *blobComponents* table. Each component has a color associated with it and can activate one or many attributes in the same table. In this example, blobs one to three do not have any additional attribute. Blobs four and five have activated the crossed-out attribute.

Note that this process is totally objective. That is, the information uploaded to the database does not require any interpretation from the first responder. With all the objective information available in the tables and the associations between the data one can produce an informed graffiti interpretation. For example, we have added components with IDs 27 (*SPV*) and 29 (*LK*). These IDs are associated with specific gang names in the *gangComponents* table. The same reasoning could be used if the graffiti did not contain any specific content with just the graffiti color being identified. Additional tables can relate gang IDs with color IDs effectively providing the results of gangs matching the specific color or colors.
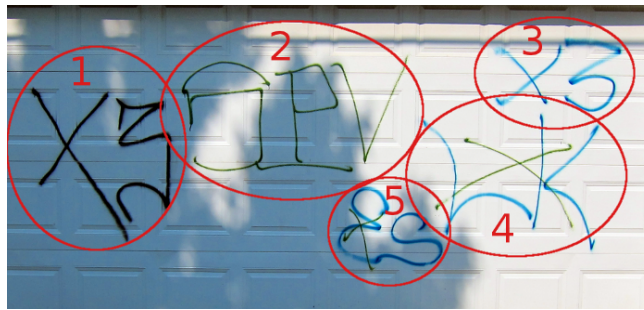
Figure 2: Example of Graffiti (manually labeled).

2.3 Android Implementation

We implemented the GARI system on an Android device as summarized in Figure 1. In this section we describe how the Android application works and describe its user interface.

A user takes an image of the gang graffiti using the embedded camera on the device via the Graphical User Interface (GUI). The EXIF data of the image, including GPS location and date and time of capture, is automatically added to the image header. The user can then choose to upload the image to the server to be included in the database of gang graffiti or do color recognition. The first option, uploading to the server, allows the user to send the image and the EXIF data to the server creating a new entry in the database. The second option, color recognition, allows the user to trace a path in the current image using the device's touchscreen. The color in the path is then automatically detected and the result is shown to the user. The database of gang graffiti can then be queried to retrieve graffiti images of the same color.

Another option is to browse the database of gang graffiti given various parameters such as the distance from current location or date and time. The thumbnail images that match the query are downloaded from the server and shown to the user on the mobile telephone. The user can then browse the results to obtain more information about the specific graffiti. Note that in order to browse the database of gang graffiti a network connection is required.

We implemented the system on a HTC Desire mobile telephone (1 GHz CPU, 576 MB RAM) running version 2.2 of the Android operating system.
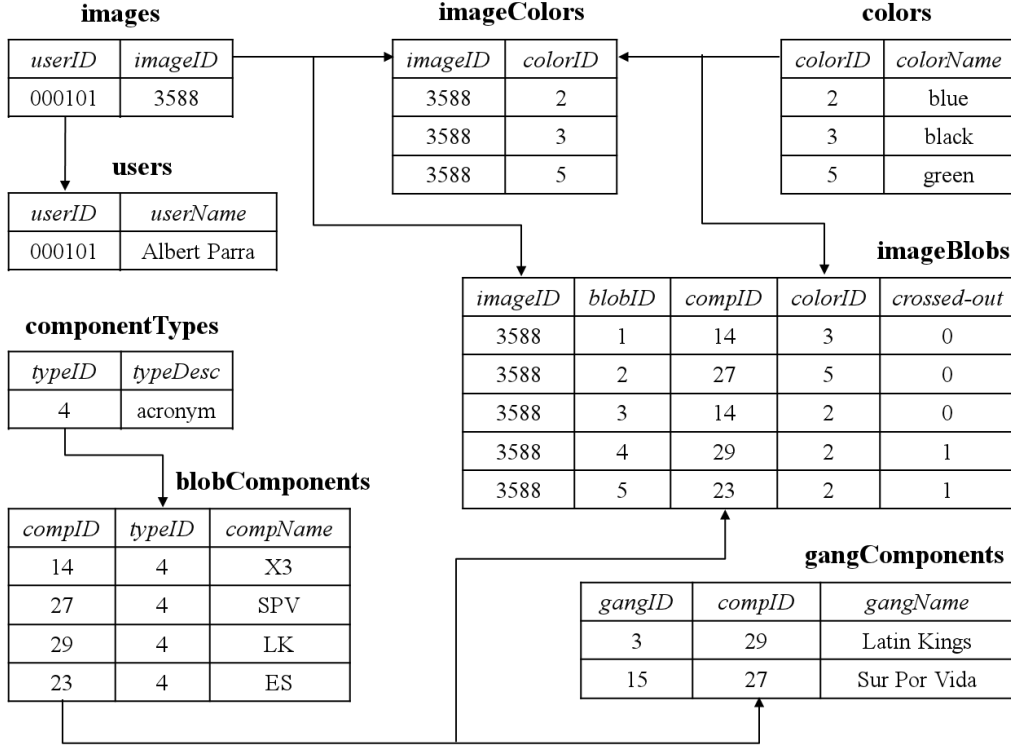
images

userID	imageID
000101	3588

imageColors

imageID	colorID
3588	2
3588	3
3588	5

colors

colorID	colorName
2	blue
3	black
5	green

users

userID	userName
000101	Albert Parra

imageBlobs

imageID	blobID	compID	colorID	crossed-out
3588	1	14	3	0
3588	2	27	5	0
3588	3	14	2	0
3588	4	29	2	1
3588	5	23	2	1

componentTypes

typeID	typeDesc
4	acronym

blobComponents

compID	typeID	compName
14	4	X3
27	4	SPV
29	4	LK
23	4	ES

gangComponents

gangID	compID	gangName
3	29	Latin Kings
15	27	Sur Por Vida

Figure 3: Database Fields with Information from the Graffiti in Figure 2.

2.3.1 User Interface

Our Android application does not require the use of a network connection. However it is mandatory if the user wants to browse the graffiti database or upload images to the graffiti database. A user must be assigned a User ID (equivalent to a First Responder ID) and a unique password in order to use GARI.

In this section we make the distinction between hardware and software keys in the application to describe some of the actions. Basically, hardware keys belong to the device, while software keys are created by the application. The user can interact with the device and the database of gang graffiti using the options described below. Some are only available when the user has captured or browsed an image.

Capture Image

The menu option "Capture Image" starts the image capture. The user can then take an image of the graffiti as shown in Figure 4a. Once the image is acquired, the user can choose either to retake the image or to tap "Done" to continue with the process.

The application then checks for the device location automatically in order to add the GPS coordinates to the image. Depending on the location system used (Network(GSM/WiFi) or GPS), it can take up to 30 seconds to acquire the location. If location services are not enabled on the device, the user is notified and taken to the location settings where the location system can be enabled.

If, despite having the location systems enabled, the location system times out, a dialog will notify the user that the location cannot be determined. It also recommends the user to manually save the location information.

When the location has been acquired (automatically or manually), the user is given the option to crop the image. The image can be cropped by scaling the orange rectangle as shown in Figure 4b. When the desired graffiti contents are contained in the cropped area, the image can be cropped by tapping "Save." The user will be returned to the main screen and the image will be set as background.

6

(a) Camera Activity. (b) Crop Dialog.

Figure 4: Image Capture.

Browse Image

If an image has been acquired without using the GARI application, or if it has been taken with the the GARI application but not analyzed or sent to the server, the user has the option "Browse Image." When tapped, a directory browsing window is opened and the user can search and select the desired image.

Browse Database

The menu option "Browse Database" allows the user to browse the graffiti database by radius or distance from their current location. The system finds in the database the images in a chosen radius from the current location. The user can select a radius between 1 mile and 20 miles. If the option "All" is selected the application use all the images in the database. If a specific radius is chosen, the application has to first acquire the user's current location. Then, the application contacts the graffiti server and checks how many graffiti thumbnail images have to be downloaded. If the user accepts the information that matches the query is retrieved.Figure 5a shows an example of the results, where each line contains a thumbnail image of a graffiti and basic information about it, including the date and time the image was acquired, and its GPS latitude and longitude. To obtain more information about a particular graffiti, the user can tap either the thumbnail or the text field, and the application will contact the server, retrieving a larger image and the information available. Figure 5b shows an example of the extended results. Currently, the text field includes the information from the *images*' table on the database. The menu hardware key has the option "Show in map". This allows the user to see the location of one or multiple graffiti on a map (Figure 5c).

Send to Server

Once the user captures an image using the "Capture Image" menu option or browses an image from the device using the "Browse Image" menu option, the menu option "Send to Server" is enabled in the main screen. This option allows the user to send the current image to the graffiti server. The application will contact the server and check if the image has been previously uploaded. If not, the image is sent to the server. If the image is successfully added to the graffiti database the application will indicate this to the user. The graffiti information is downloaded from the server and displayed to the user.

Manual Input

The user can analyze the graffiti using the option "Manual Input." This options allows a user to annotate the image manually and then store the information in the database. Figure 6a shows a graffiti image taken by a first responder. This menu option allows the user to select a symbol and its color. Figure 6 illustrates the manual input steps that identify the black 6-point star in Figure 6a. This process can be repeated for any symbol, number, or color in the graffiti. The information is saved and later sent to the database.

7
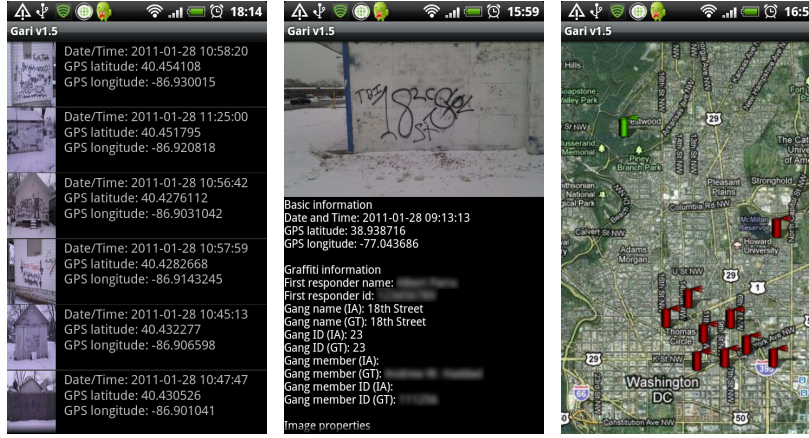
(a) Browse results.　　　(b) Extended results.　　　(c) Interactive map.

Figure 5: Browse Database - Retrieving Results.
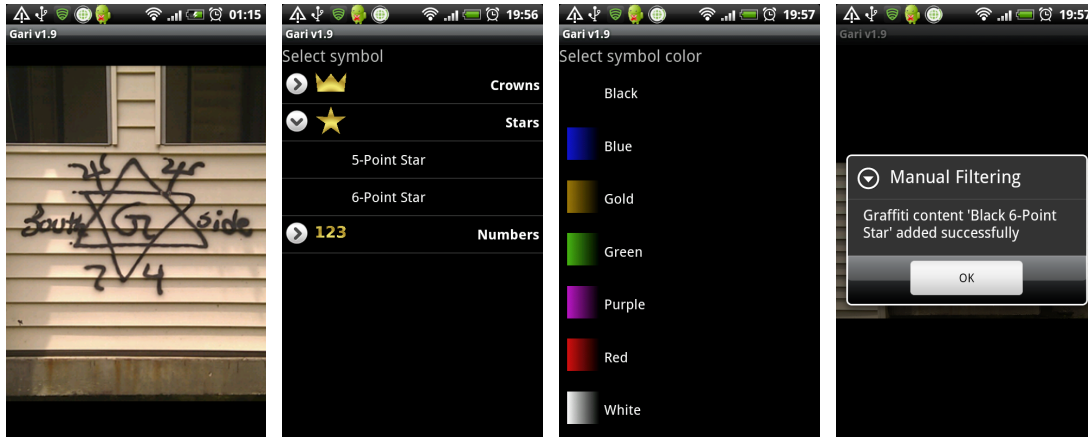
(a) Input image.　　　(b) Select symbol.　　　(c) Select color.　　　(d) Image labeled.

Figure 6: Manual Input - Labeling An Image Containing a Black 6-Point Star.

2.3.2 Security

Our Android application is used by first responders from multiple agencies. Therefore, it is necessary to ensure that only authorized users can access and use the application. The connections to the server must be secure and all the information transmitted to and from the server must be encrypted (using the SSL/TLS protocol). The user credentials are sent every time the application contacts the server to make sure the connection is made by an authorized user. An additional level of security includes the creation of two types of users:

- **Regular users**: Can switch between users, change their password, delete specific images only taken by themselves, and send crashlogs.

- **Administrative users**: Can modify the server domain name/IP address, change user IDs, change passwords, delete specific images from any user, delete all images of any specific user, and send crashlogs.

When launching the GARI application a dialog box automatically prompts the user for login credentials. The user is required to input a user ID and a password.

The first time a user logs in the credentials are checked with the server and once they are validated they are stored in the device in an encrypted file. This allows the user to use the application without needing a network

connection. Note that passwords are never stored as plaintext, neither on the device or the server. They are hashed using an MD5 cryptographic hash function[8]

All authorized users can access the "Settings" option from the main screen of the application. Note that no one can delete images from the server. At this time no one can edit the attributes of images retrieved from the server.

- **Server domain/IP**: Specifies the domain name or IP address of the server. The server is contacted to send images, browse the contents of the database, as well as to log in and change the login password. This option is only available for administrative users.

- **Switch user**: Allows a different user to log in and use the application.

- **Change password**: Allows the password of the current user to be changed. The new password is hashed and stored back in a system encrypted file for future logins. Note that this requires the application to contact the server requiring a network connection.

- **Delete specific images**: Allows specific images acquired by the user to be deleted. Administrative users can see and delete other users' images.

- **Delete user's images**: Allows all images taken by the user to be deleted. Administrative users can delete images from other users.

- **Delete all images**: Deletes all images. This option is only available for administrative users.

- **Send crashlog**: Allows the user to send crash feedback to the server when the application crashes. This helps us analyze and keep track of application errors.

2.4 Web Interface

We also implemented our system as a web interface that gives a user access to the graffiti in the database and provides the ability to upload, modify and browse most database contents. The diagram for the web-based interface is the same as for the Android application (see Figure 1). However, note that the platform to interact with the server is now any device with an Internet browser and the client-side services now correspond to services available through the website.

The user logs in into the "Archive" using authorized credentials. Note that the credentials are the same for both the Android application and the web services. The user can then either browse the database of gang graffiti or upload an image. If the choice is to browse the database, the user can check the graffiti images and their attributes or filter the database using parameters such as radius from a specific location or address, capture data, upload data, or modified date. The results are shown as a list of thumbnail images with basic information that identifies the graffiti image. The user can then browse specific images and place them on a map, so to visually track gang activity. If the choice is to upload an image, the user can select a graffiti image from their local system (i.e., any device with a web browser). Some attributes can be adjusted through guided steps before adding the information to the database, such as location, gang information, or additional comments.

The web interface is available from any device with a web browser. This includes all desktop and laptop machines and all mobile telephones capable of browsing the web (e.g., iPhone, Blackberry, Android devices). In some cases, the current location of the user is required in order to retrieve results from the database of gang graffiti such as when using the "radius" function to display graffiti on a map. Geolocation was introduced with HTML5 and it is widely implemented by many modern browsers. However, only the latest browsers support this service.

2.4.1 User Interface

Through the "Archive" link on the left sidebar the user can browse the entire database of gang graffiti. Once the user has logged in using authorized credentials, the available options are "Browse database" and "Upload image."

Browse Database

The "Browse database" page allows the user to either browse the entire graffiti database or to do a specific search given parameters. This include search by radius, search by date (captured, uploaded or modified), and search by address. When searching by radius the system retrieves from the database all the graffiti in a specific radius from the user's current location. When searching by address, we use the Google Maps API to geocode the location and query the database. The results are formatted as shown in Figure 7. At first, only a small-scale image and basic information is displayed. Depending on the search different parameters are shown including the date and time the graffiti image was captured, uploaded or modified; the address were the graffiti was found; and extended graffiti information.

Figure 7: The Results from Browsing The Database.

Each graffiti or group of graffiti can be placed on an interactive map to visually track the results of a search. Figure 8 shows an example of the interactive map. Each marker represents the location of a graffiti from the search results. From this map the user can click on any of the markers to see a thumbnail of the graffiti image, its location in GPS coordinates, and a link to obtain more information about the graffiti.

When "More information" is clicked, either from the list of results or from the interactive map, the user can see the information available in the database for the specific graffiti.

Upload Image

The "Upload image" page lets the user upload an image from the machine that is accessing the website. First, the user chooses the file to upload, which is previewed before actually adding any entries to the database Once the image is uploaded, its EXIF data are automatically extracted, and a new entry is created in the database. If the device used to acquire the image did not have a GPS receiver, the location can be manually assigned in the next step.The user can input an address, which is geocoded and shown on an interactive map. More accurate
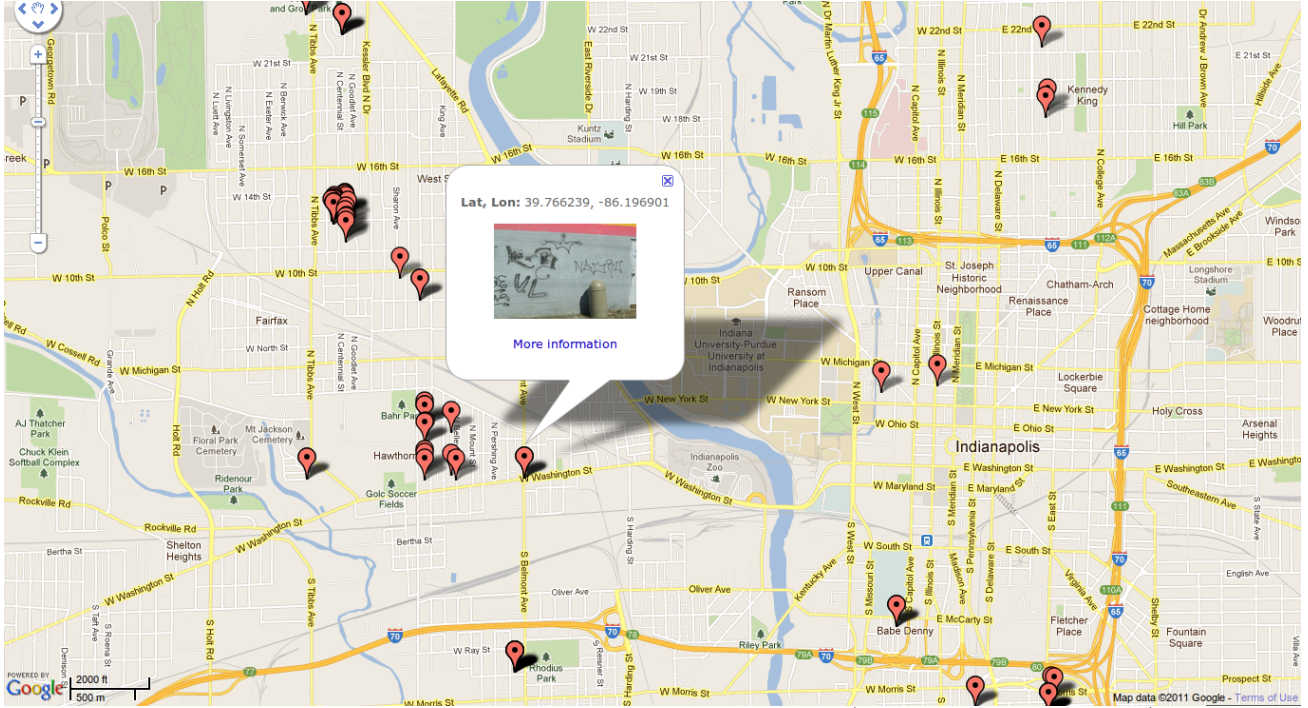
Figure 8: Interactive Map with the Graffiti Locations Noted (Map data ©2001 Google).

GPS coordinates can be obtained by directly clicking on the interactive map. Besides the GPS coordinates other information is required in order to uniquely identify the image uploaded. This include the first responder name and first responder ID. Given that the user has accessed the database of gang graffiti using authorized credentials, the first responder ID is automatically obtained. The first responder name is associated to this ID in the *users* table in the database. Finally, the user can input additional information to help analyze the graffiti. This include fields for gang name, gang member, and additional comments. The gang name can be completed using a drop-down list of known-gangs, or entering it manually. The gang member is currently entered manually. The additional comments include any kind of information that does not fit in any of the previously described fields, such as symbols or colors found, graffiti meaning, or relative location of the graffiti with respect to the surrounding streets.

Clicking on "Submit Image" completes the editing step and shows the user the final output. This output is the same as when the user uses the "More Information" option when browsing the database.

2.4.2 Security

Access and navigation to the web interface are established and managed using encrypted Secure Sockets Layer (SSL) sessions. SSL encrypts information both during the transmission. The user must log in using authorized credentials before entering the archive. Once successfully logged in an SSL session is created and maintained for the current user. The user account can be managed by clicking on the "User Settings" link on the left sidebar. Note that currently the only option available is password change.

3. EXPERIMENTAL RESULTS

In this section, we investigate the execution time and memory size of our database of gang graffiti images as well as some of our processes on the hand-held device. We tested the execution time of database queries from a HTC Desire Mobile telephone with the Android 2.0 OS. The battery life of the hand-held device is not yet considered an issue but may pose problems in the future if we add more processing functions to the Android application. Note that we have not performed any battery consumption tests, since we do not currently perform computationally intense operations on the HTC Desire.

3.1 Execution Time

We tested the database query performance of our system. That is, we tested the elapsed time between sending an image from the hand-held device, using the Android application, and receiving the results of the upload. On the client side, the process includes sending and receiving the image to the server via HTTPS and returning the graffiti image thumbnail and text retrieved to the user. On the server side, the process includes creating a session for the user, checking image existence in the database, copying the image to a specific directory, creating the thumbnail image and reduced size copies of the image, extracting up to 24 EXIF data points from the image, creating a new entry in the PosgtreSQL table and adding information in as many as 30 fields, and sending back a string with the results of the upload. Table 1 shows the details of ten graffiti image uploads using the same network conditions (WiFi). As one can see most of the elapsed time is due to the HTTPS connection since the user interface operations on the hand-held device (for the specific action of uploading an image to the server) do not slow down the process.

Table 1: Execution Time On the Hand-Held Device and the Server When Uploading an Image.

Image Size	Server Time	Total Time
146.7 KB	0.66 s	2.24 s
157.9 KB	0.65 s	2.33 s
179.8 KB	0.65 s	2.66 s
203.3 KB	0.66 s	2.42 s
207.9 KB	0.64 s	2.44 s
227.8 KB	0.65 s	2.34 s
609.9 KB	1.05 s	3.64 s
639.8 KB	1.47 s	4.71 s
653.6 KB	1.06 s	4.00 s
760.4 KB	1.07 s	4.31 s

3.2 Memory Size

We computed the memory size of the images in the gang graffiti database as well as the memory size of the Android application.

We cooperated with the Indianapolis Metropolitan Police Department to acquire graffiti images. This allows us to be able to accurately calibrate and analyze the images. Currently we have 657 images from the city of Indianapolis. These include images acquired with and without using a tripod and with and without fiducial markers. We used three digital cameras for this purpose: a 10Mpx Canon Powershot S95, a 4Mpx Panasonic Lumix DMC-FZ4, and a 5Mpx HTC Desire (Android mobile telephone). From these images, 151 images are currently contained in the database of gang graffiti. The rest are for research purposes. As of December 2011, a total of 38 users have entered 348 graffiti images to our system, resulting in a total of 499 graffiti images that can be browsed from both the Android application and the web interface. The 499 images are stored in the server and all have a thumbnail and a reduced size version. This makes a total of 1,497 images adding up to 588 MB of data. Taking into account the 506 images not in the database (containing fiducial marks - for research purposes) we have a total of 2,003 images adding up to 866 MB of data.

The Android application on the hand-held device consists of an Android application package file (APK) of 1.8 MB. This includes the compiled code and the multimedia used when the user is not connected to a network. When contacting to the server the contents of the database can be browsed. If the users chooses to retrieve thumbnail images of all the 499 browsable images in the database, 5.67 MB are added to the application, making a total of 7.47 MB.

4. CONCLUSIONS AND FUTURE WORK

As of December 2011 we have developed an Android application and a web-based interface for the GARI system. Our tests on database query performance suggest that the bottleneck for the upload and retrieval process is the network connection. This is due to the fact that an entire image is sent to the server using HTTPS which highly depends on the network speed. Our database of gang graffiti images contains 499 browsable images with associated thumbnails and reduced size versions. These 1,497 images are 588 MB of data. We have also acquired a total of 506 images for research purposes. The Android application has a memory size of 1.8 MB on the hand-held device. If all 499 thumbnail images from the database are downloaded for browsing, the application would take 7.47 MB of data.

Our long term goal is to develop a system capable of using location-based services, combined with image analysis, to automatically populate a database of graffiti images with information that can be used by law enforcement to identify, track, and mitigate gang activity. This can be done by implementing image analysis methods to segment the graffiti image in order to detect shapes, such as symbols and numbers, and orientation with respect to each other. These results can be associated to identify gangs, gang members, and track gang activity. Other future work includes enlarging the number of fields and relationships in the database so as to link gangs to their respective colors, acronyms, gang members, locations, or activity over time. The same can be done with graffiti components, in order to automatically interpret their position and alignment and the relationship between different components in the same or other graffiti. We also need to investigate the use of color calibration information for any type of future color analysis we do.

REFERENCES

[1] National Drug Intelligence Center (NDIC), *Attorney General's Report to Congress on the Growth of Violent Street Gangs in Suburban Areas*. United States Deptartment of Justice, April 2008.

[2] A. Parra, "An Integrated Mobile System for Gang Graffiti Image Acquisition and Recognition," M.S. Thesis, Purdue University, West Lafayette, IN, December 2011.

[3] A. K. Jain, J.-E. Lee, and R. Jin, "Graffiti-ID: Matching and Retrieval of Graffiti Images," *Proceedings of the 1st ACM Workshop on Multimedia in Forensics*, pp. 1–6, October 2009, Beijing, China.

[4] W. Tong, J.-E. Lee, R. Jin, and A. K. Jain, "Gang and Moniker Identification by Graffiti Matching," *Proceedings of the 3rd ACM Workshop on Multimedia in Forensics and Intelligence*, November 2011 (to appear), Scottsdale, AZ.

[5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004, Hingham, MA.

[6] "Graffiti Tracker." [Online]. Available: http://graffititracker.net

[7] B. Momjian, *PostgreSQL: Introduction and Concepts*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 2001.

[8] J. D. Touch, "Performance Analysis of MD5," *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pp. 77–86, October 1995, Cambridge, MA.

Mobile-Based Hazmat Sign Detection System

Bin Zhao, Albert Parra, Edward J. Delp
Video and Image Processing Lab (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

Abstract—In this paper we describe a mobile-based hazardous material (hazmat) sign detection system. The hazmat sign detection method is based on visual saliency models. We use saliency maps to denote regions likely containing hazmat signs in complex scenes and use a convex quadrilateral shape detector to find hazmat sign candidates in these regions. Our experimental results show that our proposed hazmat sign detection method is robust to distance, projective distortion, and blurred and shaded signs. Our method has an overall execution time of 5.09 seconds and a sign detection accuracy of 72.6%. Our image dataset consists of images taken in the field under different lighting and weather conditions, distances, and perspectives.

Index Terms—Sign detection, shape detection, saliency model.

I. INTRODUCTION

Hazardous materials can cause serious accidents and emergency situations and makes them dangerous to the society and the environment. A federal law in the US requires vehicles transporting hazardous materials to be marked with a standard sign (hazmat sign) identifying the type of hazardous material the vehicle is carrying [1]. Hazmat signs help identify the material and determine what specialty equipment, procedures and precautions should be taken in the event of an emergency. This information is contained in the Emergency Response Guidebook (ERG) [2], published by the US Department of Transportation (DOT). There exist several mobile-based applications that provide easy access to this guidebook for first responders in the field. For example, the official ERG 2012 mobile application [2] lets the user browse the ERG guidebook by UN identifier, template images from the book, and guide page. The WISER (Wireless Information System for Emergency Responders) mobile application [3] lets the user browse the ERG guidebook by known substances types and hazard classifications. However, these applications only provide different ways of manually browsing the guidebook. We have developed an integrated mobile-based system that makes use of location-based-services and image analysis methods. We call this system MERGE (Mobile Emergency Response GuidE) [4]. The MERGE mobile application is capable of detecting hazmat signs from an image and querying an internal database to provide accurate and useful information to first responders in real time.

Hazmat signs are designed to have identifying visual information that can be distinguished from their surroundings by specific colors, shapes, symbols, and numbers. However, there exist some challenges for successful detection of hazmat signs in complex scenes. The exposure of hazmat signs to different lighting and weather conditions deteriorate their shape and color over time, thus making the detection process more difficult. Moreover, since images taken by cameras on mobile device can be out of focus, additional image distortions may

occur, such as blur and change in contrast. Our hazmat sign detection system uses saliency maps and convex quadrilateral shape detection to overcome the problems mentioned above.

II. REVIEW OF EXISTING METHODS

Sign detection approaches can be divided into three categories: color-based methods [5], shape-based methods [6] and saliency-based methods [7]. Color-based methods take advantage of the fact that signs are often colored in highly visible contrasting colors. These specific colors are extracted from images for sign detection. For example, a color histogram backprojection method is used in [8] to detect interesting regions possibly containing hazmat signs. The luminance homogeneity of blocks is used in [9] to identify homogenous regions as a first step towards detection of information signs containing text. In [10] several color components are used to segment traffic signs under different weather conditions. However, these methods are not robust to lighting conditions and illumination changes. Shape-based approaches first generate an edge map and then use shape characteristics to find signs. For example, in [11] triangular, square and octagonal road signs are detected by exploiting the properties of symmetry and edge orientations exhibited by equiangular polygons. A shape classification method of a road-sign detection system in [12] is based on linear and Gaussian-kernel support vector machines (SVM). Shape-based methods are invariant to translation, rotation and scale. Saliency-based approaches make use of visual saliency models to construct saliency maps that denote areas where signs are likely to be found. For example, in [13] a saliency map of road traffic signs is constructed by the weighted sum of color and edge feature maps for image preprocessing. A traffic sign recognition system in [14] uses a visual attention system to generate regions with possible traffic sign candidates.

Visual saliency models have recently become an active research area [15]. The main idea behind computational saliency models is to generate several types of features from input images and fuse them in a scalar map, called saliency map. For example, a notable saliency-based visual attention (SBVA) model was proposed in [7] using intensity, color and orientation features with a subsampled Gaussian pyramid. In [16] a graph-based visual saliency (GBVS) method forms the activation map from each feature map based on graph theory. A dynamic visual attention (DVA) model based on the rarity of features is proposed in [17]. A multi-scale dissimilarity aggregation (MSDA) method is used to estimate the saliency of regions in [18], which is robust to the presence of significant noise. In [19] an image signature (IS) method for image figure-ground separation is developed to generate a saliency map in RGB or Lab color spaces. An saliency detector based on hypercomplex Fourier transform (HFT) is presented in [20] using the convolution of the image amplitude spectrum with a low-pass Gaussian kernel.

III. MOBILE-BASED HAZMAT SIGN DETECTION SYSTEM

A. System Overview

Figure 1 illustrates our mobile-based hazmat sign detection system. There are two basic operation modes: analysis of new or existing images and internal database browsing. The first mode includes capturing or selecting an image from the device and performing image preprocessing and image analysis. The image preprocessing step includes blur detection and color correction based on daylight references. When available, the accelerometer on the mobile device is used to detect shaking and avoid the lose of focus before taking the image. In order to save battery and lower the execution time, the hazmat sign detection is currently performed on a back-end server and the results are sent back to the mobile device. We built an internal database based on the contents of the 2012 ERG guidebook, which is queried using the detection results to retrieve emergency response information. The second mode includes browsing our internal database to obtain information about a specific hazmat sign. The internal database can be browsed by UN identifier, class, symbol, or color.

Fig. 1: Mobile-Based Hazmat Sign Detection System.

We implemented a prototype of our system as an application for Android mobile devices. Figure 2 shows the user interface and its operation workflow. The figures on the bottom row illustrate the internal database query results, where we display information about potential hazards, public safety and emergency response. We also use a map to project an evacuation area based on the chemical found, the size of the chemical spill and the time of the day. This information is extracted from our internal database.
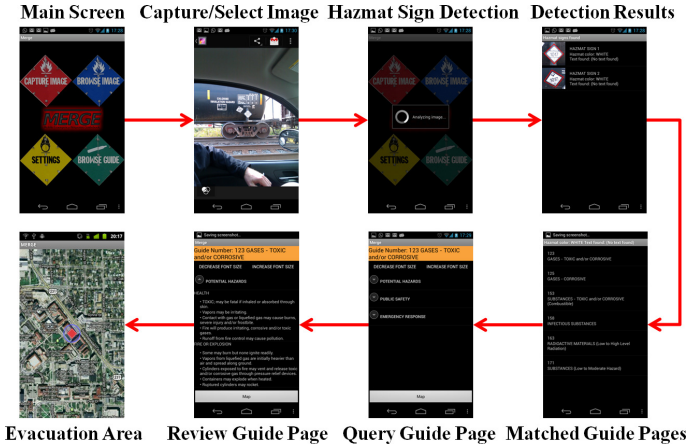
Fig. 2: User Interface of Mobile Application.

B. Hazmat Sign Detection Method

We use saliency-based methods to find regions likely containing hazmat signs in complex scenes. We use saliency maps in the RGB color space and the Lab color space. We prove that by using multiple color spaces we improve the detection accuracy (see Section IV). The block diagram in Figure 3 shows the five steps of our hazmat sign detection method:
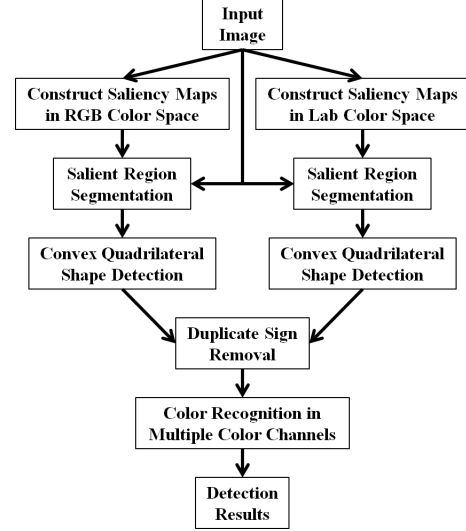
Fig. 3: The Proposed Hazmat Sign Detection Method.

(1) Construct Saliency Maps in Two Color Spaces: The input image is represented in both RGB and Lab color spaces. In each color space, two saliency maps are constructed using two computational saliency models, IS [19] and HFT [20], which are normalized and combined into a single saliency map. Note that the original HFT method uses I-RG-BY opponent color space. We modified this method to use different color features and weights in the RGB and Lab color spaces. This combined saliency map method, which we call IS+HFT(RGB+Lab), gave us the best results in the experiments. The IS+HFT(RGB+Lab) method generates two saliency maps, one for RGB and one for Lab. These saliency maps assign higher saliency value to more visually attractive regions. **(2) Salient Region Segmentation:** We threshold each saliency map to create a binary mask to segment the salient regions from the original image. The threshold T is determined as k times the mean saliency value of a given saliency map. That is, $T = \frac{k}{H \times W} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x, y)$, where W and H are the width and height of the saliency map, $S(x, y)$ is the saliency value at position (x, y) and k is empirically selected based on our experiments ($k = 4$ for RGB and $k = 3$ for Lab). **(3) Convex Quadrilateral Shape Detection:** We detect a hazmat sign with specific shape in segmented salient regions by checking for convex quadrilaterals. The salient regions in two sets are transformed into grayscale images and then thresholded into binary images using an empirically determined threshold. The binarized regions is morphologically opened to remove small objects and dilated to merge areas that may belong to the same object. We then retrieve the contours from the resulting binary image [21]. For each contour, we use the standard Hough Transform [22] to find straight lines that approximate the contour as a polygon. The intersections of these lines give us the corners of the polygon, which can be used to discard non-quadrilateral shapes. If the contour is approximated by four vertices, we find its convex hull [23]. If the convex hull still has

four vertices, we check the angles formed by the intersection of its points. If each of these angles is in the range $90° \pm 1.5°$, and the ratio of the sides formed by the convex hull is in the range 1 ± 0.5, we can assume that we have found a convex quadrilateral as a hazmat sign candidate. **(4) Duplicate Sign Removal:** For each candidate set satisfying the geometric constraints we compute its minimal bounding box. We then discard any candidate with a bounding box aspect ratio smaller than 1.3. Finally, we check the remaining candidates and remove those that correspond to the same sign. This can be done by first dividing all bounding boxes that overlap more than 50% into groups, and then finding the optimal bounding box for each group. We consider the optimal bounding box to be the one with its nodes closest to its centroid (i.e., closest to a square). Once a hazmat sign is segmented, its color is set to the average hue inside the optimal bounding box and the color is used to identify the sign. **(5) Color Recognition in Multiple Color Channels:** For each salient region found, we detect signs using specific color channels. This allows us to do both sign detection and color recognition at the same time, since we will assume that the color of any hazmat sign found in the region will correspond to the color channel associated to it. Currently the channels used are red, green, blue, and we also use the grayscale version to account for white or black signs. Note that the possible colors for hazmat signs also include yellow and orange, but these can be obtained by transforming the image from RGB to a hue-based color space and then segment the hue channel. The grayscale and the color channels are thresholded to account for highly chromatic areas using an empirically determined threshold. The recognized color information is used for sign identification based on the contents of the 2012 ERG guidebook. Figure 4a illustrates a successful detection of two signs, one of which is affected by perspective and rotation distortion. Figure 4b illustrates a true positive and a false positive.

TABLE I: Average Execution Time (in Seconds), Distribution and Score of Each Saliency Map Method.

Method	Time	Good	Fair	Bad	Lost	Score
SBVA(I-RG-BY)	2.0739	34	16	11	1	145
GBVS(I-RG-BY)	3.3626	30	15	15	2	135
DVA(RGB)	0.4309	19	2	11	30	72
MSDA(RGB)	3.7357	22	7	27	6	107
IS(I-RG-BY)	0.4277	23	4	17	18	94
IS(RGB)	0.3609	45	8	4	5	155
IS(Lab)	0.3901	27	5	20	10	111
HFT(I-RG-BY)	0.5889	33	8	12	9	127
HFT(RGB)	0.5251	38	5	8	11	132
HFT(Lab)	0.5495	37	10	8	7	139
IS(RGB+Lab)	0.7511	52	6	1	3	169
HFT(RGB+Lab)	1.0749	41	6	8	7	143
IS+HFT(RGB+LAB)	1.8260	55	4	2	1	175
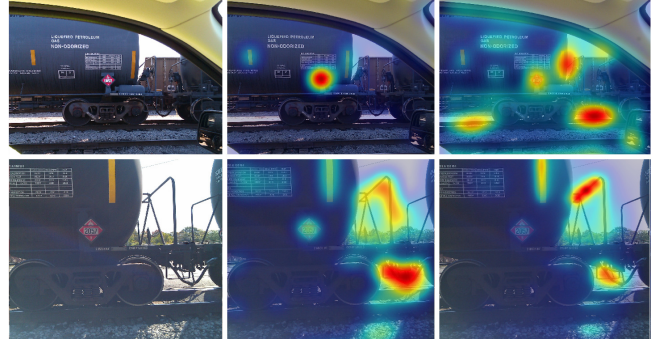
Fig. 6: Saliency map categories (top to bottom, left to right): original image, good, fair; original image, bad, lost.

(a) Two True Positives.   (b) One True Positive and One False Positive.

Fig. 4: Examples of Detection Results.

Fig. 5: Samples from the image dataset. From left to right: low resolution, projective distortion, blurred sign, shaded sign.

IV. Experimental Results

We performed two experiments to determine the speed and accuracy of our proposed method. The first experiment consisted of constructing saliency maps using different existing saliency models in the literature, and evaluating their performance based on ground-truth information. The second experiment consisted of performing hazmat sign detection on images from our dataset and manually comparing the detection results with ground-truth information. The tests were executed on an Galaxy Nexus mobile telephone with a dual-core 1.2GHz CPU and 1GB RAM and a backend server with a quad-core 2.4GHz CPU and 4GB RAM. Our image dataset consisted of 50 images, each containing one or more hazmat signs (62 hazmat signs in total). The images were acquired in the field, under various lighting and weather conditions, distances, and perspectives. Among the 50 images, 4 were taken at 10-50 feet, 23 at 50-100 feet and 23 at 100-200 feet. Among the 62 hazmat signs, 2 had low resolution, 11 had projective distortion, 8 were blurred and 6 were shaded. Figure 5 illustrates some samples from the image dataset. The images were taken using three different cameras: a 8.2 Mpx Kodak Easyshare C813, a 16 Mpx Nikon Coolpix S800c, and a 5 Mpx camera on an HTC Wildfire mobile telephone. The ground-truth information included the distance from the camera to the sign, sign color, projective distortion of the sign, image resolution, possible shadow affecting the sign, and sign location on the image. Note that we only used the color and not the text inside the sign for sign identification for these experiments.

Table I shows the results of the evaluation of different saliency map methods, including average execution times and scores. The saliency map methods selected in the experiment are: SBVA [7], GBVS [16], DVA [17], MSDA [18], IS [19], HFT [20]. The scores were calculated as follows. We manually classified the resulting saliency maps into four categories: good, fair, bad, and lost. For each of the 62 signs,

TABLE II: Hazmat Sign Detection Results.

Method	Total	Sign	Accuracy	Color	Accuracy
IS(RGB+Lab)	62	42	67.7%	36	58.1%
HFT(RGB+Lab)	62	32	51.6%	26	41.9%
IS+HFT(RGB+Lab)	62	45	72.6%	38	61.3%

we assigned 3 points to a good map (sign was mostly contained in a high saliency-valued region), 2 points to a fair map (sign was mostly contained in a middle saliency-valued region), 1 point to a bad map (sign was mostly contained in a low saliency-valued region), and 0 point to a lost map (sign was not contained in any saliency-valued region). Figure 6 illustrates examples of each case. The score of each saliency map method is the sum of the points assigned to each image in the dataset, which ranges from 0 to 186 for each saliency map method. The DVA method runs faster than the SBVA, the GBVS and even the HFT methods but it has the lowest score. The MSDA method has a higher score than the DVA method but runs slower than the SBVA and the GBVS methods. The IS and the HFT methods using a single color space have high scores and execute faster than the SBVA and the GBVS methods. The IS(RGB+Lab) and the HFT(RGB+Lab) methods using two color spaces both had higher scores, at the expense of the execution time. The IS(RGB+Lab), the HFT(RGB+Lab) and the IS+HFT(RGB+Lab) methods in two color spaces still run 2.76, 1.93 and 1.14 times faster than the SBVA method and 4.48, 3.13 and 1.84 times faster than the GBVS method respectively. Note that the SBVA and the GBVS methods use a single color space. These results indicate that we can combine the IS and the HFT methods in RGB and Lab to improve their individual score, while still having a reasonable execution time.

Table II shows the results of our proposed hazmat sign detection method using the IS and the HFT methods with two color spaces. We determined how many signs were successfully detected (*Sign*) and how many were correctly color recognized after a successful detection (*Color*). Note that the sign color recognition was done only if a sign was detected, and that multi-colored signs may also cause our method to misidentify the sign color, given that we detect signs at individual color channels. The accuracy of our proposed method is mainly influenced by the thresholding method used in the saliency region segmentation, and the non-adaptive morphological operations used in the convex quadrilateral shape detection. The overall average execution times of our hazmat sign detection method using IS(RGB+Lab), HFT(RGB+Lab) and IS+HFT(RGB+Lab) are 2.6034, 2.4871 and 5.0908 seconds in total respectively. The IS+HFT(RGB+Lab) method has a higher accuracy than the IS(RGB+Lab) and the HFT(RGB+Lab) methods, and the total execution time is still suitable for real-time applications.

V. Conclusions and Future Work

We described a mobile-based hazmat sign detection system that combines salient region segmentation and convex quadrilateral shape detection. Our experimental results show that our proposed hazmat sign detection method is robust to distance, projective distortion, and blurred and shaded signs. Our method has an overall execution time of 5.09 seconds and a sign detection accuracy of 72.6%. Our image dataset consists of images taken in the field under different lighting and weather conditions, distances, and perspectives. We will investigate the use of an adaptive threshold for the salient region segmentation. We can further increase the sign detection accuracy by improving the color correction methods. We can improve the color recognition by using a hue-based channel and accounting for multi-

colored signs. An optical character recognition method can be also used to interpret the text inside the detected hazmat signs.

References

[1] United States Department of Transportation, *Code of Federal Regulations, Title 49, DOT Hazmat*, Labelmaster, 2012 edition, October 2012.

[2] ERG, Available: http://www.phmsa.dot.gov/hazmat/library/erg.

[3] WISER, Available: http://wiser.nlm.nih.gov.

[4] MERGE, Available: https://redpill.ecn.purdue.edu/~hazmat.

[5] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, no. 3, pp. 453–464, March 1999.

[6] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1274–1286, October 2003.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, November 1998.

[8] D. Gossow, J. Pellenz, and D. Paulus, "Danger sign detection using color histograms and SURF matching," *Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics*, pp. 13–18, October 2008, Sendai, Japan.

[9] K. L. Bouman, G. Abdollahian, M. Boutin, and E. J. Delp, "A low complexity sign detection and text localization method for mobile applications," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 922–934, October 2011.

[10] L. Song and Z. Liu, "Color-based traffic sign detection," *Proceedings of the International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, pp. 353–357, June 2012, Chengdu, China.

[11] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 70–75, September 2004, Stockholm, Sweden.

[12] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264–278, June 2007.

[13] W.-J. Won, M. Lee, and J.-W. Son, "Implementation of road traffic signs detection based on saliency map model," *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*, pp. 542–547, June 2008, Eindhoven, Netherlands.

[14] R. Kastner, T. Michalke, T. Burbach, J. Fritsch, and C. Goerick, "Attention-based traffic sign recognition with an array of weak classifiers," *Proceedings of the IEEE Intelligent Vehicles Symposium (IVS)*, pp. 333–339, June 2010, San Diego, CA, USA.

[15] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, January 2013.

[16] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pp. 545–552, December 2006, Vancouver, BC, Canada.

[17] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pp. 681–688, December 2008, Vancouver, BC, Canada.

[18] Chelhwon Kim and Peyman Milanfar, "Visual saliency in noisy images," *Journal of Vision*, vol. 13, no. 4, pp. 1–14, March 2013.

[19] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, January 2012.

[20] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 996–1010, April 2013.

[21] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, April 1985.

[22] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, January 1972.

[23] Jack Sklansky, "Finding the convex hull of a simple polygon," *Pattern Recognition Letters*, vol. 1, no. 2, pp. 79–83, December 1982.

Recognition, Segmentation and Retrieval
of Gang Graffiti Images

Albert Parra, Bin Zhao, Joonsoo Kim, Edward J. Delp
Video and Image Processing Lab (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

Abstract—In this paper we describe three methods for recognition, segmentation and retrieval of gang graffiti images. The first method is color recognition based on touchscreen tracing, the second method is color image segmentation based on Gaussian thresholding and the third method is content based image retrieval. Our experimental results show an image retrieval accuracy of 92.8% for gang graffiti scene recognition and an image retrieval accuracy of 50.0% for gang graffiti component classification. The experiments also show an average image retrieval time of 0.56 seconds, from which the scoring process takes on average 984 microseconds.

Index Terms—Color recognition, image segmentation, content based image retrieval, Gaussian thresholding, SIFT.

I. INTRODUCTION

Gangs are a serious threat to public safety throughout the United States. They are responsible for an increasing percentage of crime and violence [1]. Street gang graffiti is their most common way to communicate messages, including challenges, warnings or intimidation to rival gangs. It is an excellent way to track gang affiliation, growth, and membership. The goal of our work is to develop a mobile-based system capable of using location-based-services, combined with image analysis methods, to provide accurate and useful information concerning gangs based on a network connected database of gang graffiti images. In this paper we describe three methods for recognition, segmentation and retrieval of gang graffiti images. Our first method is color recognition of gang graffiti based on touchscreen tracing. Our second method is color image segmentation based on Gaussian thresholding. Our third method is content based gang graffiti image retrieval. Our content based image retrieval method is tested for accuracy and speed in two scenarios: scene recognition and gang graffiti component classification.

Previous work on gang graffiti: In [2] methods for segmenting and retrieving graffiti images are described using global thresholding and template matching. In [3] we described a mobile-based gang graffiti system that uses location information for querying a database of graffiti images. That system does not include image segmentation or matching. Other approaches described in [4], [5], [6], [7] do not use the bag-of-words models for image retrieval of gang graffiti and tattoos and report slower matching and retrieval times than we demonstrate in our experiments. In [8], [9], bag-of-words models for image retrieval of gang or gang-like tattoos are used but are not intended for real-time retrieval in mobile-based environments.

II. REVIEW OF EXISTING METHODS

Color Recognition: Gang graffiti are often sprayed in non-uniform surfaces, which makes them difficult to distinguish from the background. Since our system is deployed on a mobile telephone, we take advantage of the touchscreen capabilities of modern mobile devices to aid the recognition of color in gang graffiti images. The touchscreen can be used to detect a path drawn with the finger on the screen for image analysis such as color recognition. Color recognition techniques using tactile feedback use thresholds based on perceptual attributes of specific color spaces. The perceptual thresholds (also known as discrimination thresholds) have been widely studied for human observers [10]. However, some methods do use thresholds based on human perceptibility, but use application based thresholds. For example, some skin detection methods use an adaptive skin color filter to detect color regions, by setting thresholds in both RGB and HSV color spaces [11], [12].

Color Image Segmentation: Image color segmentation techniques can be divided into three categories [13]: physics based, feature-space based, and image-domain based. Methods based on physics include dichromatic reflection models [14] and unichromatic reflection models [15] for single illumination sources, and a more general model of image formation [16] for multiple illuminations. Methods based on feature spaces can be sub-categorized into three groups: clustering of regions given patterns with specific properties, including methods such as k-means clustering [17] or Iterative Self-Organizing Data Analysis Technique (ISODATA) [18]; adaptive k-means clustering, including methods based on maximum a posteriori (MAP) estimation [19] or split-and-merge strategies [20]; and histogram thresholding, including methods based on RGB thresholding and hue information [21], specific skin color domains [22], or entropy thresholding [23]. Methods based on the image-domain can be subcategorized into four groups: split-and-merge, including methods such as region smoothing by Markov Random Fields (MRF) [24] or splitting by either watershed transform [25] or quad-tree image representation for segmentation of skin cancers [26], among others; region growing, including methods such as RGB color distribution growing, HSV morphological open-close growing, or color quantization growing [27]; classification based, including methods such as minimization of Hopfield networks [28], or background extraction using two three-layered neural network [29]; edge based techniques, including methods such as combination of HSI gradients [30], active contours, or the Mumford-Shah variation model [31].

Content Based Image Retrieval: Content Based Image Retrieval (CBIR) consists of four core techniques [32]: visual signature extraction, similarity measures, and classification and clustering. Visual signature extraction usually implies three steps: 1) segmenting images using methods such as k-means clustering [33], normalized cuts [34], or salient region detection [35]; 2) extracting features such as color, texture, or shape [36]; 3) constructing the signatures using distri-

butions [37] or adaptivity [38]. Similarity measure methods include manifold embedding [39], and vector quantization [40]. Classification and clustering methods include hierarchical k-means [41], support vector machine [42], or Bayesian classifiers [43].

Our color image segmentation approach falls into the feature-space based techniques. However, our approach differs from the methods mentioned above. Although there are some techniques in the literature that use only hue or luma information, either circular histogram thresholding [44] or one-dimensional histogram thresholding [45], we do not obtain the descriptors of the probability distribution from the color histogram of the image. Instead, the median and the variance obtained from the tracing-bases color recognition process are used for segmentation. Our segmentation approach does not produce binarized images, but grayscale images weighed by a Gaussian distribution, thus creating a probability map for a specific luma or hue. These types of probability maps are used for increased accuracy and robustness in some clustering techniques [46], [47]. Our content based image retrieval approach uses hierarchical k-means to build a vocabulary tree based on the method in [41].

III. GANG GRAFFITI COLOR RECOGNITION AND SEGMENTATION

One of the goals of our system is to identify the color of graffiti components of an image. We use features and a priori information specific to gang graffiti. First, gang graffiti are mostly monochromatic. This makes it easier to segment graffiti components. Second, gang graffiti are hand-written with each gang member having a different style. This rules out the use of generic Optical Character Recognition methods. Third, gang graffiti are almost always painted on non uniform surfaces with various textures, such as walls, garage doors, or trees. This makes the segmentation of the graffiti contents more challenging. We developed a method for identifying the color of a graffiti component. We call this approach color recognition based on tracing. It can be implemented on a hand-held device without the need of an network connection. We also describe a method for segmenting an image based on the colors determined by our tracing method. We call this color image segmentation based on Gaussian thresholding.

Color Recognition Based on Touchscreen Tracing: In this method the user takes an image of a gang graffiti and traces a path along a colored region using the touchscreen display. Then we recognize the color along the path, and provide a list of gangs related to the color by querying an internal database on the mobile phone. For this method we use an RGB to Y'CH color space conversion. Figure 1 shows an overview of our color recognition method. The path is drawn along a component of the graffiti image assumed to have uniform color. The RGB color components of each pixel on the path are converted to a new luma/chroma/hue color space that we call the Y'CH color space. The Y'CH color space is used because color changes are more intuitive and perceptually relevant to represent in luma or hue than in RGB triplets, in order to obtain the median and the variance of the color along the traced path. Equation 1 shows the mapping between RGB and Y'CH. Note that we use luma (Y') as opposed to luminance (Y) [48]. Third, we compute three medians on the pixel array that forms the path, namely the luma median ($\tilde{Y}$), the chroma median ($\tilde{C}$) and the hue median ($\tilde{H}$). We then define three disjoint regions in our Y'CH color space (labeled 3a, 3b and 3c in Figure 1), delimited by manually set thresholds based on luma ($T_{Y_w} = 0.12$, $T_{Y_b} = 0.85$) and chroma ($T_C = 0.05$). These thresholds were empirically obtained from our database of gang graffiti, consisting of more than 600 gang

graffiti images. Depending on the region where the medians are located, we do color recognition based on luma (3a) or hue (3b). Once we have the median, either based on luma or hue, we need to decide which color is associated with it. From all the images in our database, the possible colors for gang graffiti components are black, white, red, blue, green, gold and purple. If the median is based on luma, the color detected is either black ($\tilde{Y} \leq 0.5$) or white ($\tilde{Y} > 0.5$). If the median is based on hue, the color detected is $H_d = \min_i(\theta(\tilde{H}, H_{A_i}))$, where $\theta(\tilde{H}, H_{A_i})$ is the angular distance between the computed hue ($\tilde{H}$) and the i-th component of a set of average hues (H_A), empirically obtained from analyzing 100 color calibrated images taken from our database. These values are (color, H_A)=({Red, 6.10 rad}, {Blue, 4.00 rad}, {Green, 2.20 rad}, {Gold, 0.69 rad}, {Purple, 5.15 rad}). Once the color is detected, we provide a list of gangs related to that color by querying an internal database on the mobile phone. Finally, we also estimate the variance $\sigma_{\tilde{X}}^2$ around the computed median $\tilde{X} = \{\tilde{Y} \text{ or } \tilde{H}\}$. This variance is used as an input to the color image segmentation method described next. Note that this method can be used with multi-colored graffiti by using it on each trace on the touchscreen.
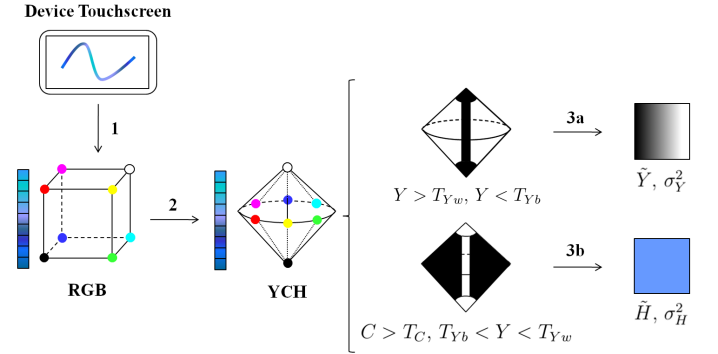

Fig. 1: Color Recognition Using Touch Screen Tracing.

$$Y' = 0.299R + 0.587G + 0.114B.$$
$$C = max(R, G, B) - min(R, G, B) = M - m$$
$$H = \begin{cases} \frac{G-B}{C} & \text{if } M = R \\ \frac{B-R}{C} + 2 & \text{if } M = G \\ \frac{R-G}{C} + 4 & \text{if } M = B \\ 0 & \text{if } C = 0 \end{cases} \tag{1}$$

Color Image Segmentation Based on Gaussian Thresholding: For the segmentation we use a Gaussian threshold near a specific luma or hue value in the Y'CH color space, in order to produce a segmented image where each pixel is given a weight depending on its distance from a median. Figure 2 shows an overview of our color segmentation method divided in 5 steps. We assume that, given a graffiti image, we have the median $\tilde{X}$ and the variance, $\sigma_{\tilde{X}}^2$, of a traced path (step 1a). We then transform the entire RGB image to the our Y'CH color space (steps 1b and 2). We finally segment the image using Gaussian thresholding (steps 3 to 5). The segmentation works as follows. We first ignore all pixels in the image that fall outside our manually set thresholds in the Y'CH color space (step 3). Note that these thresholds are the same as the ones used for the color recognition process. We apply a weight to the rest of the pixels using a normal distribution centered at $\tilde{X}$ and a confidence interval of $2\sigma_{\tilde{X}}$ (step 4). The output

of this process is a grayscale image where each pixel is given a probability based on a normal distribution (step 5). This probability is higher as the pixel value gets closer to $\tilde{X}$. The image is then scaled to $[0, 255]$. Figure 3 shows an example where the color recognition process is performed by tracing a path along the blue numbers "2" and "5" in the graffiti component. Figure 4 shows the effect of the Gaussian thresholding process on the letters "Hill". Note that this method produces a probability map, where the values in a graffiti component decrease as the spray paint fades.
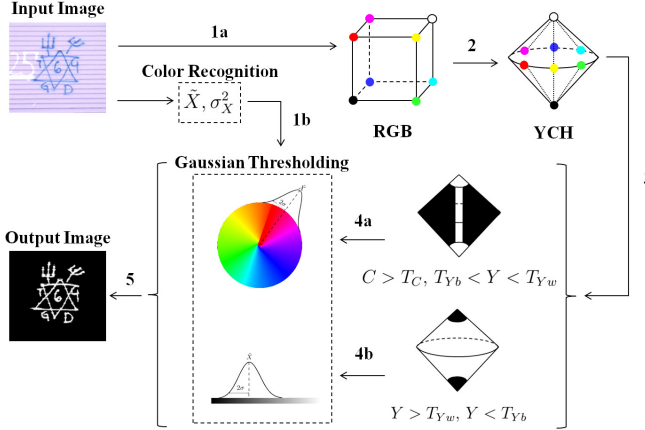
Fig. 2: Color Image Segmentation Using Gaussian Thresholding.
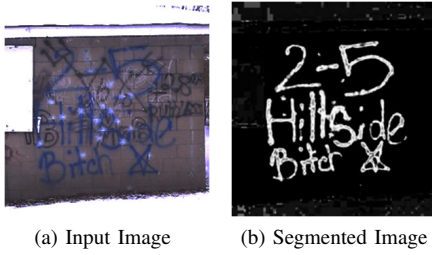
(a) Input Image (b) Segmented Image

Fig. 3: Gaussian Thresholding on blue. $(\tilde{H}, \sigma_{\tilde{H}}^2) = (4.19, 0.05)$.
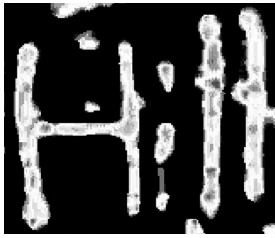
Fig. 4: Probability map created by the Gaussian Thresholding.

IV. CONTENT BASED GANG GRAFFITI IMAGE RETRIEVAL

We describe a method to recognize gang graffiti by matching image features from query images against our database of gang graffiti. The method is currently used in two scenarios: 1) recognize scenes containing graffiti (Figure 5) and 2) classify individual graffiti components (Figure 6). In both cases we use a vocabulary tree [41] to retrieve query images. The vocabulary tree is obtained as follows. First, we extract SIFT features from a set of training images to get sets of 128 dimensional vectors. To recognize scenes containing

graffiti, we extract SIFT features from the entire image, similar to the work done in [4], [5], [6], [7], [8], [9] for graffiti and tattoo images. To recognize individual graffiti components, we extract features from segmented graffiti components. By graffiti components we mean the objects and shapes contained in a graffiti image, such as stars, pitchforks, crowns, and arrows. These components are found by first performing color recognition based on touchscreen tracing, then applying our color image segmentation based on Gaussian thresholding, and finally segmenting each graffiti component. Note that at this stage we binarize the probability map returned by the Gaussian thresholding method. That is, we do not take into account the strength of the graffiti trace. Also note that once we segment the gang graffiti from the background we segment the graffiti components manually. Then, we use hierarchical k-means clustering to recursively divide the $\mathbb{R}^{128}$ space into sub-clusters, as in [9]. Each sub-cluster contains the set of descriptors closest to its center. We call each of these sub-clusters a word. This clustering can be interpreted as a vocabulary tree, where k corresponds to the branching factor at each level, and each word corresponds to a leave in the tree. Figure 7 illustrates this equivalence. Each black dot corresponds to a descriptor from a training image. Note that we keep track of the image corresponding to each descriptor. Therefore, each word can be associated to a number representing a path down the vocabulary tree. At the end of the training, each image i can be represented as an n_w dimensional vector d_i, n_w being the total number of words in the tree. At each index j, an entropy weighting [41] is applied so that $d_i[j] = N_j^i \ln \frac{N}{N_j}$, where N_j^i is the the number of descriptors of the i-th training image associated with the j-th word, N is the total number of training images, and N_j is the number of training images with at least one descriptor belonging to the j-th word. Note that d_i is normalized to make it invariant to the total number of descriptors found on the i-th image. Based on the results of [41] we chose $k = 3$ and $10,000$ leaves to create our vocabulary tree. In order to match a query image to an image in our database we first extract SIFT descriptors from the query image. Each of the query descriptors is pushed down the vocabulary tree to find its closest word, and an n_w dimensional vector q is created following the same criteria as in the training process, such that $q[j] = N_j^q \ln \frac{N}{N_j}$, where N_j^q is the the number of descriptors of the query image associated with the j-th word. The closest match to the query image is then $\min_i ||q - d_i||$. Since all the images in our database have location information, we can improve the performance by comparing q to a set of training images in a certain physical radius from the query image. Note that instead of computing norms between the query vector and each training image we could use inverted files in memory to speed up the process [41]. With the use of the location information we reduce the search on average to 10 images instead of the entire training set. The main advantage of using a vocabulary tree for image retrieval is that its leaves define the quantization, thus making the comparison dramatically less expensive than previous methods in the literature. Also, once the vocabulary tree is built, new images can be added by just pushing down its descriptors. Currently, SIFT features are used for both scenes containing graffiti and individual graffiti components. However, note that the k-means clustering accepts any type multi-dimensional vector, hence we can use gang graffiti related features in the future to improve the retrieval performance.

V. EXPERIMENTAL RESULTS

We did two experiments to determine the accuracy and the speed of our image retrieval approach. The tests were executed in a desktop computer with a 2.8GHz CPU and 2GB RAM. The goal of the first
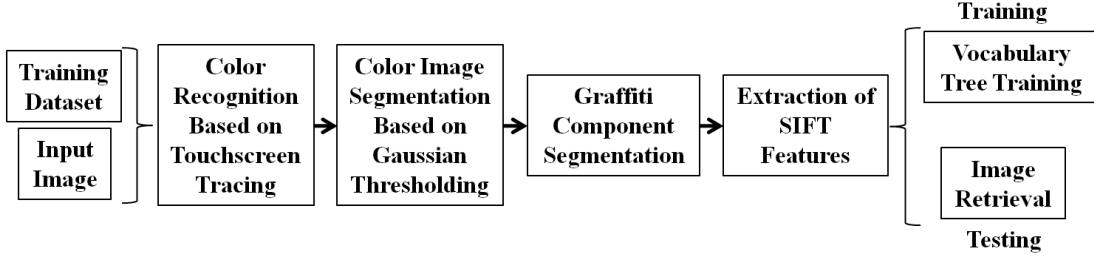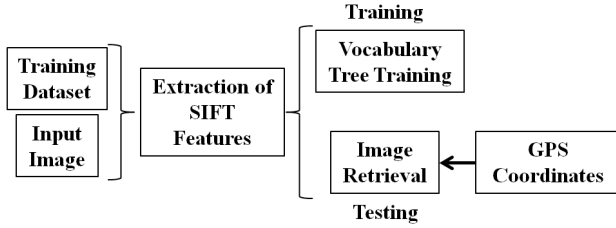
Fig. 6: Gang Graffiti Component Classification.

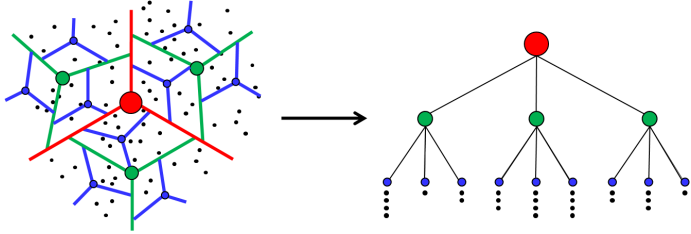

Fig. 5: Gang Graffiti Scene Recognition.

Fig. 7: Vocabulary Tree From Hierarchical k-Means.

TABLE I: Experimental Results For Scene Recognition.

	Score						
Images	5	4	3	2	1	0	Accuracy
100	91	1	1	1	0	6	92.8%

point star, pitchfork, or arrow. Figure 10 illustrates an example of each class. For each test image we retrieved its 5 closest matches on the training set and we associated a class to it based on a scoring scheme. Given the scores of the 5 for closest matches $s = \{s_1, s_2, s_3, s_4, s_5\}$ in ascending order, we normalize them and invert them so that the new scores become $p = \{p_1, p_2, p_3, p_4, p_5\}$, where $p_i = 1 - \frac{s_i}{\sum s_i}$. Then, we manually group the 5 closest matches into N classes, $N \in \{1, ..., 33\}$. We add up the new scores associated to each class, and we assign the class C with the highest score to the query image, such that $C = \underset{n}{\mathrm{argmax}}\{\sum_k p_k^{(n)}\}$, where k is the set of indeces of s belonging to the n-th class, $n \in \{1, ..., N\}$. Given the 40 testing images, containing multiple examples from the 33 classes, the classification accuracy is 50.0%. The classification accuracy for the class *6-point star* is 66.7%, while the classification accuracy for the class *pitchfork* is 33.6%. Thus, using SIFT features for graffiti component classification is good for some classes but poor for others. The fundamental reason why SIFT is not good enough for component classification is that the some classes overlap. For example, *5-point star* includes *X*, and *pitchfork* includes *3*. The average execution time of the second experiment was 0.57 seconds. Note that although the training set for the second experiment is smaller than the training set for the first experiment, the running times for the two experiments are similar. This is because the process that takes longer (an average of 0.45 seconds in both experiments) is finding the paths of the query image features down the vocabulary tree. The scoring process takes an average of 984 microseconds in both experiments.

experiment was to match query images to images in our database based on the scene. That is, by finding features not only from the graffiti in the image, but also of the background. We trained 708 images from our database and used hierarchical k-means to create a vocabulary tree. Based on the results of [41] and the size of our dataset we set $k = 3$ and the number of leaves in the vocabulary tree to 300. A separate set of 100 images was used for testing. Figure 8 shows some sample images. Each of the test images corresponded to one of the scenes in our database, but with different viewpoint, rotation, and illumination. For each test image we retrieved its 5 closest matches from the training set and we gave it a score from 5 to 0, 5 meaning that the matching scene was in first position, and 0 meaning that there was no matching scene in the top 5 results. Table I summarizes the results of the first experiments. It is worth noting that although this experiment only accounted for scene recognition, we found that sometimes the results returned included scenes close to the query. Figure 9 illustrates an example. These results can be used to recognize nearby graffiti or even graffiti that have been removed. The average execution time of the first experiment was 0.55 seconds. The goal of the second experiment was to classify query images into categories based on a set of gang graffiti symbols. We trained 162 images, each one consisting of one graffiti component in black with white background. A separate set of 40 images was used for testing. Each of the test images also consisted of one graffiti component in black with white background. We considered 33 distinct graffiti components (i.e., classes), including 5-point star, 6-

Fig. 9: Query images (left) and similar retrieved scenes (right).

Fig. 8: Image dataset sample.

Fig. 10: Gang Graffiti Component Classes.

VI. Conclusions and Future Work

We described two methods for color recognition and segmentation of gang graffiti and one method for content based gang graffiti image retrieval. We evaluated the accuracy and speed of the content based gang graffiti image retrieval in two scenarios: scene recognition and gang graffiti component classification. The experimental results showed that the use of SIFT features for scene recognition produce very accurate outcomes, but the use of SIFT features for component classification does not produce accurate results. The experiments also showed that the image retrieval is fast in both scenarios. We can increase the accuracy of the component classification by creating our own set of gang graffiti component features. The set would include information such as the component color, aspect ratio, or position and alignment with respect to other components. We can also use the probability map created by the Gaussian thresholding method to estimate the direction of the graffiti trace, and use that information as a component feature. These new features can still be used to train a vocabulary tree using the current approach. We are also investigating methods for automatic graffiti component segmentation based on elastic shape recognition [2]. Finally, we are working on implementing a mobile-based version of our classification system by optimizing the feature extraction (i.e. using Dense SIFT features instead of SIFT) and the image retrieval method (i.e. using inverted file lookup methods).

References

[1] National Drug Intelligence Center (NDIC), *Attorney General's Report to Congress on the Growth of Violent Street Gangs in Suburban Areas*, United States Deptartment of Justice, April 2008.

[2] C. Yang, P. C. Wong, W. Ribarsky, and J. Fan, "Efficient graffiti image retrieval," *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, pp. 36:1–36:8, June 2012, Hong Kong, China.

[3] A. Parra, M. Boutin, and E. J. Delp, "Location-aware gang graffiti acquisition and browsing on a mobile device," *Proceedings of the IS&T/SPIE Electronic Imaging on Multimedia on Mobile Devices*, pp. 830402–1–13, January 2012, San Francisco, CA, USA.

[4] A. K. Jain, J.-E. Lee, R. Jin, and N. Gregg, "Content-based image retrieval: An application to tattoo images," *Proceedings of the International Conference on Image Processing*, pp. 2745–2748, 2009.

[5] J.-E. Lee, A.K. Jain, and R. Jin, "Scars, marks and tattoos (SMT): Soft biometric for suspect and victim identification," *Proceedings of the Biometrics Symposium*, pp. 1–8, September 2008, Tampa, Florida, USA.

[6] W. Tong, J.-E. Lee, R. Jin, and A. K. Jain, "Gang and Moniker Identification by Graffiti Matching," *Proceedings of the 3rd ACM Workshop on Multimedia in Forensics and Intelligence*, November 2011, Scottsdale, AZ.

[7] A. K. Jain, J.-E. Lee, and R. Jin, "Graffiti-ID: Matching and Retrieval of Graffiti Images," *Proceedings of the 1st ACM Workshop on Multimedia in Forensics*, pp. 1–6, October 2009, Beijing, China.

[8] D. Manger, "Large-scale tattoo image retrieval," *Proceedings of the Conference on Computer and Robot Vision*, pp. 454–459, May 2012, Toronto, Ontario, Canada.

[9] J.-E. Lee, R. Jin, A. K. Jain, and W. Tong, "Image retrieval in forensics: Tattoo image database application," *IEEE MultiMedia*, vol. 19, no. 1, pp. 40–49, 2012.

[10] J. Krauskopf and G. Karl, "Color Discrimination and Adaptation," *Vision Research*, vol. 32, no. 11, pp. 2165–2175, January 1992.

[11] K.-M. Cho, J.-H. Jang, and K.-S. Hong, "Adaptive Skin-Color Filter," *Pattern Recognition*, vol. 34, no. 5, pp. 1067–1073, May 2001.

[12] R.M. Jusoh, N. Hamzah, M.H. Marhaban, and N.M.A. Alias, "Skin Detection Based on Thresholding in RGB and Hue Component," *Proceedings of the 2010 IEEE Symposium on Industrial Electronics Applications*, pp. 515–517, October 2010, Penang, Malaysia.

[13] S. R. Vantaram and E. Saber, "Survey of contemporary trends in color image segmentation," *Journal of Electronic Imaging*, vol. 21, no. 4, pp. 040901–1–040901–28, October 2012.

[14] R.T. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 178–193, February 2005.

[15] G. Healey, "Segmenting Images Using Normalized Color," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, pp. 64–73, January 1992.

[16] B. A. Maxwell and S. A. Shafer, "Physics-Based Segmentation of Complex Objects Using Multiple Hypotheses of Image Formation," *Computer Vision and Image Understanding*, vol. 65, no. 2, pp. 269–295, November 1997.

[17] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, pp. 604–610, October 2005, Montbonnot, France.

[18] Y. Tarabalka, J.A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2973–2987, August 2009.

[19] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 313–320, June 2003, Urbana, IL, USA.

[20] A.L.N. Fred and A.K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 835–850, June 2005.

[21] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, and S. Lafuente-Arroyo, "Goal evaluation of segmentation algorithms for traffic sign recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 917–930, December 2010.

[22] S.L. Phung, A. Bouzerdoum, and Sr. Chai, D., "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 148–154, January 2005.

[23] C-I Chang, Y. Du, J. Wang, S-M Guo, and P.D. Thouin, "Survey and comparative analysis of entropy and relative entropy thresholding techniques," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 153, no. 6, pp. 837–850, December 2006.

[24] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 787–800, July 2003.

[25] V. Grau, A. U J Mewes, M. Alcaniz, R. Kikinis, and S.K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 447–458, April 2004.

[26] A.J. Round, A.W.G. Duller, and P.J. Fish, "Colour Segmentation for Lesion Classification," *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 582–585, November 1997, Chicago, IL, USA.

[27] Y. Deng and B.S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800–810, August 2001.

[28] M.E. Plissiti, D.I. Fotiadis, L.K. Michalis, and G.E. Bozios, "An automated method for lumen and media-adventitia border detection in a sequence of ivus frames," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 2, pp. 131–141, June 2004.

[29] N. Funakubo, "Feature Extraction of Color Texture Using Neural Networks for Region Segmentation," *Proceedings of the 20th Annual Conference of IEEE Industrial Electronics*, vol. 2, pp. 852–856, September 1994, Bologna, Italy.

[30] T. Carron and P. Lambert, "Color Edge Detector Using Jointly Hue, Saturation and Intensity," *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 977–981, November 1994, Austin, TX, USA.

[31] T.F. Chan and L.A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, February 2001.

[32] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 5:1–5:60, May 2008, New York, NY, USA.

[33] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, July 2002.

[34] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, August 2000.

[35] F. Zhu, M. Bosch, N. Khanna, C.J. Boushey, and E.J. Delp, "Multilevel segmentation for food classification in dietary assessment," *Proceedings of 7th International Symposium on Image and Signal Processing and Analysis*, pp. 337–342, September 2008, Dubrovnik, Croatia.

[36] D.E. Ilea and P.F. Whelan, "CTex - an adaptive unsupervised segmentation algorithm based on color-texture coherence," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1926–1939, October 2008.

[37] J. Li and J.Z. W., "Studying digital imagery of ancient paintings by mixtures of stochastic models," *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 340–353, March 2004.

[38] H. Muller, T. Pun, and D. Squire, "Learning from user behavior in image retrieval: Application of market basket analysis," *International Journal of Computer Vision*, vol. 56, pp. 65–77, January 2004.

[39] J. He, H. Tong, M. Li, H-J. Zhang, and C. Zhang, "Mean version space: a new active learning method for content-based image retrieval," *Proceedings of the ACM SIGMM international workshop on Multimedia information retrieval*, pp. 15–22, October 2004, New York, NY, USA.

[40] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "An efficient and effective region-based image retrieval framework," *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 699–709, May 2004.

[41] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2161–2168, June 2006, Washington, DC, USA.

[42] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," *Proceedings of the ACM international conference on Multimedia*, pp. 107–118, October 2001, Ottawa, Canada.

[43] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 924–937, August 2003.

[44] D.-C. Tseng, Y.-F. Li, and C.-T. Tung, "Circular Histogram Thresholding for Color Image Segmentation," *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, vol. 2, pp. 673–676, August 1995, Montreal, Canada.

[45] D.-C. Tseng and C.-H. Chang, "Color Segmentation Using Perceptual Attributes," *Proceedings of the 11th IAPR International Conference on Pattern Recognition*, vol. 3, pp. 228–231, September 1992, La Haye, Holland.

[46] J.D. Brand and J.S.D. Mason, "Skin Probability Map and Its Use in Face Detection," *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, pp. 1034–1037, October 2001, Thessaloniki, Greece.

[47] Z. Xue, D. Shen, and S. Wong, "Tissue Probability Map Constrained CLASSIC for Increased Accuracy and Robustness in Serial Image Segmentation," *Proceedings of the 2009 SPIE Symposium on Medical Imaging*, vol. 7258, pp. 725904–1–9, February 2009, Lake Buena Vista, FL, USA.

[48] C. Poynton, *Digital Video and HDTV Algorithms and Interfaces*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2003.

A Correlative Analysis Process in a Visual Analytics Environment

Abish Malik*
Purdue University, USA

Ross Maciejewski†
Arizona State University, USA

Yun Jang‡
Sejong University, South Korea

Whitney Huang*
Purdue University, USA

Niklas Elmqvist*
Purdue University, USA

David S. Ebert*
Purdue University, USA

ABSTRACT

Finding patterns and trends in spatial and temporal datasets has been a long studied problem in statistics and different domains of science. This paper presents a visual analytics approach for the interactive exploration and analysis of spatiotemporal correlations among multivariate datasets. Our approach enables users to discover correlations and explore potentially causal or predictive links at different spatiotemporal aggregation levels among the datasets, and allows them to understand the underlying statistical foundations that precede the analysis. Our technique utilizes the Pearson's product-moment correlation coefficient and factors in the lead or lag between different datasets to detect trends and periodic patterns amongst them.

Keywords: Visual analytics, correlative analysis

1 INTRODUCTION

With more and more agencies collecting and storing data pertaining to different processes, modern datasets often become complex and too large to handle, triggering a need for solutions that make the exploration and analysis of these datasets more manageable. One specific opportunity created by such multivariate and multisource data is the potential to look for correlations and explore possible causal or predictive links from these datasets. The challenges with such analysis, though, include end-users and decision makers understanding the underlying statistical algorithms, applying the algorithms at the appropriate temporal and spatial scale, and dealing with the noisiness of the real-world data. Therefore, a visual analytics environment to enable analysts to explore and understand potential correlations at different scales and with various certainties is necessary for effective decision making.

In this paper, we present a visual analytics approach for exploring correlations among multivariate spatiotemporal datasets. Our correlative analysis framework enables users to detect trends and patterns among datasets using statistical techniques that are oftentimes complex and difficult to understand by both novice and advanced users. Furthermore, we argue that if the users were to rely on mere statistical results without the aid of any interactive visualization or visual analytic tools, there remains the potential that they may miss certain important details. The use of visual representations of data and of the different analytical processes minimizes these risks and provides more insights to users than traditional methods.

Our correlative visual analytics framework also enables users to understand the benefits and harms of performing their analyses over aggregate statistics, where datasets are aggregated and compared over spatial (e.g., by census blocks or census tracts), or temporal (e.g., by day, week, month or year) units. For example, users may

---

*e-mail: {amalik|huang251|elm|ebertd}@purdue.edu

†e-mail: rmacieje@asu.edu

‡e-mail: jangy@sejong.edu

find no relevant patterns or anomalies when they perform an analysis over large spatial or temporal regions but may discover trends over customized spatiotemporal regions. We incorporate this concept in our system, and allow users to interactively correlate any data subsets to assist them in the discovery of underlying patterns and anomalies.

Our current work builds upon our previous framework [26] that has been developed to explore and analyze spatiotemporal datasets. A screenshot of our system is shown in Figure 1. This system provides users with a suite of analytical tools that are coupled with interactive visual interfaces for exploring criminal, traffic and civil (CTC) datasets. Our system allows users to observe patterns and quickly identify regions with higher probabilities of activity. Our current work focuses on interactive correlative visualization and analysis methods of spatiotemporal datasets in a visual analytics environment. We emphasize that although the examples provided in this paper are based in the law enforcement domain, our approach is extendible to any spatiotemporal dataset framework.

2 RELATED WORK

In recent years, there has been much work exploring multivariate datasets in the spatiotemporal domain that enables the discovery and detection of trends, patterns and anomalies (e.g., [14, 25, 30, 32]). In this section, we discuss previous work in the areas of spatiotemporal analysis and visualization.

2.1 Time series visualizations

The visualization of time series data is one of the most common problems in any data domain, and, as such, much work has been done to address the challenges associated with visualizing time series data. Common methods of visualizing time series datasets include line graphs and bar charts. An overview of some existing techniques for visualizing time series datasets can be found in [1]. Our work primarily utilizes line and bar graphs, clock views and calendar view visualizations [33] to support a temporal visualization of the data.

Researchers have also explored different techniques that provide interactive visual analysis methods for time series datasets. Bade et al. [4] present different interactive visualization techniques that reveal data at different levels of detail and abstraction. They augment the standard temporal visualizations to include implicit information with regards to the data as well as *a priori* knowledge pertaining to the data. Zhao et al. [34] present an exploratory time-series visualization technique that allows users to focus on areas of interest in the data through magnification and visual filtering. Javed et al. [19] also explore user performance for different line graph techniques that involve multiple time series to provide guidelines for designers designing temporal visualization applications. Buono et al. [8] provide a review of techniques that enable users to interactively query time series datasets and present a data exploration tool that allows users to explore multidimensional datasets in an Overview+Detail design. Their system enables users to interactively search for existing patterns to find similar temporal occurrences and trends. We adapt an Overview+Detail design for the visualization of datasets by providing users with different temporal aggregation levels.
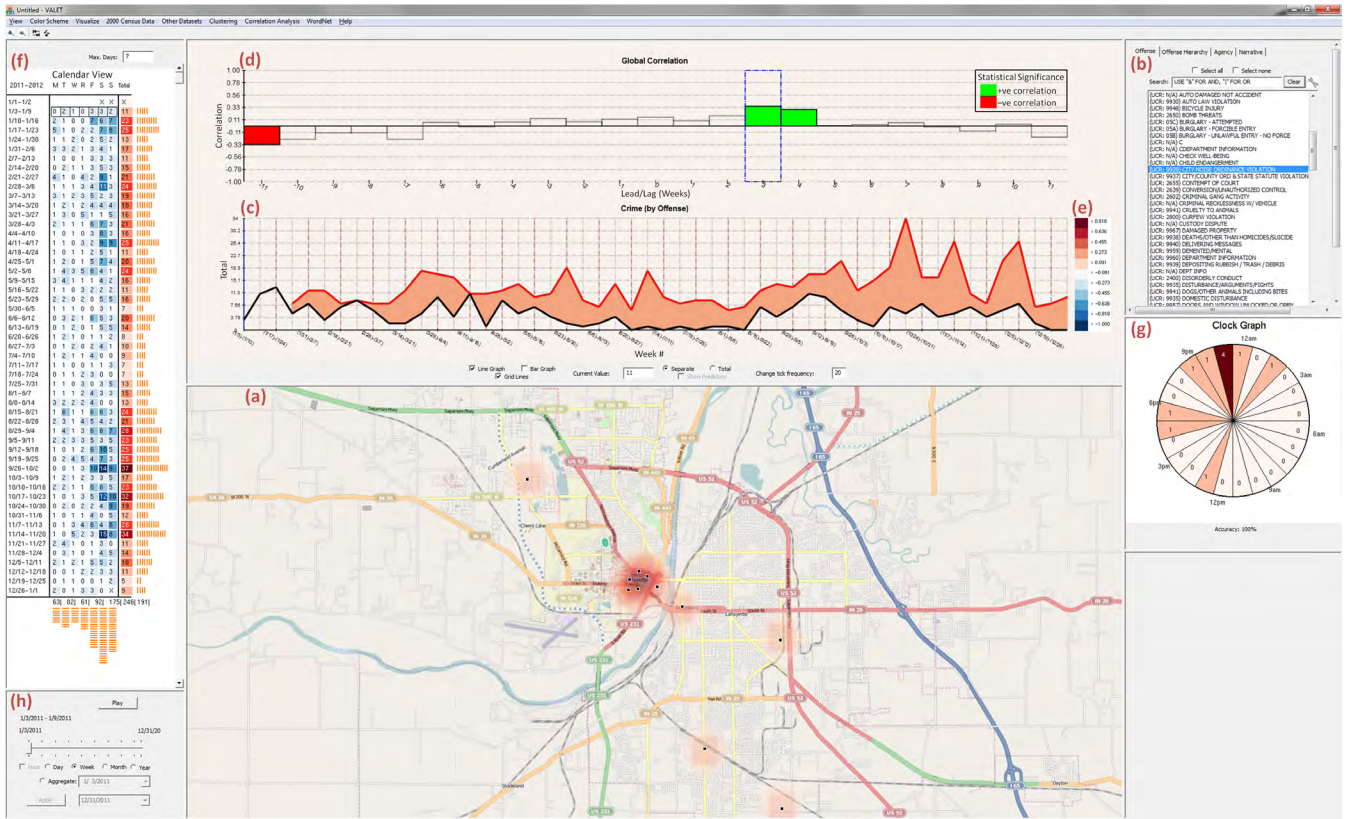
33

Figure 1: A screenshot of our correlative visual analytics system. Here, the user is visualizing all crimes related to noise complaints and drunkenness/public intoxication in Tippecanoe County, IN. The main viewing area (a) shows the map view with the points showing the locations of the selected offenses on the map for the week of 1/3/2011 to 1/9/2011. The top-right window (b) shows an interactive menu that allows users to filter through the variables of the datasets. The middle window (c) shows the weekly time-series view of the selected incident reports, where the user is performing a correlative analysis of the system, the results of which are displayed in the correlation vs. lead/lag graph (d). The region of overlap between the time series in (c) is colored by the corresponding correlation value based on a divergent color scale (e). The left window (f) shows the calendar view for the year 2011, and the right window (g) shows a clock view of the selected incident report data for the week of 1/3/2011 to 1/9/2011. Finally, the bottom-left window (h) shows the time slider with radio buttons that allow different temporal aggregation levels.

Datasets that are large in size often tend to be incomplete, un-reliable and contradictory. This often leads to wrong conclusions and results and it becomes imperative for system designers to make users aware of data reliability issues [20]. Aigner et al. [2] pro-vide a visualization technique that represents uncertainties in the temporal domain. Their approach provides users with temporal glyphs that support planning and controlling tasks. We make a note here that data accuracy and reliability becomes an important measure for analysts to review and understand in any correlative analysis process. This becomes especially important to consider when the datasets being analyzed are sparse, as correlating such datasets may produce unreliable results. Our system allows users to visualize sparse datasets during the analysis process by showing a percent of cases that have zero values in the time series. This is especially helpful in case of incident report datasets (e.g., CTC in-cidents) where there may be a certain categories that have very few incidents thereby leading to sparse datasets. In addition, we also allow users to choose different data aggregation levels as a method of aggregating sparse data to appropriate levels, thereby allowing them to utilize their domain knowledge to determine the appropri-ate scales for analyses.

2.2 Time series data pattern and anomaly detection

Works in the field of time series data exploration, pattern and anomaly detection includes work by Krstajic et al. [22] who present an incremental time-series visualization technique that interactively

distorts time-based representations of multiple event datasets. Their technique uses a timeline distortion technique to accommodate for recent data within individual items. The authors note that many analytical tasks can be more efficiently solved by using interactive visual analytic tools that include the analyst at different stages of the exploration process. We incorporate the same concept in our sys-tem, and provide a correlative analysis framework that incorporates the analyst in the visual analytics loop. Hao et al. [16] also present a framework for visualizing large sets of time series datasets. Their approach involves importance-driven space-filling layout schemes for time series datasets. Lin et al. [23] present a time series pat-tern discovery and visualization system that summarizes the global and local structures of time series datasets. Their system enables the discovery of motifs and anomalies and uses a coefficient to measure the dissimilarity between any two time series. Kincaid [21] uses a Focus+Context approach to visualize electronic test and measurement system datasets and allows users to filter and detect and visualize computationally detected motifs in the data. Malik et al. [27] utilize a visual analytics approach for exploring linear cor-relations across a variety of spatial aggregations to allow analysts to explore global temporal correlations, as well as explore the un-derlying spatial factors that may potentially influence trends among their datasets. Corman and Mocan [12] utilize high frequency time series criminal incident datasets and perform time series analyses to explore the relationship among crime, deterrence and drug use. Our approach deals with determining the correlations among datasets in

a visual analytics environment that enable analysts to compare different datasets by factoring in temporal leads/lags among them.

Other methods in statistics for detecting relationships between time series datasets include computing the cross-correlation of two different signals [15]. This approach applies a time-lag function to one of the time series signals while keeping the other time series fixed. The process of computing the correlation of a time series signal with itself (with lag) is called autocorrelation, and the resulting plot of the sample autocorrelations plotted against the time lags is called a correlogram [10]. This technique is widely used in the detection of repeating patterns and randomness within the datasets. We adapt a similar process of correlating the datasets by sliding the datasets against one another and providing an interactive interface that enables users to explore and understand potential correlations between the datasets at different spatiotemporal scales.

Recent work in exploring time series similarity detection methods include Morse and Patel [28] who propose a method that can be used to evaluate different threshold value techniques that measure the similarity between time series datasets. Bollobas et al. [5] present deterministic and randomized algorithms for determining the similarity between two time series datasets. Das et al. [13] also consider the problem of finding rules that relate different patterns within the same time series. They focus on the discovery of local patterns in multivariate time series datasets. Many of these data mining techniques rely on comparing elements of one time series dataset with the elements of the other datasets.

2.3 Spatiotemporal exploration of predictive links

Researchers have also utilized multivariate datasets for forecasting purposes and have developed systems that factor in categorical datasets in order to generate alerts and projections into the future. Buono et al. [9] present a data driven forecasting method that employs pattern matching search using historic datasets. Maciejewski et al. [24] also implement predictive analytical tools in analyzing syndromic surveillance data. Toole et al. [18] utilize cross-correlation measures and other techniques to identify spatiotemporal patterns using multivariate datasets. Rodrigues and Diggle [29] adapt Log-Cox processes to model the spatiotemporal crime intensities where the probabilistic predictions of the crime intensities are made in the Bayesian framework. Andrienko et al. [3] apply combinations of clustering and dimensionality reduction methods to support the visual analysis of spatiotemporal datasets.

3 CORRELATIVE EXPLORATION IN A VISUAL ANALYTICS ENVIRONMENT

Understanding relationships between spatial and temporal trends among multivariate datasets is an important part in the analysis process of spatiotemporal datasets. Such analyses often act as a precursor to creating predictive models from the data. Moreover, such analyses can provide invaluable insights into the workings of real world environments and can aid analysts in their hypothesis generation and exploration process. However, this process of determining these potential relationships remains a challenging problem due to reasons ranging from datasets being noisy to analysts being strained due to the nature and complexity of their datasets.

Our approach focuses on determining the correlations between different spatiotemporal datasets. Correlation is a single number that describes the relationship between two variables. In the case of temporal datasets, we use correlation to describe the degree to which two variables are related to each other. Our method applies to collections of categorical spatiotemporal datasets and allows users to compute the correlations between any pair of variables among any of the given datasets, each of which may consist of different number of variables. We use the Pearson product-moment correlation coefficient (*Pearson's correlation* - r_{xy}) in our analysis, which

is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{(N-1)S_x S_y} \quad (1)$$

where N is the length of the time series datasets, $\bar{x}$ and $\bar{y}$ are the means of the time series x and y respectively, and S_x and S_y are the standard deviations of the time series x and y between the regions of overlap. We also apply a two sided t-test to check if the correlation obtained is significant within a 95% confidence interval with N-2 degrees of freedom. This two sided t-test is given by:

$$t = \frac{r_{xy}}{\sqrt{(1 - r_{xy}^2)/(N-2)}} \quad (2)$$

We also observe that there may be different forces that determine the nature of the relationships between different datasets. For example, one event may be a precursor to a series of other events. This would require the exploration of leads and lags between different datasets in order to understand if one dataset is related to the other at a different temporal location (i.e., a lead or lag value). Our system incorporates this concept in the analysis process, and allows users to determine the lead or lag that maximizes the correlation, thereby enabling them to determine the lead/lag values that maximize the similarity among the datasets.

3.1 Visual Analytics Environment

Our visual analytics framework builds upon our previous work [26] and provides analysts and decision makers with the ability to visualize and model criminal, traffic and civil (CTC) incidents. Figure 1 shows a screenshot of our correlative visual analytics system. Our system has been organized into a dashboard view and provides users with linked dynamic windows to explore their spatiotemporal datasets. The main viewing region (Figure 1 (a)) is the map view that plots all selected incidents on a map. We utilize a kernel density estimation technique [24] to generate heatmaps to allow users to quickly detect hotspots among the CTC incidents. Figure 1 (b) shows an interactive menu that is used to filter through the different dataset variables (e.g., offenses, agencies). Figure 1 (c) shows a time series view of the selected incidents, along with a correlation vs. lead/lag graph (Figure 1 (d)) that shows the correlation of one dataset against the other at different lead and lags between them. The region of overlap between the time series (Figure 1 (c)) is also encoded by a color on a divergent color scheme [7] (Figure 1 (e)) to reflect the correlation value at the current lead/lag value. These points will be expanded upon later in Section 3.2 of the paper. Figure 1 (f) implements the calendar view approach adapted from [31] and displays the time series data in the format of a calendar, with the rows corresponding to weeks and the columns corresponding to the days of the week. The bar charts at the end of each row and column of the calendar view encode the sum of incidents for a particular week and day, respectively. Figure 1 (g) shows a clock view that shows the hourly temporal data in the form of a clock. Finally, Figure 1 (h) shows the interactive time slider that allows users to temporally slide through their datasets and provides them with multiple temporal data aggregation levels.

3.2 Temporal Correlation Exploration

We provide three different interactive options for users to determine the temporal correlation between datasets. The first option involves providing the user with control over the correlation computation process and allows him/her to directly set the lead or lag between the datasets for use in correlation calculations by manually dragging any of the selected time series datasets. The second option automatically computes the correlations at different leads or lags between the selected datasets and outputs the lead/lag that maximizes the correlation. The third option utilizes a brute force algorithm and computes the correlations between all possible permutations of the different data variables and saves the results in the decreasing order
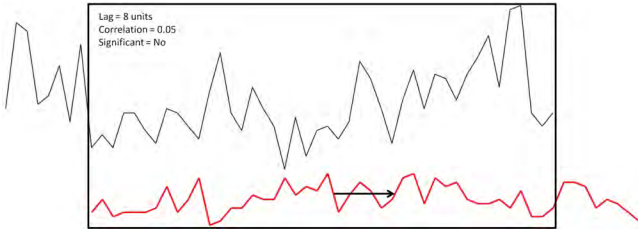
Figure 2: Interactive calculation of correlation between datasets. The user has clicked and dragged the time series data in red to the right by 8 units. The system automatically computes the correlation and significance (using the t-test) of the two time series data between regions of overlap (shown by the black rectangle).

of the number of significant correlation values. This allows analysts to review all potentially significant correlations among their datasets. We describe these options in detail below.

3.2.1 Manual dragging of datasets

This option allows users to determine the correlation between any selected datasets at any desired lag/lead. In order to determine the correlations between these different time series datasets, our system allows users to simply click on a line on the time series graph (Figure 1 (c)) and drag it either to the left or the right. This action automatically computes the correlation between the selected time series dataset, and the other selected datasets and interactively displays the results to the user. In addition to displaying the correlation value, the system also colors the region of overlap between the time series datasets on a divergent color scale (Figure 1 (e)) to encode the current correlation value. The system also displays the current lead/lag value corresponding to the length of the difference between the time series datasets (which, in turn, corresponds to the length of the drag by the user). In our implementation, if the user selects multiple datasets at the same time and clicks on a line graph to compute the correlation, we take a sum of all the other temporal datasets and correlate the selected time series dataset with this summed dataset. We plan on investigating other methods of computing the correlations among multiple datasets and leave this as future work. We also note that this process of manually dragging the datasets provides users with complete control over the correlation process by setting any desired lead/lag between the time series datasets. This enables them to visually explore the potential correlations between the datasets. Furthermore, users may apply any desired spatiotemporal filters to their datasets and acquire the time series for any of the desired variables pertaining to their datasets.

Since Pearson's correlation calculation requires the two time series datasets be of equal lengths, we utilize two different approaches with regards to the non-overlapping regions of the datasets when the user introduces a lead/lag by dragging the time series graph either to the left or right:

- *Calculating the correlation between the regions of overlap*: This approach calculates the correlation between only the overlapped regions of the time series datasets, and ignores all the non-overlapping temporal regions. Correspondingly, when the time series x lags behind the time series y, the correlation calculation (Equation 1) changes to:

$$r_{xy}(d) = \frac{\sum_{i=1}^{N_d}(x_{i+d} - \bar{x})(y_i - \bar{y})}{(N_d - 1)S_x S_y} \quad (3)$$

where $N_d = (N - d)$, $\bar{x}$ and $\bar{y}$ are the means of the time series x and y respectively, S_x and S_y are the standard deviations of the time series x and y between the regions of overlap, and $d = \{0, 1, ..., \lceil \frac{N}{4} \rceil\}$ is the *lag* between time series x with respect
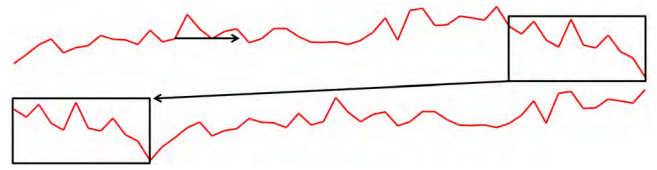
Figure 3: Cycling the non-overlapping region of the data. The user has clicked the top time series data and dragged it to the right by 12 units. The 12 units at the end are cycled to the beginning of the time series, resulting in the new time series data shown at the bottom.

to y. Similarly, when the time series x *leads* y by a value of d units, Equation 1 changes to:

$$r_{xy}(d) = \frac{\sum_{i=1}^{N_d}(x_i - \bar{x})(y_{i+d} - \bar{y})}{(N_d - 1)S_x S_y} \quad (4)$$

The corresponding t-test equation (Equation 2) thus becomes:

$$t = \frac{r_{xy}(d)}{\sqrt{(1 - r_{xy}^2(d))/(N_d - 2)}} \quad (5)$$

An example of this method can be seen in Figure 2, where the user has chosen to visualize two temporal datasets, and has dragged the series in the red color to the right by a distance of 8 units. The other time series (black color) remains stationary. We then compute the correlation of the two time series between the overlapping region only and display this value as a number, along with the corresponding significance as a result of applying the t-test.

- *Cycling the non-overlapping regions*: This method allows users to cycle the non-overlapping temporal regions of a selected time series dataset to either its beginning or end as the user drags the time series to the right or left on the graph, respectively. This cycling strategy is demonstrated in Figure 3, where the user has dragged the top time series to the right by 12 units, resulting in the time series shown at the bottom of the Figure. The system automatically cycles the points at the end of the top time series to its beginning, resulting in the time series shown at the bottom. This newly generated time series is then utilized in the correlation calculations (using Equations 1 and 2). This approach, coupled with temporal region selection described in Section 3.2.3, allows analysts to determine the cyclic trends in their datasets.

3.2.2 Automatic correlation computation against lead/lags

In addition to computing the correlation at a desired lead/lag dictated by the user (Section 3.2.1), the system also provides users with the option to automatically determine the lead/lag that maximizes the correlation between the selected datasets. When this option is selected, the user starts by simply clicking on the desired time series plot that he/she wants to perform the correlative analysis on. Figure 1 (c) shows an example where the user is visualizing the weekly drunkenness/public intoxication offenses against noise complaints for Tippecanoe County, Indiana, U.S.A. for all 52 weeks of the year 2011. In this example, the user has selected the time series plot of drunkenness/public intoxication offenses (shown in red color), and is analyzing how these compare against noise complaints. The system computes the correlation values at corresponding lead/lag values (ranging between $\pm N/4$, where N is the total length of the datasets), and displays them as a bar graph of the computed correlation values with respect to the lead/lag (Figure 1 (d)). The $\pm N/4$ metric has been adapted from [6], and is chosen because as the lead/lag value increases, the region of overlap between the datasets (and hence the correlation sample size) decreases, thereby making the correlation estimate of the overlapping

datasets less reliable. We also note that the resulting plot (Figure 1 (d)) resembles a correlogram in that we plot the results of the correlation process as correlation values against time lags.

The system also applies the t-test for the correlation values obtained at each lead/lag value and checks whether the correlation is significant or not at a given lead/lag. The statistically significant correlation values are highlighted with green colored bars (for positive correlation), or red colored bars (for negative correlation) at the corresponding lead/lag value. This is done in order to reduce the risk of users making wrong inferences from the correlation results, as positive and negative correlation values have different statistical meanings and implications in real world scenarios. The system also animates the two graphs by iterating through lead/lag values with a step size of 1 time unit, updating the two graphs simultaneously. This animation starts from a lead value of $-N/4$ and ends at a lag value of $+N/4$, and lasts for about 5 seconds for $N = 365$ days. We find this animation length to be enough time for users to understand the underlying process [17]. The corresponding correlation vs. lead/lag graph is dynamically linked to the main graph, and the correlation value corresponding to the current lead/lag value is highlighted on the global correlation graph (using a blue colored outline at the current lead/lag value).

Finally, when the current lead/lag value reaches the maximum lag value of $+N/4$, the animation proceeds and ends at the lead/lag value that maximizes the correlation between the datasets (this happens at a lead/lag value of 3 weeks in Figure 1 (d)). Note that the system also allows users to interactively click on any of the time series datasets and drag it to the left or right to interactively move it (just like in the manual correlation option described in Section 3.2.1), and visualize the datasets at a desired lead/lag corresponding to the correlation vs. lead/lag graph. We note that the user may choose to disable the animation at any time in which case the system resorts to displaying only the end results (e.g., Figure 1 (d)).

We further note that all the options described in Section 3.2.1 apply equally in the automatic correlation computation mode. Specifically, in the auto-computation mode, we allow users to: (a) discard the non-overlapping temporal regions while calculating the correlation values, and (b) cycle the non-overlapping region to either the beginning or the end of the selected time series.

3.2.3 Temporal windows for correlation computation

We provide users with options to restrict the temporal correlation calculations to a user specified temporal window. This feature allows users to interactively click and hold the left mouse button, and drag the mouse on the time series graph to select the temporal region between the two clicked locations. The user can then choose to perform the correlation analyses described in either of sections 3.2.1 and 3.2.2. The system automatically restricts the correlation analysis to the user selected temporal window. Furthermore, if the user chooses to cycle the non-overlapping temporal regions of their datasets in the correlation computation procedure, and selects a temporal window, the system replicates the selected time series within the temporal window and concatenates it to its beginning and end, thereby forming a new time series dataset. The system then uses this newly generated temporal dataset and performs the correlation analysis with the other (unaltered) user-selected datasets.

We observe that this temporal windowing method is especially useful when the user is interested in determining the correlations between different datasets within a specific temporal window. This may be because certain temporal distributions may be highly correlated only within certain temporal periods (e.g., holiday periods, weekends). For example, noise complaints and drunkenness/public intoxication offenses may have a high correlation pattern within university academic semesters, and may not have significant patterns within the summer months when most students are not on-campus.

3.2.4 Automatic computation of correlation between different datasets

In order to assist analysts in the discovery of potential high correlation between datasets, we utilize a brute-force algorithm that applies the method described in Section 3.2.2 to all possible combinations of the user-selected datasets. We then count the total number of significant correlation values at different lead/lag values (using the t-test), and save the results in descending order of the number of significant correlation values in a text file. In addition to this text file, we also provide users the option to save the corresponding global correlation graphs obtained for all the combinations as images on their local machines for the analysts to review. This method allows analysts to easily analyze and discover the datasets that are potentially correlated to one another. We note, however, that this process acts only as a starting point for analysis and only provides a hint towards possible correlations among the datasets. An analyst who has domain knowledge of the underlying datasets remains a necessary component in this correlative visual analytics process.

3.3 Correlation Exploration in Geospace

In addition to performing a temporal analysis of data, our system also provides analysts with a suite of geospatial data aggregation tools that allows them to compare the correlations between the incident distribution levels within and across selected spatial regions. This method allows users to restrict their analysis to specific regions and generate and test hypotheses pertinent to their domain knowledge of the datasets. We also allow analysts to factor in the lead/lag values obtained as a result of performing a temporal correlation analysis and visualize their datasets in geo-space by displaying the incidents at different temporal lead or lag levels. We describe both these methods in detail below.

3.3.1 Spatial region selection

We provide analysts with the ability to spatially aggregate their datasets by either drawing arbitrary regions on the map, or by an underlying spatial boundary (e.g., census block, county). This action aggregates the incidents falling within the selected spatial regions and extracts the temporal datasets of an aggregate of all user selected data categories for these regions. This allows users to compare the incidents between multiple spatial regions and determine how closely related the selected incident distributions are to both within, and across the spatial regions. We note that this spatial region selection also enables the correlation analysis and comparison of a smaller spatial region against the entire region. For example, an analyst may choose to compare the burglary crimes of a certain neighborhood by drawing a shape around the neighborhood, and draw another boundary around the entire city, and explore how the two selected regions compare against one another.

3.3.2 Visualizing the incident distributions in space at selected temporal lead/lag values

In order to visually compare the incidents across multiple temporal lead/lag values in space, we allow users to interactively select a temporal lead/lag value using the global correlation graph (Figure 1 (d)). Once the lead/lag value is selected, the spatial map display (Figure 1 (a)) shows the geo-spatial distribution of the incidents of the selected category at an offset applied to the current time. This offset is given by the selected lead/lag value, and the current time is dictated by the time slider (Figure 1 (g)). The geospatial distribution of the incidents of the other selected categories, on the other hand, are shown on the map at the current time. This allows the user to visualize the spatial incident distributions of one category with respect to the other selected categories at a selected lag, and allows them to visually observe spatial patterns among the
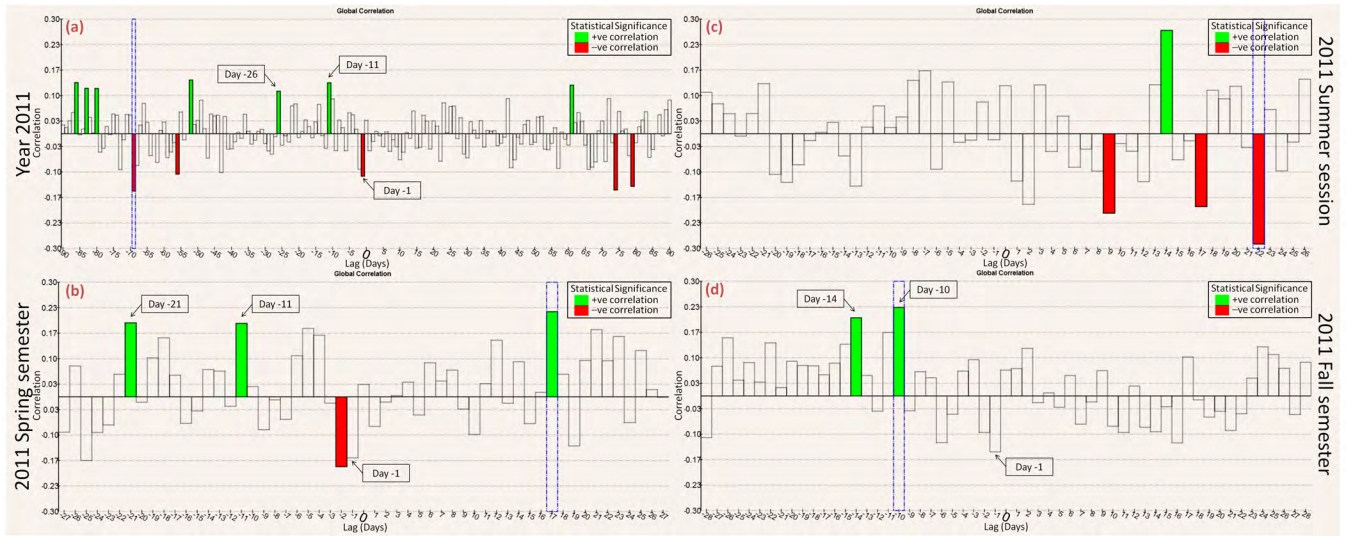
Figure 4: Correlating drug abuse violations and burglary crimes in Tippecanoe County, IN. The top-left graph (a) shows the correlation vs. lead/lag graph of the two crimes for the year of 2011. The correlation vs. lead/lag graphs shown in (b-d) correspond to correlating the two crimes within temporal windows of university academic Spring, Summer and Fall semesters.

different incident categories across time to explore potential relationships and causations.

4 CASE STUDY: EXPLORING CORRELATIONS IN CRIMINAL, TRAFFIC AND CIVIL (CTC) INCIDENT REPORT DATASETS

In this section, we demonstrate our work by applying the methods described so far to CTC incident reports from Tippecanoe County, IN, U.S.A. This dataset provides several categorical incident distributions including offense reported, entities against whom crimes are committed and agencies that respond to the incident. In addition, every incident has an associated geographic location and a time stamp. We now provide several examples that show how our correlative visual analytics system can be used and highlight some of the features of our system. We also discuss the insights that the system provides into the spatiotemporal CTC incident activity levels as a result of the analysis done using our system.

4.1 Drug abuse violations vs. burglaries

In this example, we provide a hypothetical scenario where an analyst is interested in exploring the correlations among drug abuse violations and burglaries (both residential and commercial) in order to find patterns and trends among the two crimes. To begin the process, the analyst filters and selects the offenses related to drug abuse violations and burglaries using the system's interactive menu (Figure 1 (b), and visualizes the daily incident temporal distributions of the offenses for the year 2011. The analyst now chooses to perform a correlative analysis over the entire region, and allows the system to determine the correlations between the two crimes at different temporal lead/lag values using the automatic computation method described in Section 3.2.2. Specifically, the analyst selects drug abuse violations (by clicking on its time series graph), and allows the system to compute the temporal correlations with respect to burglary crimes within a temporal lead/lag values of $\pm N/4$ days (where N is the total temporal duration of 365 days). This action animates the main graph by moving the time series data for drug abuse violations from a lead value of $-N/4$ to a lag value of $+N/4$, and computes the correlation between the overlapping regions of the time series datasets. The results are shown in the correlation vs. lead/lag graph alongside the main graph, as shown in Figure 4 (a).

The analyst first notes that there occurs a significant negative correlation between drug abuse violations and burglaries at a lead value of -1 days between the two offenses, providing an indication that an increase in drug abuse violation arrests or reports lead to a decrease in burglaries after one day. He hypothesizes that this may be because with more drug abuse violations and arrests that follow these drug busts, there may be less criminals out in the streets to commit crimes, thereby suggesting that these criminals may be have a relationship with the burglary crimes in the county. He also notes a significant correlation when drug abuse violations lead burglary crimes by 11 and 26 days (seen by the significant correlation at a lag value of -11 and -26 days from Figure 4 (a)). The analyst notes that these may be attributed to these drug offenders getting bailed out of jail and burglarizing a house in order to get some fast cash to satisfy their drug needs. He further observes other significant correlations at the outliers of the correlation vs. lead/lag graph. The analyst notes that these may be a consequence of the offenders serving jail time, or simply due to random chance. In this process however, he is quick to note that correlation does not imply a causation, and just because there is a strong positive or negative correlation between the two offenses, it does not imply that one offense directly causes the other. He notes that this is especially true since the offenses are aggregated over the entire county as different locations are governed by different spatial demographics and variables.

Temporal windowing by academic calendar

The analyst now notes that the population in Tippecanoe County fluctuates with the university academic calendar year as it is home to a large university, and investigates whether the academic calendar has an effect on drug abuse violations and burglaries in the county. As was done earlier, he repeats the automatic correlation computation analysis, but now restricts the analysis to the temporal windows of academic Spring, Summer and Fall semesters (using the technique described in Section 3.2.3). The correlation vs. lead/lag plots have been shown in Figures 4 (b-d). He first investigates whether any of the patterns discovered between drug abuse violations and burglaries in the time analysis for the entire year hold with respect to the academic semesters, which would hint to a pattern among the offenses that could be attributed to the academic calendar. He first observes a significant negative correlation when drug abuse violations lead burglary crimes by two days (at a lead/lag
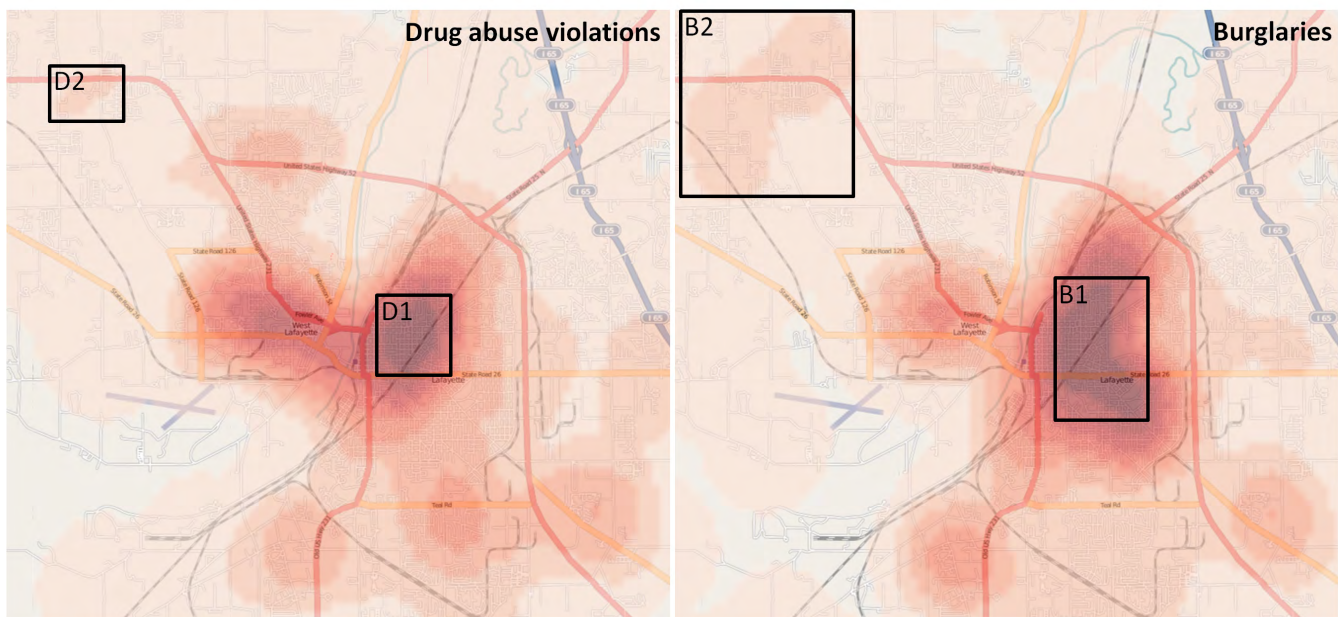
Figure 5: Geospatial heatmaps of drug abuse violations (Left) and burglary crimes (Right) for Tippecanoe County, IN for the year 2011.

value of -2 days in Figure 4 (b)) in the Spring semester. He also finds a negative correlation at a lead/lag value of -1 day (although the t-test indicates that this is not statistically significant). This pattern, however, does not hold in the Fall semester (Figure 4 (d)) (although a slightly non-significant negative correlation does exist at a lead/lag value of -2 days of the Fall semester). The analyst observes that the correlation vs. lead/lag graph for the summer session (Figure 4 (c)) shows no observable pattern as compared to the rest of the academic year, which he attributes to the fact that a majority of college students leave town during the summer months causing an economic shift in the region.

The analyst also observes that the significant positive correlation observed at a lead/lag value of -11 days in the global temporal analysis (Figure 4 (a)) is also observed around the same time in the Spring and Fall semesters. This provides further supporting evidence to his hypothesis that the drug offenders may be getting bailed out of jail by that time, resulting in a positive correlation value at around 11 days (thereby hinting to an increase of burglary offenses). With these results, the analyst may decide to investigate further whether there is a pattern between drug offenders getting bailed out and an increase in burglaries after a period of around 11 days in Tippecanoe county.

Exploring crime correlations over geospatial hotspots

The analysis done so far considers the drug abuse violations and burglaries for the entire county. However, different neighborhoods in a large spatial region have different attributes, and as such, behave differently in terms of crime patterns and trends. With this in mind, the analyst now decides to use the system and find the spatial hotspots of both drug abuse violations and burglary crimes. To achieve this, he aggregates the data first for drug violations for the year 2011, and uses our system's spatial heatmap feature to visualize the hotspots for drug violations. He interactively selects the region with the highest drug violations (region *D1* in Figure 5 (left)) for the year 2011 by drawing a rectangle over it. This action lets the system automatically extract the time series plot of all drug violations for this region. Similarly, the analyst visualizes the heatmap for burglary crimes, and interactively selects the hotspot (given by region *B1*) in the neighborhood of the selected drug violation hotspot, thereby extracting the time series plot for all the burglary

offenses for this region. He then performs the correlative analysis for these subsets of data, comparing the temporal trends of the drug abuse violation hotspot (region *D1*) against the neighboring burglary crime hotspot (region *B1*). The correlation vs. lead/lag graph resulting from this is shown in Figure 6 (a). The analyst observes something very unusual from these results. He finds a high significant positive correlation at a lag value of zero days between the two spatial regions, giving an indication that as drug abuse violation rates increase in the selected region, burglary rates also increase in the other neighborhood (and vice versa).

The analyst next selects the drug abuse violations for the region selected previously (region *D1*), and compares these to the burglary offenses for another burglary hotspot (region *B2*) in another city. The results of these are shown in Figure 6 (b). He however finds that these two regions do not share the same pattern that regions *D1* and *B1* do. Specifically, he finds no positive correlation at a lag of zero days, hinting to the fact that the drug abuse violations and burglary crimes do not share the same pattern across two different cities. The analyst now compares the drug abuse violations and burglary offenses for two other regions in the same neighborhood in the other city, that is, he compares the drug abuse violations for region *D2* and burglary offenses for region *B2*. These results are shown in Figure 6 (c). He finds that these two regions share the same pattern as regions *B1* and *D1*. That is, there exists a significant positive correlation between these two regions at a lag of zero days. Finally, he compares the drug abuse violations for region *D2* and the burglaries for region *B1*. The resulting correlation vs. lag/lead graph is shown in Figure 6 (d). Again, as was the case with regions *D1* and *B2*, there are no significant correlations at a lag value of zero days. These results suggest that drug abuse violations are highly correlated to burglary crimes in the *same* neighborhoods in Tippecanoe County at a lag of zero days, but not across different neighborhoods. He infers that this may be because the criminals lack transportation methods to travel across different cities, or that these criminals may have a tendency to target their own neighborhoods (a fact known from domain knowledge). This approach thus allows the analyst to analyze and compare the crime patterns of individual neighborhoods, and test and form hypothesis using their domain knowledge. As can be observed from this process, domain expert knowledge is a critical part in this analytical process.
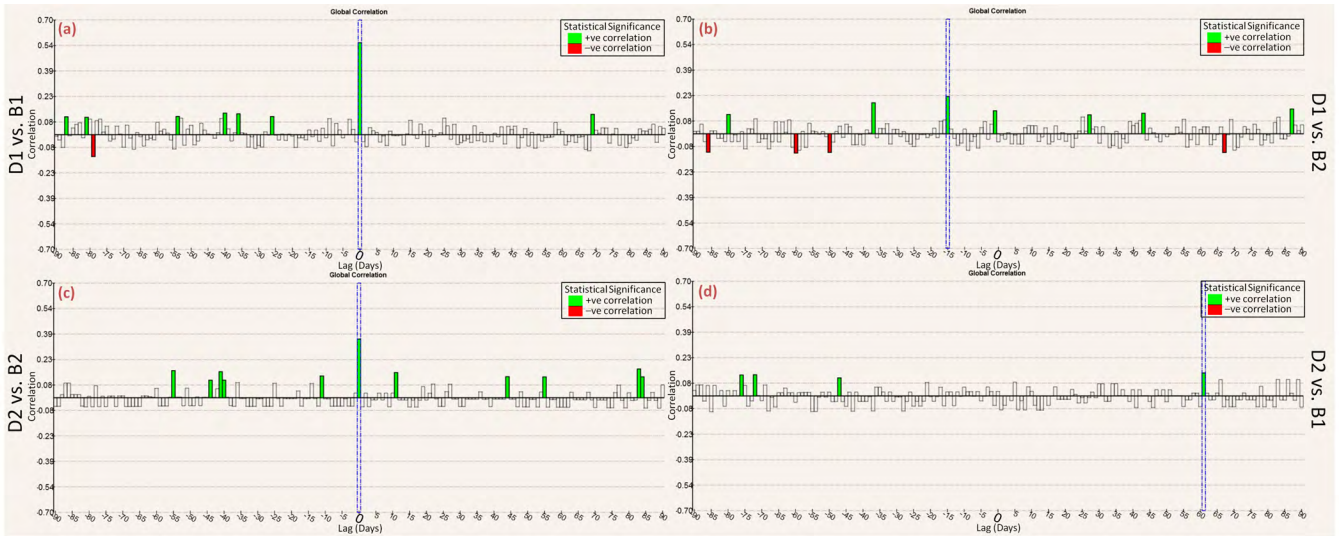
Figure 6: Correlating the temporal distributions of drug abuse violation hotspots against burglary crime hotspots in Tippecanoe County, IN. The graphs (a-d) correlate the drug abuse violation hotspots of regions *D1* and *D2* against the burglary hotspots of regions *B1* and *B2* in Figure 5.
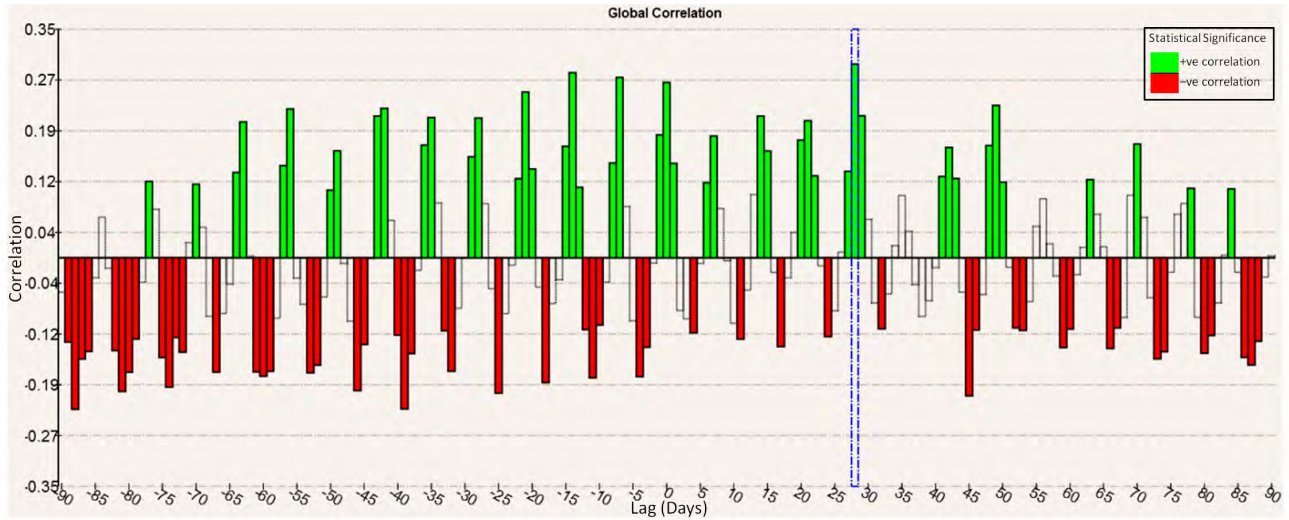

Figure 7: The correlation vs. lead/lag graph obtained as a result of correlating the drunkenness/public intoxication offenses against noise complaints for the year 2011 in the cities of Lafayette and West Lafayette, IN.

4.2 Drunkenness/public intoxication vs. noise complaints

We now provide an example in which an analyst is investigating the correlations between drunkenness/public intoxication offenses and noise complaints in the cities of Lafayette and West Lafayette, Indiana. He works under the hypothesis that there exists a causal relationship between the two offenses that repeats periodically by a period of seven days. This is because drunkenness and noise complaints generally go hand in hand, and usually have higher occurrences over the weekends. To begin this analysis, the analyst first selects both these offenses and visualizes the distributions of the two offenses by a weekly temporal aggregation level to detect any significant patterns over a weekly temporal scale. He first visualizes the offenses in geospace by interactively scrolling in time using the system's time slider (Figure 1 (h)), and observes that the hotspots for the two offenses occur near the university campus and have a strong co-occurrence with the locations of the local pubs and bars. He also uses the system's calendar view and observes that the daily

distributions are higher over the weekends, with most complaints occurring between Thursdays and Sundays. This is obvious to the analyst as he knows from his domain knowledge that college students tend to party over the weekends.

The analyst now applies the automatic correlation computation method (Section 3.2.2) on these two datasets with a weekly temporal aggregation scaling for the year 2011. The results from this analysis are shown in Figure 1 (d). In this image, the analyst has selected drunkenness/public intoxication offenses (shown in red color). The system automatically animates and moves this time series graph from a lead value of $-N/4$ to a lag value of $+N/4$ (here, $N = 52$ weeks) with a step size of 1. The system computes the correlation at each lead/lag step, and displays the resulting correlation in the correlation vs. lead/lag graph (Figure 1 (d)). However, the analyst notices no significant patterns as a result of this analysis. He notes however that the problem lies with the weekly temporal aggregation level chosen. He then chooses a daily temporal scale level, and performs the same analysis for the two datasets over a period of one year. The resulting correlation vs. lead/lag graph has
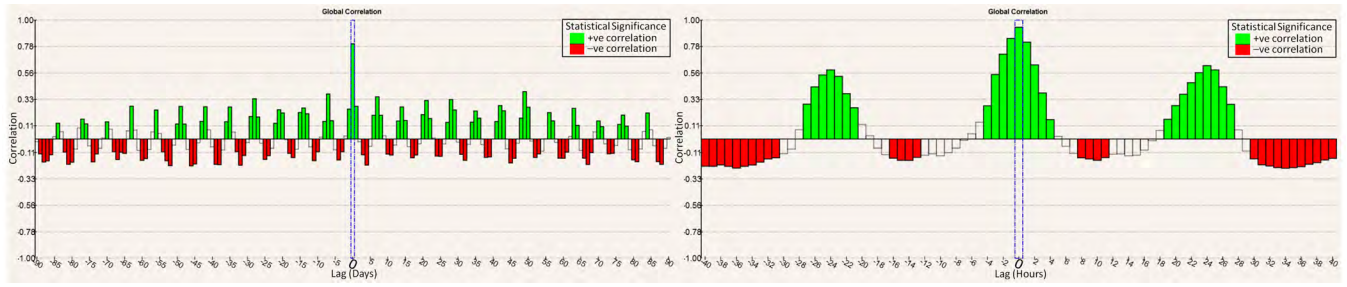
Figure 8: Correlating an aggregate of drunkenness/public intoxication and noise complaints against traffic accidents for the cities of Lafayette and West Lafayette, IN for the year 2011 at different temporal lead/lag values between the datasets. The datasets are correlated on a daily temporal scale in the left plot, and on an hourly temporal scale in the right plot.

been shown in Figure 7. These results show a clear day-of-the-week effect between the two datasets, with the correlation (significantly) peaking after every 7 days. The dashed blue region in the graph highlights the lag value that maximizes the correlation (28 days). We note here that if the system only displayed the lag value that maximizes the correlation, the analyst would not get the insights that are provided by the visualization resulting from this visual analytics process. This further demonstrates the value of visualization and visual analytics in this analysis process.

4.3 Drunkenness/public intoxication and noise complaints vs. traffic accidents

In our final example, we demonstrate our visual analytic correlation technique with a scenario in which an analyst is investigating the correlations between drunkenness/public intoxication offenses, noise complaints and traffic accidents in Tippecanoe County, IN. Using his domain knowledge, the analyst decides to compare the datasets for the spatial hotspots (as was done in the example provided in Section 4.1) for traffic accidents against drunkenness/public intoxication and noise complaints combined, and selects a temporal scaling by day for the year 2011. The correlation vs. lead/lag graph is shown in Figure 8 (Left), which shows a similar day-of-the-week effect as was observed in the example given in Section 4.2, with the pattern repeating every 7 days. Also, the correlation peak occurs at a lag of zero days, suggesting that drunkenness and noise complaints go hand-in-hand with traffic accidents, and are highly correlated to one another.

The analyst now decides to compare drunkenness/public intoxication and noise complaints combined against traffic accidents for the two hotspot regions selected for the year of 2011. In this case however, he decides to visualize the data by hour by binning the data from Monday to Sunday, with the total bins equaling 7×24. By this convention, the incidents for all Mondays for the year of 2011 fall in the bin range of 0-23, the incidents occurring on Tuesdays fall in the bin range of 24-47 hours, and so on. The results of running the correlative analysis on these two time series datasets have been shown in Figure 8 (Right). These results show a high positive correlation between drunkenness/public intoxication and noise complaints combined and traffic accidents within the selected regions within a lead/lag values between -5 to +4 hours, indicating that drunkenness/public intoxication, noise complaints and traffic accidents go hand-in-hand with each other within a 24 hour period.

Exploring geospatial patterns between selected offenses

The analyst now utilizes the method described in Section 3.3.2, and chooses to visualize the geospatial distribution of drunkenness/public intoxication and noise complaints against traffic accidents at an hourly lag between -5 to 0 hours to see if a spatial pattern emerges that shows a spatial association between these offenses and traffic accidents. He observes that there occur many traffic accidents after a few hours of noise/drunkenness complaints around

the neighborhoods of local pubs and bars. The analyst notes that this may indicate that more police officers should be on the lookout for traffic accidents when the rates of noise complaints and drunkenness increase around the regions of local pubs and bars.

5 CONCLUSIONS AND FUTURE WORK

In this work, we have presented a correlative analysis framework for computing spatiotemporal correlations in a visual analytics environment. We utilize the Pearson's correlation coefficient and factor in the lead or lag between different datasets for pattern and anomaly detection, thereby providing the tools for analysts and decision makers to quickly obtain the potential relationships between different datasets. We provide interactive tools that enable users to compute the correlations between multiple multivariate datasets. Our work explores the potential of visual analytics in the correlation exploration domain, and provides users with an interactive way of understanding and applying the underlying statistical algorithms at different appropriate temporal and spatial scales.

Our future work includes expanding our current work further in the spatial domain, and allowing users to interactively explore the correlations among the datasets in the spatial view. We plan on utilizing advanced statistical methods to enable users to detect spatial anomalies, that would provide a starting point in the correlative analysis process. We also plan on investigating different methods of computing and displaying correlations among multiple time series datasets at different lead/lag values (as was noted in Section 3.2.1). We note that Pearson's correlation coefficient enables a linear correlation between two variables, and we plan on investigating and incorporating non-linear correlation techniques among multivariate datasets in the future. We also note that a time series can be viewed as the sum of multiple components of variation, and can be separated into its various components using, for example, Seasonal-trend decomposition methods based on *loess* [11]. We plan on utilizing these techniques to separate the time series into its components and then applying our methods to further reveal finer trends among the datasets. We also plan on expanding our system to enable a correlation analysis of multiple data sources (e.g., CTC incidents vs. climate and moon phase datasets). We also plan on developing natural scale templates for automatically aggregating data to appropriate spatiotemporal levels for use in the correlation analysis process. Furthermore, we plan on conducting a formal user-evaluation of our system in the future to evaluate the efficiency of our system with varied amounts of data. Finally, we plan on deploying our correlative analysis framework to our public safety consortium members for further refinement and testing.

6 ACKNOWLEDGEMENTS

REFERENCES

[1] W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1), January/February 2008.

[2] W. Aigner, S. Miksch, B. Thurnher, and S. Biffl. Planninglines: Novel glyphs for representing temporal uncertainties and their evaluation. In *Proceedings of the 9th International Conference Information Visualisation*, 2005.

[3] G. Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. Von Landesberger, P. Bak, and D. Keim. Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns. *Computer Graphics Forum*, 29(3):913–922, 2010.

[4] R. Bade, S. Schlechtweg, and S. Miksch. Connecting time-oriented data and information to a coherent interactive visualization. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 105–112, New York, NY, USA, 2004. ACM.

[5] E. Bollob'as, G. Das, D. Gunopulos, and H. Mannila. Time-series similarity problems and well-separated geometric sets. In *Nordic Journal of Computing*, pages 454–456, 1997.

[6] G. Box. *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, 1970.

[7] C. A. Brewer. *Designing Better Maps: A Guide for GIS users*. ESRI Press, 2005.

[8] P. Buono, A. Aris, C. Plaisant, A. Khella, and B. Shneiderman. Interactive pattern search in time series. In *Proceedings of Conference on Visualization and Data Analysis*, pages 175–186, 2005.

[9] P. Buono, C. Plaisant, A. Simeone, A. Aris, G. Shmueli, and W. Jank. Similarity-based forecasting with simultaneous previews: A river plot interface for time series forecasting. In *Proceedings of the 11th International Conference Information Visualization*, pages 191–196, Washington, DC, USA, 2007. IEEE Computer Society.

[10] C. Chatfield. *Time-Series Forecasting*. Statistics (Chapman and Hall/CRC). Chapman & Hall/CRC, 2001.

[11] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. STL: A seasonal-trend decomposition procedure based on loess (with discussion). *Journal of Official Statistics*, 6:3–73, 1990.

[12] H. Corman and H. N. Mocan. A time-series analysis of crime, deterrence, and drug abuse in New York City. *American Economic Review*, 90(3):584–604, June 2000.

[13] G. Das, K. ip Lin, H. Mannila, G. Renganathan, and P. Smyth. Rule discovery from time series. pages 16–22. American Association for Artificial Intelligence, 1998.

[14] D. Guo, J. Chen, A. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1461 –1474, Nov.-Dec. 2006.

[15] J. D. Hamilton. *Time-series analysis*. Princeton Univerity Press, 1 edition, Jan. 1994.

[16] M. C. Hao, U. Dayal, D. A. Keim, and T. Schreck. Importance-driven visualization layouts for large time series data. In J. T. Stasko and M. O. Ward, editors, *Proceedings of the International Conference of Information Visualization*, page 27. IEEE Computer Society, 2005.

[17] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, Nov. 2007.

[18] J. P. J. Toole, N. Eagle. Quantifying behavioral data sets of criminal activity. Technical report, AAAI Symposium on Artificial Intelligence for Development, 2010.

[19] W. Javed, B. McDonnel, and N. Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE InfoVis 2010)*, 16(6):927–934, 2010.

[20] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. *Mastering the information age: Solving problems with Visual Analytics*. EuroGraphics, 2010.

[21] R. Kincaid. Signallens: Focus+context applied to electronic time series. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):900 –907, Nov.-Dec. 2010.

[22] M. Krstajic, E. Bertini, and D. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2432 –2439, Dec. 2011.

[23] J. Lin, E. Keogh, and S. Lonardi. Visualizing and discovering nontrivial patterns in large time series databases. *Information Visualization*, 4(2):61–82, 2005.

[24] R. Maciejewski, R. Hafen, S. Rudolph, S. Larew, M. Mitchell, W. Cleveland, and D. Ebert. Forecasting hotspots: A predictive analytics approach. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):440 –453, April 2011.

[25] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, M. Wade, and D. S. Ebert. A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16:205–220, Mar. - Apr. 2010.

[26] A. Malik, R. Maciejewski, T. F. Collins, and D. S. Ebert. Visual analytics law enforcement toolkit. In *IEEE International Conference on Technologies for Homeland Security*, pages 222–228, 2010.

[27] A. Malik, R. Maciejewski, E. Hodgess, and D. S. Ebert. Describing temporal correlation spatially in a visual analytics environment. In *Proceedings of the Hawaii International Conference on System Sciences*, pages 1–8, 2011.

[28] M. D. Morse and J. M. Patel. An efficient and accurate method for evaluating time series similarity. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, SIGMOD '07, pages 569–580, New York, NY, USA, 2007. ACM.

[29] A. Rodrigues and P. J. Diggle. Bayesian estimation and prediction for inhomogeneous spatiotemporal log-gaussian cox processes using low-rank models, with application to criminal surveillance. *Journal of the American Statistical Association*, 107(497):93–101, 2012.

[30] J. Stasko, C. Gorg, Z. Liu, and K. Singal. Jigsaw: Supporting investigative analysis through interactive visualization. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 131–138, 2007.

[31] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *INFOVIS '99: Proceedings of the 1999 IEEE Symposium on Information Visualization*, page 4, Washington, DC, USA, 1999. IEEE Computer Society.

[32] C. Weaver. Cross-filtered views for multidimensional visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 16:192–204, Mar. - Apr. 2010.

[33] J. J. V. Wijk and E. R. V. Selow. Cluster and calendar based visualization of time series data. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 4–9, San Francisco, CA, USA, 1999. IEEE Computer Society.

[34] J. Zhao, F. Chevalier, E. Pietriga, and R. Balakrishnan. Exploratory analysis of time-series with chronolenses. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2422 –2431, dec. 2011.

A Visual Analytics Process for Maritime Resource Allocation and Risk Assessment

Abish Malik *   Ross Maciejewski*, *Member, IEEE*   Ben Maule†   David S. Ebert*, *Fellow, IEEE*

*Purdue University Visualization and Analytics Center (PURVAC)   †United States Coast Guard

ABSTRACT

In this paper, we present our collaborative work with the U.S. Coast Guard's Ninth District and Atlantic Area Commands where we developed a visual analytics system to analyze historic response operations and assess the potential risks in the maritime environment associated with the hypothetical allocation of Coast Guard resources. The system includes linked views and interactive displays that enable the analysis of trends, patterns and anomalies among the U.S. Coast Guard search and rescue (SAR) operations and their associated sorties. Our system allows users to determine the potential change in risks associated with closing certain stations in terms of response time, potential lives and property lost and provides optimal direction as to the nearest available station. We provide maritime risk assessment tools that allow analysts to explore Coast Guard coverage for SAR operations and identify regions of high risk. The system also enables a thorough assessment of all SAR operations conducted by each Coast Guard station in the Great Lakes region. Our system demonstrates the effectiveness of visual analytics in analyzing risk within the maritime domain and is currently being used by analysts at the Coast Guard Atlantic Area.

Keywords: Visual analytics, risk assessment, Coast Guard

1 INTRODUCTION

As modern datasets increase in size and complexity, it becomes increasingly difficult for analysts and decision makers to extract actionable information for making effective decisions. In order to better facilitate the exploration of such datasets, tool sets are required that allow users to interact with their data and assist them in their analysis. Furthermore, such datasets can be utilized to explore the consequences and risks associated with making decisions, thereby providing insights to analysts and aiding them in making informed decisions.

Besides the sheer volume and complexity of such datasets, analysts must also deal with data quality issues, including uncertain, incomplete and contradictory data. Moreover, analysts are often faced with different decisions and are required to weigh all possible consequences of these decisions using such datasets in order to arrive at a solution that minimizes the associated risks within a given time constraint. Using traditional methods of sifting through sheets of data to explore potential risks can be highly inefficient and difficult due to the nature and size of these datasets. Therefore, advanced tools are required that enable a more timely exploration and analysis. Our work focuses on the use of visual analytics [17, 31] in the realm of risk assessment and analysis and demonstrates the effectiveness of visual analytics in this domain. The work described in this paper is based on the application of visual analytics to analyze historic response operations and assess the potential risks in

---

*e-mail: {amalik|rmacieje|ebertd}@purdue.edu

†email: Ben.J.Maule@uscg.mil

the maritime environment based on notational station closures. Our work was done in collaboration with the U.S. Coast Guard's Ninth District and Atlantic Area Commands that are responsible for all Coast Guard operations in the five U.S. Great Lakes. In particular, we focused on the Auxiliary stations that are staffed by Coast Guard volunteers and civilians. These Auxiliary stations assist their parent stations in their operations and usually operate on a seasonal basis using a small fleet of boats for conducting their operations. However, the number of Auxiliary personnel that volunteer their time at these stations has decreased over recent years. This has required Coast Guard analysts to develop possible courses of action and analyze the risks and benefits with each option. Several options include seasonal or weekend only staffing of these units, or at worst, closure. Closure, however, may involve increased risks to the boating public and a complete analysis of the risks associated with closing an Auxiliary station needs to be evaluated. The results of this type of analysis would assist the decision makers in determining the optimal course of action.

In particular, the analysts are interested in determining the spatial and temporal distribution of response cases and their associated sorties (a boat or an aircraft deployed to respond to an incident) for all SAR operations conducted in the Great Lakes and how closing certain Auxiliary stations affects the workload of the stations that absorb these cases. Coast Guard policy mandates the launch of a sortie within 30 minutes and have an asset (boat or aircraft) on scene within two hours of receiving a distress call [32]. Closing these stations implies a potential for longer response times that could potentially translate into the loss of lives and property.

To address these challenges, we developed a visual analytics system that supports decision making and risk assessment and allows an interactive analysis of trends, patterns and anomalies among the U.S. Coast Guard's Ninth District operations and their associated sorties. Our system, shown in Figure 1, allows enhanced exploration of multivariate spatiotemporal datasets. We have incorporated enhanced tools that enable maritime risk assessment and analysis. Our system includes linked spatiotemporal views for multivariate data exploration and analysis and allows users to determine the potential increase or decrease in risks associated with closing one or more Coast Guard stations. The system enables a thorough assessment of all operations conducted by each station. In addition, the system provides analysts with the tools to determine which Coast Guard stations are more optimally suited to assume control of the operations of the closed station(s) by comparing the distances from available stations to all SAR cases previously handled by the closed station(s). Our system features include the following:

- Risk profile visualizations and interactive risk assessment tools for exploring the impact of closing Coast Guard stations

- Optimization algorithms that assist with the interactive exploration of case load distribution in resource allocation

- Linked filters combined with spatial and temporal views for interactive risk analysis/exploration

Our work focuses on providing analysts with interactive visual analytics tools that equip them to deal with risk assessment scenar-
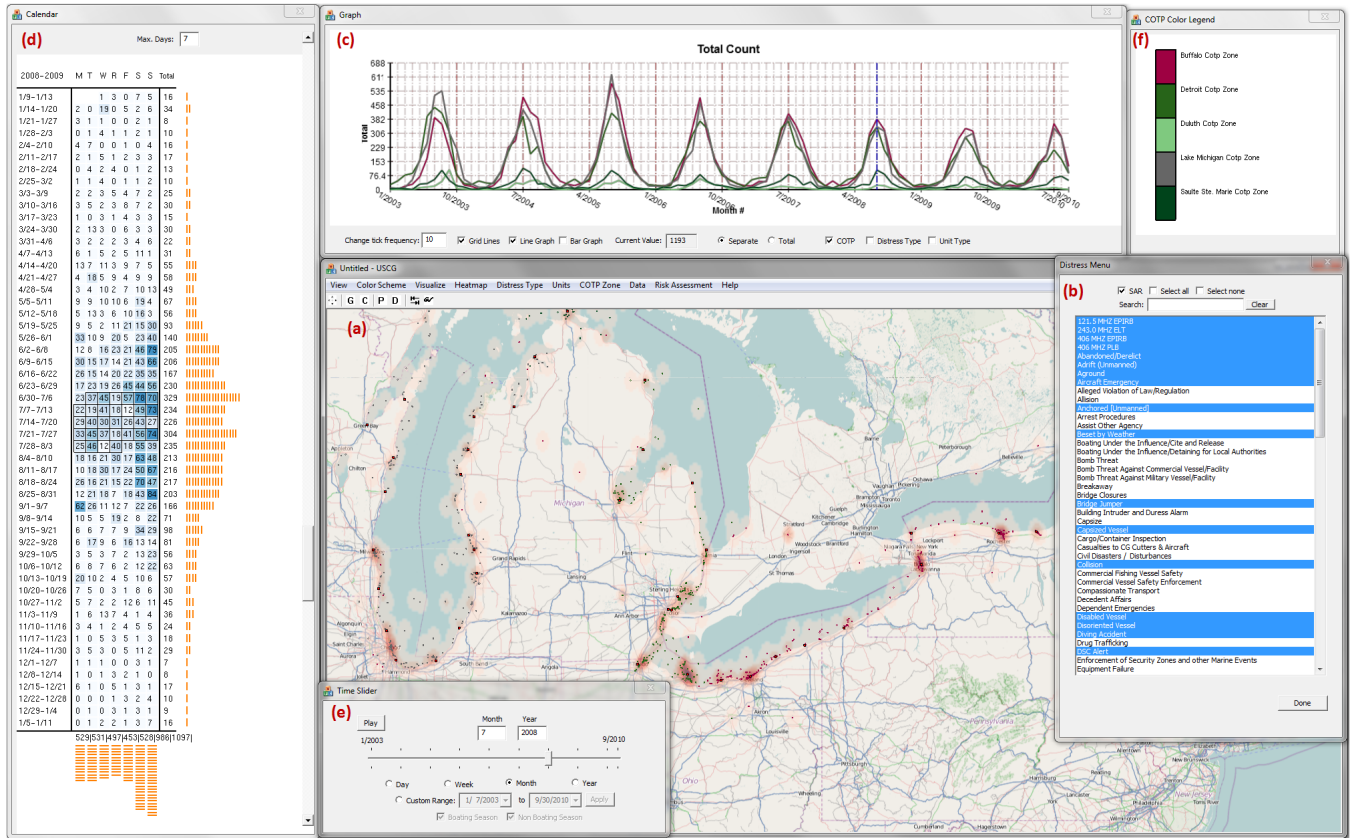
Figure 1: A screenshot of our risk assessment visual analytics system. Here, the user is visualizing all search and rescue (SAR) operations conducted by the U.S. Coast Guard in the Great Lakes region in July 2008. The main viewing area (a) shows the map view with the circles showing the locations of SAR incidents in the Great Lakes. The right-most window (b) shows an interactive menu showing all distress types with SAR cases selected in blue. The top window (c) shows the time-series and the left window (d) shows the calendar views of the SAR incident report data. The bottom-left window (e) shows the time slider with radio buttons that allow different temporal aggregation levels. A legend for all District Nine maritime zones is shown in the upper right (f).

ios associated with closing Coast Guard stations. We emphasize that although our risk assessment toolkit and the examples given in this paper have been based in the maritime domain, these techniques apply equally as well to other domains (e.g., criminal offense analysis, syndromic surveillance).

2 RELATED WORK

In recent years, there has been a rapid growth in the development of new visual analytics tools and techniques for advanced data analysis and exploration (e.g., [30, 34]). From traditional scatterplots [8] and parallel coordinate plots [16] to tools like Theme River [15] and spiral graphs [7], these systems incorporate different forms of visualizations to provide enhanced analytical tools to users. Although these tools allow users to explore their data and assist them in their decision making process, researchers have only recently started to employ visual analytics techniques for risk-assessment and decision-making domain that allow users to perform a thorough analysis of risks associated with different decisions.

Migut and Worring [21] propose an interactive approach to risk assessment where they demonstrate a risk assessment framework that integrates interactive visual exploration with machine learning techniques to support the risk assessment and decision making process. They use a series of 2D visualizations including scatterplots and mosaic plots to visualize numerical and ordinal attributes of the datasets. While the authors demonstrate the effectiveness of using

visual analytics in the field of risk assessment, their work is mainly focused on building classification models that users may interactively use to classify their data entities and visualize the effects on their classification. Gandhi and Lee [13] also apply visual analytics techniques to the realm of requirements-driven risk assessment. Specifically, they use cohesive bar and arc graphs to illustrate the risks due to the cascading effects of non-compliance with Certification and Accreditation requirements for the U.S. Department of Defense. Sanusi and Mustafa [25] introduce a framework to develop a visualization tool that may be used for risk assessment in the software development domain. Their proposed framework allow users to identify the components of the software system that are likely to have a high fault rate. Direct visualizations of risk use tools like bar graphs and confidence interval charts to visualize measures of risk and are usually constructed using spreadsheet programs like Microsoft Excel [12, 13]. Although widely used, these techniques fail to work for our purposes primarily due to the nature of the risk analysis that is required. The Coast Guard SAR dataset is spatiotemporal in nature and the exploration of risk requires domain knowledge that is difficult to incorporate algorithmically.

With respect to the temporal nature of risk assessment, researchers have also developed different visualization systems that allow users to explore risks associated with financial decisions related with investments and mutual funds, among other financial planning scenarios. Rudolph et al. [24] propose a personal finance

decision making visual analytics tool that allows users to analyze both short-term and long-term risks associated with making investment decisions. Savikhin et al. [26] also demonstrate the benefits of applying visual analytics techniques to aid users in their economic decision making and, by extension, to general decision making tasks. Both of the previous examples only explore temporal datasets. In this work, we apply visual analytics techniques to explore risks using multivariate spatio-temporal datasets that guide analysts in making complex decisions.

As is the case with most multivariate datasets, data tends to be inherently unreliable, incomplete and contradictory. In order to reach to correct conclusions, analysts must take these into account in their analysis. In this regard, Correa et. al. [9] describe a framework that supports uncertainty and reliability issues in the different stages of the visual analytics process. They argue that with an explicit representation of uncertainty, analysts can make informed decisions based on the levels of confidence of the data. Our system factors data reliability issues in the risk assessment process and provides confidence levels at all stages of risk assessment that, in turn, enable analysts to better understand the underlying nature of the data and guides them in making effective decisions.

There also exist many geospatial and temporal analytical systems that provide users with the ability to explore their spatiotemporal datasets in order to find patterns and provide an overview of the data in a visual analytics platform (e.g., [1, 2, 3, 14]). As the needs of our end users are unique, this warrants developing a stand alone system to address the challenges faced by the Coast Guard analysts. We plan on further examining these robust geo-temporal analysis tools and the degree to which they can be extended to meet the Coast Guard requirements that have been identified in this paper.

There has also been much work done in visualizing large datasets using interactive cross-filtered and linked views that allow users to explore their datasets. Stasko et. al. [30] use multiple coordinated views of documents to reveal connections between entities across different documents. Eick and Johnson [10] utilize multiple linked views to visualize abstract, non-geometric datasets in order to reduce visual clutter and provide users with insights into their datasets. Eick and Wills [11] also demonstrate the effectiveness of linking and interaction techniques in the visualization of large networks. Our system utilizes these practices and allows users to interactively explore their multi-dimensional and multi-attribute datasets using a series of multi-coordinated linked views.

Researchers have also explored different methods to address the challenges posed to maritime security and safety. Willems et. al. [34] introduce a novel geographic visualization that supports coastal surveillance systems and decision making analysts in gaining insights into vessel movements. They utilize density estimated heatmaps to reveal finer details and anomalies in vessel movements. Scheepens et. al. [27] also present methods to explore multivariate trajectories with density maps and allow the exploration of anomalously behaving vessels. Lane et. al. [18] present techniques that allow analysts to discover potential risks and threats to maritime safety by analyzing the behavior of vessel movements and determining the probability that they are anomalous. Some other models for anomaly detection in sea traffic can be found in [19, 22]. Researchers have also proposed several approaches to maritime domain awareness. For example, Roy and Davenport [23] present a knowledge based categorization of maritime anomalies built on a taxonomy of maritime situational facts involved in maritime anomaly detection. We observe that these methods and models may help in risk analysis and understanding the impact of weather and varying speeds of Coast Guard vessels in the Great Lakes to identify high risk regions.

There has also been much work done to assess and mitigate risks to critical infrastructure and transportation in the maritime domain. Adler and Fuller [5] provide dynamic scenario- and simulation-based risk management models to assess risks to critical maritime infrastructure and strategies implemented for mitigating these risks. Mansouri et. al. [20] also propose a risk management-based decision analysis framework that enables decision makers to identify, analyze, and prioritize risks involved in maritime infrastructure and transportation systems. Their framework is based on risk analysis and management methodologies that allows understanding uncertainty and enables analysts to devise strategies to identify the vulnerabilities of the system. Furthermore, work has been done to quantify risks in the maritime transportation domain, a summary of which can be found in [29]. While these methods facilitate maritime infrastructure risk analysis, our work is focused on assessing maritime risks from multivariate spatiotemporal SAR data sets. In this paper, we present a visual analytics approach to maritime risk assessment and provide examples that demonstrate the advantages of applying visual analytics in this domain.

3 VISUAL ANALYTICS RISK ASSESSMENT ENVIRONMENT

Our visual analytics system provides enhanced risk assessment and analytical tools to analysts and has been built to operate for SAR incident report data. Our system has been implemented in a custom Windows-based geographical information system that allows drawing on an OpenStreetMap map [4], using Visual C++, MySQL and OpenGL. The system displays geo-referenced data on a map and allows users to temporally scroll through their data. We provide linked windows that facilitate user interaction between the spatial and temporal domains of the data. We also provide advanced filtering techniques that allow users to interactively explore through data. In addition, we have adapted the calendar view presented by vanWijk and Selow [33] and extended it to explore seasonal and cyclical trends of SAR operations and also as means to filter data to support advanced analysis.

Figure 1 presents a screenshot of our system. The main viewing window (Figure 1 (a)) shows the map view where the user can explore the spatial distribution of all cases handled by the Coast Guard. We utilize density estimated heatmaps (Section 3.2) to quickly identify hotspots. Users may draw a bounding box over incident points on the map that generates a summary of all incidents enclosed by the box. We also provide tape measure tools that allow users to measure the distance between two points on a map. The top-most window (Figure 1 (c)) shows the time-series view of the data where multiple lines graphs can be overlaid for comparison and analysis. Users may visualize time-series plots by department, distress type and Coast Guard Captain of the Port (COTP) zone to explore summer cyclical patterns. The left-most window (Figure 1 (d)) shows the calendar view of the selected Coast Guard cases. The total number of columns on the calendar may be changed as desired to reveal seasonal trends and patterns. The bottom window (Figure 1 (e)) shows the time-slider widget that is used to temporally scroll through the data while dynamically updating all other linked windows. The radio buttons beneath the time slider provide several temporal aggregation methods for the data. The right-most window (Figure 1 (b)) shows the distress type menu where all SAR cases (highlighted in blue) have been selected for visualization. Users may select multiple distress types using this menu, dynamically updating all linked views. We use similar menus to filter cases by other data fields. Users may also interactively search the menu using the search box provided on top of the menu. Finally, the top-right window (Figure 1 (f)) shows an interactive legend of the different Coast Guard District Nine maritime zones. This legend allows users to click on any of the zones that highlights all cases falling in the zone by filling the circles on the map with a solid color and dimming out the other cases being displayed on the map.

A key feature of our system is the interactive distress, station and COTP zone filtering component. Users interactively generate combinations of filters that are applied to the data being visualized

through the use of menus (like the one shown in Figure 1 (b)) and edit controls. The choices of filters applied affects both the geospatial viewing region and all temporal plots.

3.1 Coast Guard SAR data

The SAR data is collected by all U.S. Coast Guard stations and stored in a central repository. When the Coast Guard is called into action, a response case is generated, usually by the maritime zone that has authority in that region that receives the distress call (referred to by the Coast Guard as the Search Mission Coordinator or SMC). Upon receiving the call, this authority will determine if resources will be applied, including which unit will provide the resource, the resource type and number. Therefore, a response case may generate zero, one, or many sorties to respond to an incident. While analyzing risks associated with the various mitigation options, including station closure, analysts are interested in analyzing the spatiotemporal distribution of both the response cases and their associated sorties.

The SAR data consists of two main components: (1) response cases and (2) response sorties. Each entry in the response case and sortie dataset contains information that provides details of the incidents (e.g., number of lives saved, lost, assisted) and contains the geographic location of the distress.

Uncertainty in decision making

As is the case with most large datasets, anomalies and missing data introduce errors and uncertainty. The SAR data is no exception. We find that many SAR cases do not have an associated geographic location, or have a wrong geographic location associated with them. These inherent errors in data affect the spatial probability estimates and introduce a certain amount of uncertainty in the decisions that must be considered for an effective risk analysis and assessment. As noted in [17], visual analytics methods help people make informed decisions only if they are made aware of data quality problems. In this regard, we incorporate uncertainty and confidence levels associated with the SAR dataset in our visualizations by displaying the accuracy of the results at each step of the risk assessment process. This is shown as a percentage that shows the total cases with reliable data that can be used in the decision making process (Figure 2). This percentage is calculated by using the following formula:

$$Accuracy = \frac{N - G}{N} \times 100 \tag{1}$$

Here, N is the total number of cases and G is the number of cases with unreliable values (e.g., unknown geographic coordinates, swapped negative signs). When such errors are not obvious, the data is assumed to be correct and is displayed to the analyst on the map. The analyst can further report errors in the data and contribute to the data cleaning process.

3.2 Geospatial displays

Our system provides analysts with the ability to plot incidents as points on the map and as density estimated heatmaps (Figure 1 (a)). In addition, we provide users with the option of coloring each incident circle with a color on a sequential color scale [6] that represents its data value. For example, users may choose to visualize the average response time to respond to an incident for all SAR cases on the map and identify cases with higher response times. Furthermore, to explore the spatial distribution of the SAR cases and quickly identify hotspots, we employ a modified variable kernel density estimation technique (Equation 2) that scales the parameter of estimation by allowing the kernel scale to vary based upon the distance from the point X_i to the kth nearest neighbor x in the set comprising of N [28].

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{min_{(h,d_{i,k})}} K\left(\frac{x - X_i}{min_{(h,d_{i,k})}}\right) \tag{2}$$

Here, N is the total number of samples, $d_{i,k}$ is the distance from the i-th sample to the k-th nearest neighbor and h is the maximum allowed kernel width. We choose the maximum kernel width based on asset speed and travel time. Furthermore, we use the Epanechnikov kernel [28] (equation 3) to reduce calculation time:

$$K(\mathbf{u}) = \frac{3}{4}(1 - \mathbf{u}^2)1_{(||u|| \leq 1)} \tag{3}$$

where the function $1_{(||u|| \leq 1)}$ evaluates to 1 if the inequality is true and to zero otherwise.

3.3 Time series displays

Along with the graphical interface, our system provides a variety of visualization features for both spatial and temporal views. For temporal views, we provide line and stacked bar graphs and calendar views to visualize time series SAR incident report data.

The line graph visualization allow users to overlay multiple graphs for easy comparison and to visualize trends. Both line graph and stacked bar graph visualizations are supported and can be interchanged using the radio buttons provided. Users may choose to visualize SAR cases handled by individual stations or maritime zones, or visualize them by distress types. The data is plotted based on a temporal aggregation level that the user selects on the time-slider widget (Figure 1 (e)). In Figure 1 (c), we show the line graph display of all SAR cases aggregated by month. We can easily observe peaks in the number of SAR cases in the summer months for all maritime zones in the Great Lakes region.

The calendar view visualization was first developed by van Wijk and Selow [33]. This visualization provides a means to allow the visualization of data over time, laid in the format of a calendar. In our implementation (Figure 1 (d)), we shade each date entry based on the overall yearly trend. Users may interactively change the total number of columns of the calendar thereby changing the cycle length of the calendar view, enabling users to explore both seasonal and cyclical trends of their datasets. The system also draws histograms for each row and column. This allows analysts to visualize weekday and weekly trends of SAR incidents and further assists them in determining an effective resource allocation scheme. Furthermore, we have modified our calendar view to support an interactive database querying method for easily acquiring summary statistics from the SAR database.

4 RISK ASSESSMENT PROCESS

In this section, we describe the different methods and techniques that we apply in the Coast Guard risk assessment process.

4.1 Problem description

To bound the problem, the Coast Guard analysts provided a series of questions for use in their analysis. These questions are briefly summarized below.

1. Assuming a maximum transit speed of 15 nautical miles per hour, how many cases occur per year in which a parent station could not have a surface asset on scene within two hours?

2. For each Auxiliary station, what are the types (by percentage) of SAR response cases occurring per year?

3. For each Auxiliary station, what is the temporal (by hour, month and day of week) distribution of the response case load?
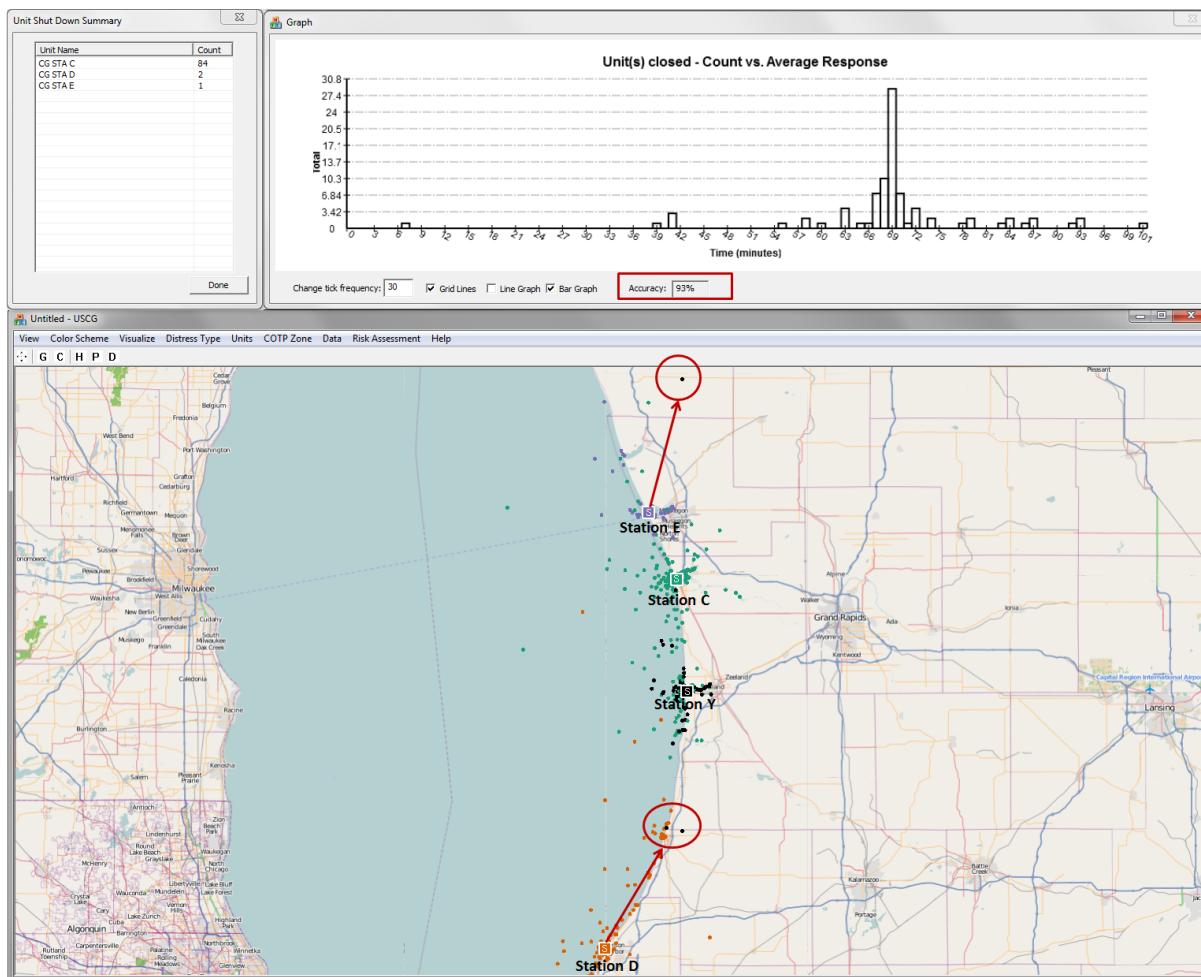
Figure 2: Average response time risk assessment when Auxiliary station Y is closed. The system automatically chooses the stations (shown in the upper-left window) that are optimally suited to respond to cases previously handled by station Y, along with a count of cases that each station absorbs. The station Y cases (black circles) to be handled by stations D and E are circled in red. The top graph shows the average response time distribution of these stations to respond to station Y's cases. We find that only 93% of the cases responded to by station Y have an associated geographic location.

4. What is the average annual case load that would be absorbed by each parent station in the absence of the Auxiliary station and what percentage increase would this represent to the parent station's annual case load?

5. Based on the historical data for all cases (SAR and others), what is the expected annual response case demand broken down by response type (i.e., Person in Water, Vessel Flooding, etc.)?

6. Assess the potential risks associated with closing certain Auxiliary stations in terms of additional case load absorbed, lives potentially lost, and other available factors.

Our visual analytics system was developed to assist the Coast Guard analysts in answering these questions and to model the potential risks of closing one or more Auxiliary stations. Furthermore, we allow analysts to explore the effects of closing multiple stations and provide a summary of stations that are most optimal to absorb the work load of the closed stations. Analysts may restrict the stations that absorb the work load of the closed stations to determine the stations that prove most effective, thereby informing optimal operational execution for the station that is nearest to respond to the distress case.

We perform our analysis under the assumption that the path between a station and a distress location is a straight line. While this assumption presents a best-case scenario to the analyst, discussions with our Coast Guard partners indicated this was an acceptable approximation as using channel and waterway information would result in a large computational overhead. With this assumption in place, if a station absorbing an Auxiliary station's cases increases the maritime risks in the region (e.g., if the average response time exceeds the two hour time limit for most SAR incidents), then closing the Auxiliary station could prove to be dangerous for the maritime and public safety of the region. This straight line approximation provides details on the best case scenario.

4.2 Average response time for SAR incidents

As stated before, a Coast Guard policy mandates the rescue resource to be on scene within two hours of a distress (e.g., disabled vessel, person in water). Given the cold water temperatures in the Great Lakes, even in the summer, increase in response time can potentially impact the success of a case. Therefore, given the option of closing a station, the analysts desire to know the nearest available resource to respond and calculate the time to respond to the scene. A typical Coast Guard vessel travels at a speed of 15 nautical miles
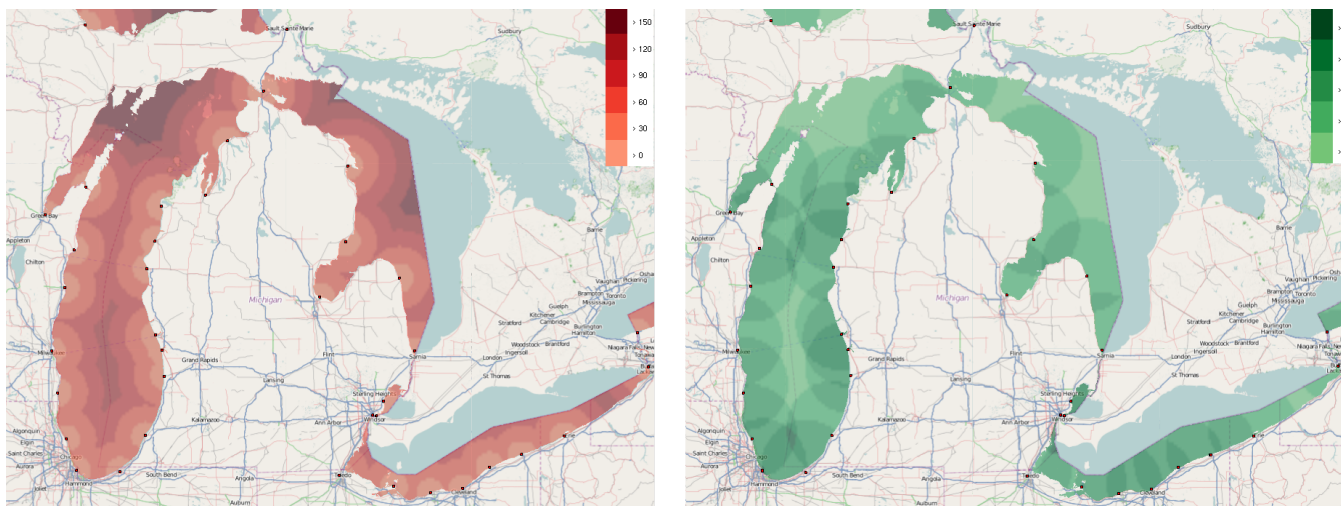
Figure 3: Risk profile. (Left) A heatmap showing the time taken (in minutes) by the Coast Guard stations to deploy an asset to the Great Lakes to respond to a SAR incident, assuming a speed of 15 nautical miles per hour. (Right) A heatmap showing the Coast Guard SAR coverage (i.e., the number of stations that respond to a particular region) in the Great Lakes. The squares along the coast show the locations of the stations.

per hour. After an Auxiliary station is closed, the parent station should still be able to reach most of the cases handled by the Auxiliary station within the two hour limit. In this section, we describe how our system can be used to determine the average response time for cases if a parent station (or any combination of stations) absorbs an Auxiliary station's cases.

In order to generate the average response time for the station(s) that absorb the work load of the closed station, we sift through all the incidents that the closed station handled and find the closest station (excluding the closed station) for each incident by comparing the distance between all stations and the incident. This distance between the closest station and incidents may also be visualized separately to reveal more details. Once the closest station is found, we obtain the time for an asset to reach the incident location using the distance formula $Time = Distance/Speed$. Users may also change the speed of the asset, changing the results dynamically.

We provide users with several filtering options while performing average response-time analysis. Users may choose to analyze the average response-time temporal distribution of incidents by applying any possible filters on distress type, department or maritime zone. Users may also analyze the distribution of only the non-SAR cases. Moreover, users may choose to close several stations all at once and model the resulting effects. They may also specify which stations absorb the cases of the closed stations and thus determine the stations best suited for closing and the optimal methods for reallocating available resources. We also note that our system can be easily modified to incorporate other risk metrics including, for example, normalizing SAR cases by the underlying population density, correlating SAR incidents with other parameters, etc.

Figure 2 shows the output generated when the analyst opts to close Auxiliary station Y. In this example, we examine all cases responded to by station Y between January 2004 and September 2010. The system automatically suggests the stations that should absorb Auxiliary station Y's cases along with the total number of cases that each station absorbs. We find that stations C (the parent station of Y), D and E absorb Auxiliary station Y's cases, with each absorbing 84, 2 and 1 cases, respectively. The analyst may instead select a specific station to absorb station Y's cases and analyze the results generated. In Figure 2, the map view shows all cases that each of the four stations responds to during this time period (shown as circles, with each case color coded by its station). We have also

highlighted the two cases that station D and the one case that station E responds to in Figure 2. It may be noted that the one case absorbed by station E appears to be out of place (possibly due to a human error in entering the geographic coordinates for that particular case). The top-right bar graph shows the count of all SAR cases handled by station Y during this time period versus the average response time (in minutes) taken by the resulting stations to reach these cases, assuming a transit speed of 15 nautical miles per hour. From this time-series plot, we observe that all cases responded to by the Auxiliary station would fall well within the tolerance level of 120 minutes when the suggested stations take over. The system also determines the accuracy of the results dynamically by determining the number of cases that have no associated geographic coordinates. We find that 93% of the cases responded to by station Y in the time range January 2004 and September 2010 have an associated geographic coordinate (as seen from the accuracy percentage in Figure 2-top-right). Data integrity is a necessary parameter to report to the analysts and decision makers. Thus, the user is made aware of these uncertainties at every step of the risk assessment process.

4.3 Temporal distribution of response case load

One important aspect of risk assessment is analyzing the work load and distribution of response cases of the stations being analyzed over different temporal ranges. This becomes necessary to determine the feasibility of a station to be closed and to determine how the available resources may be reallocated (e.g., what times of day and what months would the stations need to have more personnel deployed). Analysts also use their domain experience and expertise to determine whether a particular station can absorb a closing station's cases. In particular, the Coast Guard officials were interested in understanding the hourly, daily and monthly trends of SAR cases occurring in the Great Lakes.

Using traditional methods of sifting through SAR datasets turns out to be highly inefficient for determining the temporal distribution of the SAR cases and, as such, advanced database querying tools are necessary to facilitate this process. To this end, we adapt the calendar view for querying the SAR database. We provide three different interaction methods within the calendar view widget (Figure 1 (d)) to obtain a detailed summary of response cases occurring over the selected date-range. Users can select date ranges by simply clicking on the start and end dates that selects all the dates between the two clicked dates. Users may also select one or more columns
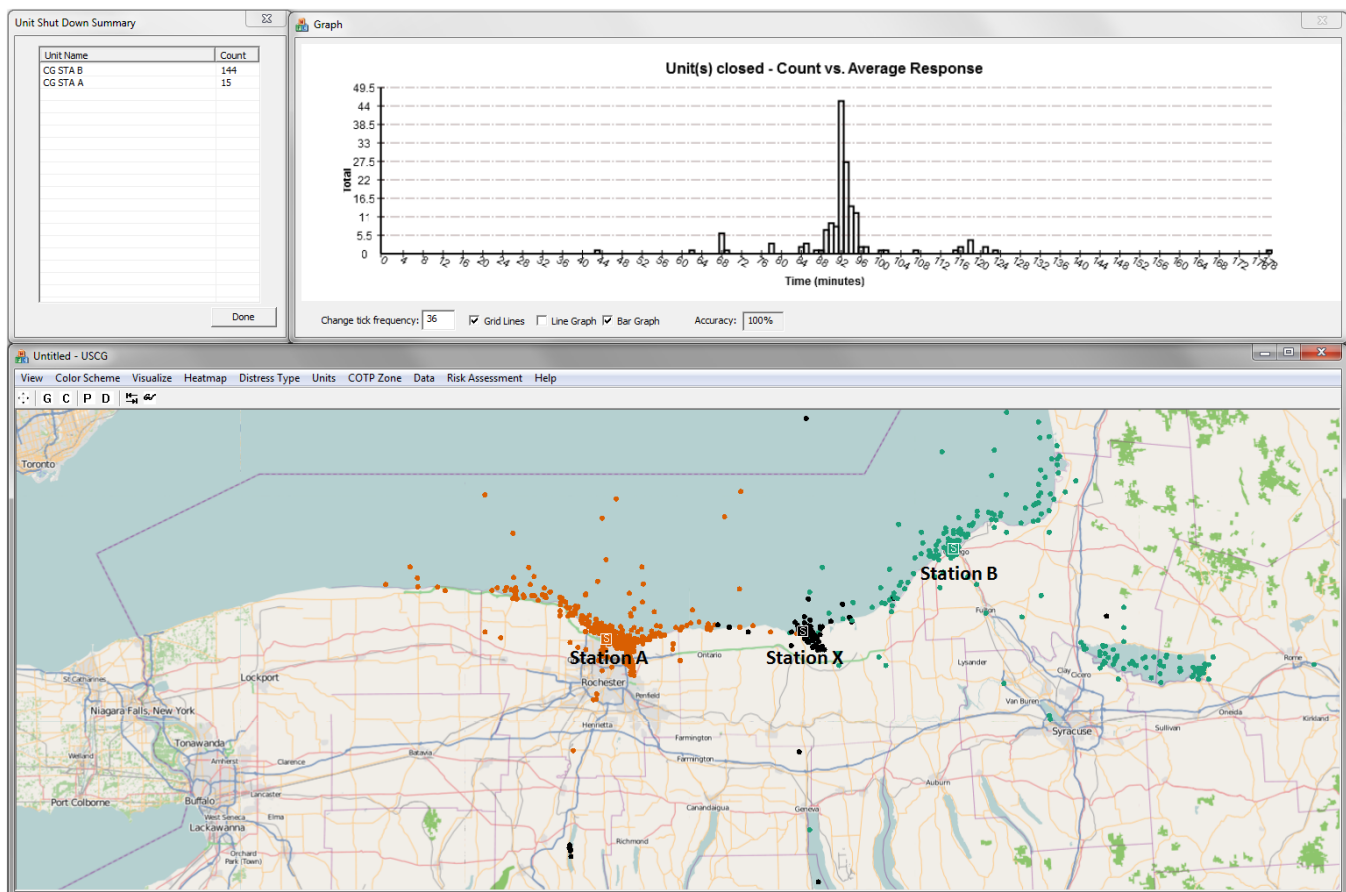
Figure 4: Risk assessment using linked views. Here, the analyst has chosen to close Auxiliary station X and is analyzing the risks associated with this decision. Each circle on the map represents a SAR case and has been color coded by its owner station. The system shows that stations A and B are best suited to absorb station X's cases, with station B absorbing 144 cases and station A absorbing 15 cases.

of the calendar to generate the summary statistics. This allows them to query the database and acquire the summary of events occurring, for example, only on a particular week day. Finally, users may select any combination of individual dates and obtain the summary of all selected response cases on those dates. These querying methods allow analysts to easily determine the temporal patterns of response cases over any date range. The system provides summary statistics of SAR incidents for all stations and includes the total number of lives saved, assisted, affected, total property damaged and saved and the count of all cases occurring over the selected date range. Users may select any date, row, column, or combinations thereof in the calendar view using the mouse to access the summary statistics. Furthermore, the system also allows users to visualize the hourly and monthly distribution of cases for any time period after all filters are applied.

4.4 Risk profile

Our system also provides users with the ability to interactively generate risk profiles that can be used to identify regions with little SAR coverage by the Coast Guard stations in the Great Lakes. Figure 3 illustrates the risk profile heatmaps that present an overview of the Coast Guard SAR coverage in the Great Lakes. Selected filter settings affect the visual output, and in this case, we are looking exclusively at small boat station coverage. When areas of low coverage exist, resources with additional capability (e.g., aircraft) are often provided to ensure coverage of all areas. Figure 3 (Left) shows the time (in minutes) that the Coast Guard stations would

take to respond to a SAR incident in the Great Lakes, assuming a transit speed of 15 nautical miles per hour. This profile is generated assuming that the station closest to a location responds to an incident in the Great Lakes. The regions in the Great Lakes that take the longest time for the Coast Guard to respond to a SAR case can be clearly seen in this figure. Users may interactively close stations, filter on a different resource type (e.g., boat, aircraft), or change the asset speed, updating the risk profile interactively. This further enables the analysts to visualize the increase or decrease in risk when a station is closed. Moreover, analysts can set the lower threshold of the color scale to 120 minutes (or any arbitrary time), thereby allowing them to easily identify regions that may take more than 120 minutes to respond. We plan on incorporating contour lines into our system to demarcate the regions that may take more than the set threshold response time.

Figure 3 (Right) provides another risk profile visualization that allows officials to identify regions with low Coast Guard coverage for SAR operations in the Great Lakes. Regions with high SAR coverage by the Coast Guard stations are shown by darker colors. This further allows analysts to identify stations where resources may be reallocated without increasing maritime risk.

5 EXPLORING RISK USING SPATIOTEMPORAL LINKED VIEWS

While examining which Auxiliary stations are most suitable to close, analysts need to weigh all options and analyze the potential increase or decrease in associated risks. They must also consider
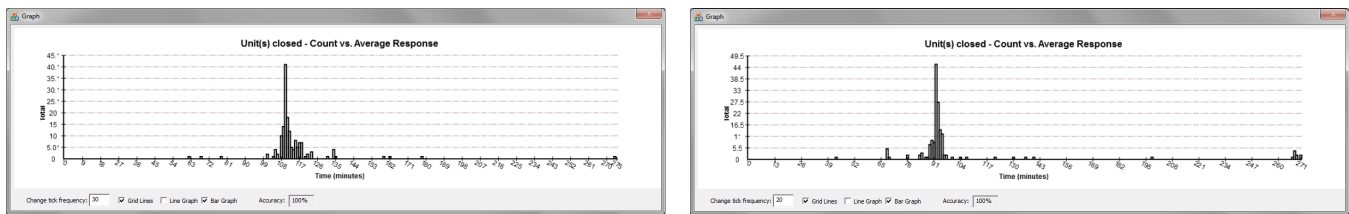
Figure 5: The average response time distribution for the additional cases when parent station A absorbs station X (Left) and when station B absorbs station X (Right). We see a left shift in the median time indicating that station B may be a better candidate to absorb station X's cases.
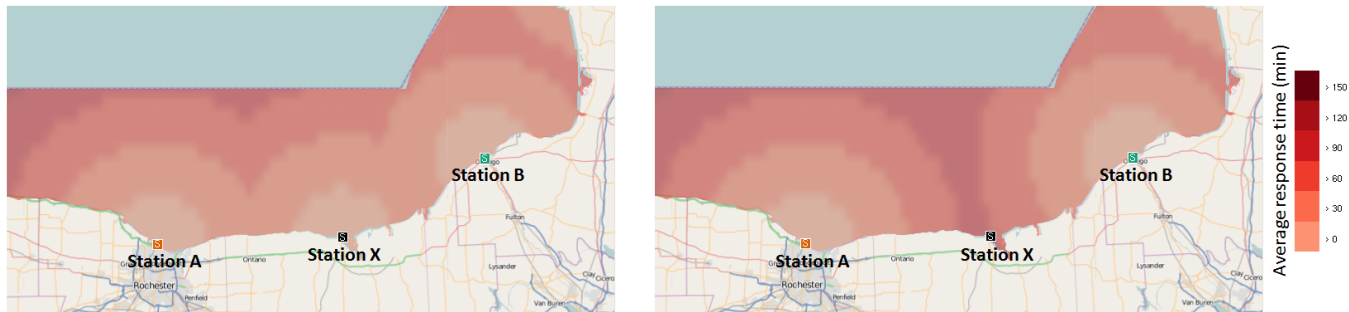
Figure 6: The change in average response time risk profile when Auxiliary station X is closed. (Left) Risk profile when station X is operational. (Right) Risk profile when station X is closed. We see an increase in risk in station X's area of operation when the station is closed.

the increase in workload of the stations that absorb the closed station's cases to effectively determine the optimal response of available resources. In this section, we describe a typical scenario where an analyst is trying to determine the risks associated with closing an Auxiliary station in one of the sectors in the Great Lakes and the stations that absorb the work load of this closed station.

Suppose the analyst chooses to close Auxiliary station X whose parent station is A. Since the parent station of X is station A, the analysts' first inclination may be to assign all cases to station A after station X is closed. The analyst first uses the system's calendar view and finds that the maximum number of cases that station X responds to in one day is 7 in the peak boating season. He also visualizes the hourly distribution of cases for station X and determines that the incidents are spread out during the day. Next, the analyst uses our risk assessment system to perform an average response time risk analysis over a time range of January 2006 to September 2010 and selects Auxiliary station X to be closed. Once station X is closed, the system automatically generates the result seen in Figure 4. As seen in the figure, the system determines stations A and B to be the optimal stations to respond to the cases handled by Auxiliary station X. In this figure, we see the spatial distribution of all SAR cases that the three stations responded to during this time range (seen as circles that are color coded by station). We observe that station B absorbs 144 additional SAR cases as opposed to parent station A which only absorbs 15 cases. Moreover, with this case load distribution, we find that 154 out of the total 159 cases are responded to within the 120 minute limit (as seen from the time series plot in Figure 4-top-right). These results suggest that station B is better suited to absorb most of the cases of Auxiliary station X.

In order to get a better picture, the analyst now restricts the stations that respond to the cases handled by station X to first, its parent station A, and then to station B and analyzes the average response time distribution of cases for each of the two stations separately. The results of this step are shown in Figure 5, with the left graph showing the average response time taken by station A, and the right graph corresponding to station B. As the analyst compares the two graphs, he realizes that if only station B is allowed to absorb station X, there are 14 cases that take more than 2 hours

for the Coast Guard to arrive on scene. On the other hand, if station A is allowed to absorb station X, 16 cases take longer than 2 hours. However, as we can clearly see from Figure 5, station A takes a longer time to respond to most cases than station B, with the median time of station A being 110 minutes, and that of station B being 92 minutes. The analyst also notes that station B takes between 264-271 minutes to respond to about nine cases, whereas station A takes 275 minutes to respond to one case. To get a better understanding of why this may be happening, the analyst explores the spatial distribution of station X's cases on the map and discovers that some cases of station X get mapped to an inland lake (which requires trailering the boat to the scene). In order to confirm that these cases do not occur as a result of errors in the database, he draws a bounding box over these incidents on the map and obtains a summary of these incidents. The summary confirms that these incidents do indeed occur in that particular lake. As station A is closer to these cases, the analyst concludes that station A would be a better candidate to absorb these cases as opposed to station B. But for the rest of the cases, the results clearly suggest that station B would be a better candidate to absorb station X's cases, and that station A would increase the maritime risks if allowed to absorb station X's cases alone. Thus, this analysis confirms that a combination of these two stations yields the best results. With these results at hand, the analyst may also recommend using an aircraft to respond to cases that take more than 120 minutes. Or, as our preceding analysis showed, stations A and B may absorb the work load of station X together, with station B receiving a higher share of resources than station A. We also note that in the future, our system could be modified to perform a real time analysis of SAR cases and could then be used to assign each case to the correct station in real time.

The analyst now uses the risk profile tools to observe the increase in risks when station X is closed. This is shown in Figure 6, with the left figure showing the average response time risk profile when Auxiliary station X is functional whereas the right figure shows the risk profile when station X is closed. The analyst explores the new average response time on the map when station X is closed and determines the potential increase in maritime risks in the region. The
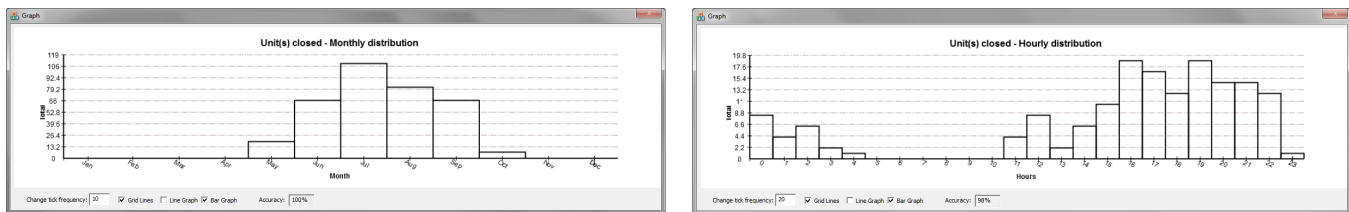
Figure 7: The monthly (left) and hourly (right) distributions of SAR incidents handled by Auxiliary station X between Jan. 2006 and Sept. 2010.

analyst thus identifies the regions in station X's area of operation that take a greater time to respond.

The analyst also visualizes the monthly and hourly distributions of all SAR cases responded to by the closed Auxiliary station X between January 2006 and September 2010 (shown in Figure 7, where the left graph shows the monthly distribution and the right graph shows the hourly distribution). We see that all activity occurs in the summer months with a peak occurring in July and the station responds to most cases mainly during evening hours. The analyst also visualizes the temporal activity of stations A and B, and determines the potential work load increase for both stations. This helps the analyst determine how the available resources must be reallocated if Auxiliary station X is to be closed. Furthermore, the analyst chooses to visualize the case distribution of Auxiliary station X between January 2006 and September 2010 using the interactive calendar view widget. This generates a summary dialog box that provides the details of the SAR cases responded to by station X and includes details including the total number of lives assisted, saved and lost. This further helps the analyst understand the risks associated with closing this station by providing an overview of the cases occurring in this region. With all these results at hand, the analyst uses his domain knowledge to make an informed decision.

6 DOMAIN EXPERT FEEDBACK

Our system was assessed by an analyst at the U.S. Coast Guard's Ninth District who is currently using the system to determine the potential risks in the maritime domain associated with the hypothetical allocation of Coast Guard resources. The analyst emphasized the need of such systems in the maritime domain that allow users to quickly and easily process large datasets in order to derive actionable results. The analyst noted that processing the desired queries took him a fraction of the time when using our system as compared to using other software (e.g., Microsoft Excel) that he had been previously using in his analysis. He was impressed by the fact that the system is intuitive to use and requires little user training. He observed that the system's ability to process large datasets allows him to quickly filter the data into manageable subsets while providing interactive spatiotemporal displays that further aid him (and ultimately the senior level decision makers) in making a decision using the best information available.

7 CONCLUSIONS AND FUTURE WORK

Our current work demonstrates the benefits of visual analytics in analyzing risk and historic resource allocation in the maritime domain. Our visual analytics system provides analysts with a suite of tools for analyzing risks and consequences of taking major decisions that translate into important measures including potential lives and property lost. Our results show how our system can be used as an effective risk assessment tool when examining various mitigation strategies to a known or emergent problem.

Before this system was developed, Coast Guard officials explored possible mitigation strategies, including the implementation of seasonal or weekend only Auxiliary duty stations, but the sheer volume of data and information inhibited the efficient processing of the data. However, using our system, the decision makers were quickly made aware that most response cases happened on Mondays/Tuesdays at some of the units. This further asserts the benefits of the use of visual analytics in the maritime domain.

In addition to performing risk analysis on the Coast Guard SAR cases, our system can also be used to conduct a thorough review of the operations (i.e. non-distress cases) conducted by different Coast Guard stations. Users may choose to visualize different datasets and analyze how each station performs in terms of factors including average response times, average distance to target, lives saved, lives assisted, lives affected, etc. Hence, the officials may analyze the efficiency of each Coast Guard station and identify problem areas that may require further attention.

Future work includes deploying our system to assist in the analysis and optimization of all operations conducted by the U.S. Coast Guard Ninth District and expanding the use of our system to other Coast Guard districts. We plan on implementing algorithms that factor the geography of the coast line in the risk assessment process in order to get accurate response times by the Coast Guard assets. We also plan on employing prediction algorithms in the temporal domain as well as spatiotemporal correlation algorithms that correlate different datasets (e.g., weather, water temperature) with the response dataset to provide insights into the operation of the Coast Guard stations. Furthermore, we plan on incorporating additional risk metrics to provide insights into different risk scenarios.

8 APPENDIX

In this section, we briefly provide some domain specific terms and definitions:

Coast Guard Auxiliary: Volunteers that support the Coast Guard.

Coast Guard Ninth District: The area of Coast Guard operations that encompasses the Great Lakes.

Atlantic Area Command: The area of Coast Guard operations East of the Rocky Mountains.

Captain of the Port (COTP) Zone: Further division of Coast Guard operations within a Coast Guard District.

Unit or Station: The operational execution arm of the Coast Guard. For example, the small boat station provides the boat and personnel to execute the assigned mission.

Coast Guard asset: A boat or an aircraft reserved to perform Coast Guard operations.

Coast Guard sortie: An asset that responds to an incident.

9 ACKNOWLEDGMENTS

10 DISCLAIMER

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security or the U.S. Coast Guard.

REFERENCES

[1] Command Post Of The Future (CPOF). Internet: `http://www.gdc4s.com/documents/cpof_datasheet_web.pdf`, [June 20, 2011].

[2] ESRI, ArcView. Internet: `http://www.esri.com/software/arcgis/arcview/index.html`, [June 9, 2011].

[3] Oculus, GeoTime. Internet: `http://www.oculusinfo.com/SoftwareProducts/GeoTime.html`, [June 25, 2011].

[4] OpenStreetMap. Internet: `http://www.openstreetmap.org`, [June 15, 2011].

[5] R. Adler and J. Fuller. An integrated framework for assessing and mitigating risks to maritime critical infrastructure. In *IEEE Conference on Technologies for Homeland Security*, pages 252 –257, May 2007.

[6] C. A. Brewer. *Designing Better Maps: A Guide for GIS users*. ESRI Press, 2005.

[7] J. V. Carlis and J. A. Konstan. Interactive visualization of serial periodic data. In *Proceedings of the Symposium on User Interface Software and Technology*, pages 29–38, 1998.

[8] W. S. Cleveland and M. E. McGill, editors. *Dynamic Graphics for Statistics*. Wadsworth and Brooks/Cole, 1988.

[9] C. Correa, Y.-H. Chan, and K.-L. Ma. A framework for uncertainty-aware visual analytics. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 51 –58, Oct. 2009.

[10] S. G. Eick and B. S. Johnson. Interactive data visualization at AT&T Bell laboratories. In *Conference companion on human factors in computing systems*, pages 17–18, New York, NY, USA, 1995.

[11] S. G. Eick and G. J. Wills. Navigating large networks with hierarchies. In *Proceedings of the 4th conference on visualization*, Visualization 1993, pages 204–209, Washington, DC, USA, 1993. IEEE Computer Society.

[12] M. Feather, S. Cornford, J. Kiper, and T. Menzies. Experiences using visualization techniques to present requirements, risks to them, and options for risk mitigation. In *First International Workshop on Requirements Engineering Visualization*, Sept. 2006.

[13] R. A. Gandhi and S.-W. Lee. Visual analytics for requirements-driven risk assessment. In *Second International Workshop on Requirements Engineering Visualization*, October 2007.

[14] F. Hardisty and A. C. Robinson. The geoviz toolkit: using component-oriented coordination methods for geographic visualization and analysis. *International Journal Geographical Information Science*, 25:191–210, February 2011.

[15] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002.

[16] A. Inselberg. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer, September 2009.

[17] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. *Mastering the information age: Solving problems with Visual Analytics*. EuroGraphics, 2010.

[18] R. Lane, D. Nevell, S. Hayward, and T. Beaney. Maritime anomaly detection and threat assessment. In *13th Conference on Information Fusion*, pages 1 –8, July 2010.

[19] R. Laxhammar. Anomaly detection for sea surveillance. In *11th International Conference on Information Fusion*, pages 1 –8, July 2008.

[20] M. Mansouri, R. Nilchiani, and A. Mostashari. A risk management-based decision analysis framework for resilience in maritime infrastructure and transportation systems. In *3rd Annual IEEE Systems Conference*, pages 35 –41, March 2009.

[21] M. Migut and M. Worring. Visual exploration of classification models for risk assessment. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 11–18, October 2010.

[22] B. Ristic, B. La Scala, M. Morelande, and N. Gordon. Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In *11th International Conference on Information Fusion*, pages 1 –7, July 2008.

[23] J. Roy and M. Davenport. Categorization of maritime anomalies for notification and alerting purpose. *NATO Workshop on Data Fusion and Anomaly Detection for Maritime Situational Awareness*, pages 15–17, September 2009.

[24] S. Rudolph, A. Savikhin, and D. Ebert. Finvis: Applied visual analytics for personal financial planning. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 195 –202, Oct. 2009.

[25] N. M. Sanusi and N. Mustafa. A visualization tool for risk assessment in software development. In *International Symposium on Information Technology*, pages 1–4, August 2008.

[26] A. Savikhin, R. Maciejewski, and D. Ebert. Applied visual analytics for economic decision-making. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 107–114, October 2008.

[27] R. Scheepens, N. Willems, H. van de Watering, and J. J. van Wijk. Interactive visualization of multivariate trajectory data with density maps. In *Proceedings of the Pacific Visualization 2011*, pages 147–154, March 2011.

[28] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1986.

[29] C. G. Soares and A. P. Teixeira. Risk assessment in maritime transportation. *Reliability Engineering and System Safety*, 74(3):299 – 309, 2001.

[30] J. Stasko, C. Gorg, Z. Liu, and K. Singal. Jigsaw: Supporting investigative analysis through interactive visualization. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 131–138, 2007.

[31] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The R&D Agenda for Visual Analytics*. IEEE Press, 2005.

[32] U. S. C. G. U.S. Department of Homeland Security. *U.S. Coast Guard Addendum to the United States National Search and Rescue Supplement (NSS) to the International Aeronautical and Maritime Search and Rescue Manual (IAMSAR)*. September 2009.

[33] J. J. V. Wijk and E. R. V. Selow. Cluster and calendar based visualization of time series data. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 4–9, San Francisco, CA, USA, 1999. IEEE Computer Society.

[34] N. Willems, H. van de Wetering, and J. J. van Wijk. Visualization of vessel movements. *Computer Graphics Forum*, 28(3):959–966, 2009.

Abstracting Attribute Space for Transfer Function Exploration and Design

Ross Maciejewski, *Member, IEEE,* Yun Jang, Insoo Woo, Heike Jänicke, *Member, IEEE,*
Kelly P. Gaither, *Member, IEEE,* and David S. Ebert, *Fellow, IEEE*

Abstract—Currently, user centered transfer function design begins with the user interacting with a one or two-dimensional histogram of the volumetric attribute space. The attribute space is visualized as a function of the number of voxels, allowing the user to explore the data in terms of the attribute size/magnitude. However, such visualizations provide the user with no information on the relationship between various attribute spaces (e.g., density, temperature, pressure, x, y, z) within the multivariate data. In this work, we propose a modification to the attribute space visualization in which the user is no longer presented with the magnitude of the attribute; instead, the user is presented with an information metric detailing the relationship between attributes of the multivariate volumetric data. In this way, the user can guide their exploration based on the relationship between the attribute magnitude and user selected attribute information as opposed to being constrained by only visualizing the magnitude of the attribute. We refer to this modification to the traditional histogram widget as an abstract attribute space representation. Our system utilizes common one and two-dimensional histogram widgets where the bins of the abstract attribute space now correspond to an attribute relationship in terms of the mean, standard deviation, entropy, or skewness. In this manner, we exploit the relationships and correlations present in the underlying data with respect to the dimension(s) under examination. These relationships are often times key to insight and allow us to guide attribute discovery as opposed to automatic extraction schemes which try to calculate and extract distinct attributes a priori. In this way, our system aids in the knowledge discovery of the interaction of properties within volumetric data.

Index Terms—Transfer function design, volume rendering, information theory

✦

1 INTRODUCTION

ONE common method that allows scientists to visually query their volume rendered data is the interactive transfer function widget. Typically, the transfer function widgets rely on either a one-dimensional (1D) or 2D histogram frequency plot of the volumetric data (e.g., [1], [2], [3]), and it is often assumed that users will have an underlying knowledge of regions of interest within the data. By using this underlying knowledge, users will then be able to create an appropriate transfer function that maps voxel values to a given attribute structure. While scientists exploring their volumetric data do have an underlying knowledge of their data, they are often novice transfer function designers.

- *R. Maciejewski is with the Arizona State University, PO Box 878809 Tempe, AZ 85287-8809. E-mail: rmacieje@asu.edu.*
- *Y. Jang is with the Sejong University, 98 Gunja-Dong, Gwangjin-Gu, Seoul 143-747, South Korea. E-mail: jangy@sejong.edu.*
- *I. Woo and D.S. Ebert are with the School of Electrical and Computer Engineering, Purdue University, Electrical Engineering Building, 465 Northwestern Avenue, West Lafayette, Indiana 47907-2035. E-mail: iwoo@purdue.edu, ebertd@ecn.purdue.edu.*
- *H. Jänicke is with the University of Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany. E-mail: heike.leitte@iwr.uni-heidelberg.de.*
- *K.P. Gaither is with the Texas Advanced Computing Center, University of Texas at Austin, Research Office Complex 1.101, J.J. Pickle Research Campus, Building 196, 10100 Burnet Road (R8700), Austin, Texas 78758-4497. E-mail: kelly@tacc.utexas.edu.*

A typical 1D transfer function widget is shown in Fig. 1(top). This widget was presented to five novice users (but domain experts) as part of an informal user opinion survey. The users were asked to interactively select an area of interest based solely on the histogram distribution. All users selected the singular peak of the 1D histogram. However, as shown in the resultant volume rendering of Fig. 1(top), the peak in this particular histogram is actually the least interesting portion of the data. When asked to further suggest areas of interest, the users indicated that they would be unsure of where to explore next within this data. Three users selected various ranges that would be of interest to them; however, all users indicated that they were likely to start exploring such data at peak values.

In order to overcome such issues, we propose to enhance the conventional 1D and 2D histogram transfer function widgets to better capture attributes and correlations that exist in the data. To do this, we create a new mapping of the statistical properties found within the histogram bins and overlay this as a new visual representation within the traditional histogram views. Essentially, an underlying statistical property of the voxels within each histogram bin is computed with respect to a user selected attribute of interest. This statistical value is then used to guide interaction within the transfer function space, providing a visualization of the *abstracted attribute space* representation. No new attribute spaces are being created; instead, the user is presented with more information about the statistical relationships between volumetric attributes.

The same five users were presented with the new abstracted attribute space shown in Fig. 1(bottom). Again, they were asked to interactively select an area of interest. In the abstracted attribute space, the users now have a variety
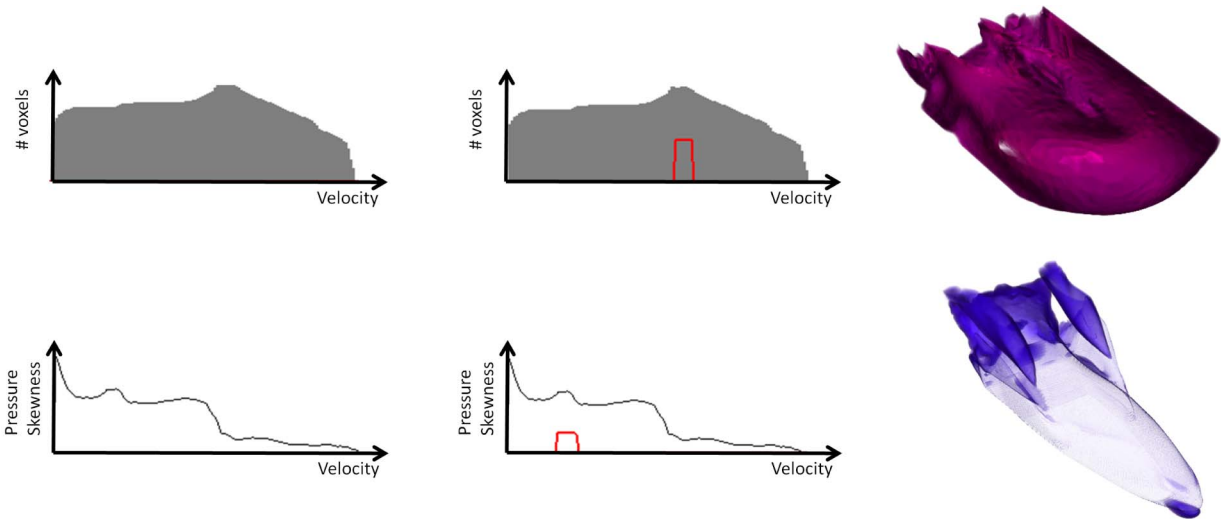
Fig. 1. Comparing explorations of the X38 experimental aircraft data using a traditional 1D histogram widget (top) and the pressure skewness abstracted 1D histogram widget (bottom). In the top image, the user is presented with only the velocity histogram. By selecting the area with the largest number of voxels, no interesting structures will be visible. By selection, we mean that the red line shown will now map opacity to the corresponding voxels. In bottom image, the user is presented with the abstracted histogram showing the skewness of the velocity histogram bins with respect to the pressure distribution in each bin. Here, the user explores the noticeable peak in the data and finds that it corresponds to the vortex cores in the data. Note that some of these regions correspond to interesting isosurface values that would be less intuitive to select when looking at the nonabstracted histogram plot. However, the same transfer function design in either widget would result in the same volume rendering.

of peaks to choose from, and many of these peaks map to interesting features of their data. Fig. 1(bottom) illustrates the resultant rendering of the most common peak chosen by the users. Other areas chosen include the leftmost area of high slope and the rightmost area of high slope.

Such an abstraction does not improve the user's ability to separate previously inseparable features. In fact, the traditional 1D histogram widget and the abstracted 1D histogram widget map the exact same voxels to the same histogram bin. Thus, if features of the volume (e.g., skin, bone) were to map to the same density bin in the original histogram, they would still map to the same bin in the abstracted space. The benefit of abstracting the histogram space is that it can provide insight to novice users with regards to the initial creation of an appropriate transfer function. Furthermore, such an abstraction is also beneficial for expert users as these statistics provide more information helping target search and analysis while exploring relationships between variables that cannot be shown with the traditional 1D histogram.

For example, extracting, representing, and understanding the properties and interactions within computational fluid dynamic (CFD) simulations has generally proven to be a difficult problem. Scientists are interested in locating critical points, vortices, shocks, and other attributes within their data and understanding the effects that various attributes (temperature, pressure, velocity, etc.) have on the volumetric flow structure. However, locating and visualizing such attributes is extremely difficult as the criteria defining flow features (e.g., shock) are often not well understood, imprecisely defined, and complex to extract volumetrically. For example, Banks and Singer [4] review eight different schemes for identifying vortices, and Ma et al. [5] listed three data properties that indicate the presence of a shock within the data. Thus, in extracting flow properties there is a need for enhanced information in understanding the interactions between the volumetric flow attribute values.

Furthermore, scientists are not only interested in the volumetric shapes and locations of the attributes, but also how these attributes interact and influence other attributes. Examples of questions that scientists may want to ask of their data are: "What effect do pressure and temperature have on velocity, and what does this correlation reveal about the physical domain?" By providing domain experts with an overlay of statistical relationships between variables in the histogram space, hypotheses can be formed and insights confirmed.

Thus, our work focuses on the mapping of statistical properties in combination with the traditional histogram widgets, thereby providing users with more information and cues on where they can begin volume exploration, as well as aiding in both transfer function design and knowledge discovery. The contribution of this work is the algorithm for generating the abstracted attribute space as well as the introduction of this space into the traditional 1D and 2D histogram widgets. By using our algorithm and modified widgets, users are able to explore the 1D abstracted transfer function space combining a variety of attributes (e.g., density, temperature, pressure, or x, y, and z dimensions simultaneously). Users can also toggle between the conventional 2D transfer function view, (in which the entries in the 2D histogram are colored by the number of voxels that map to a location) and the abstracted transfer function view, (in which the entries are colored by a derived statistical relationship). In this manner, we explore the usefulness of abstracting statistical properties in transfer function widgets and illustrate the effects on the exploration of the attribute space. We discuss the benefits and drawbacks of such an approach and illustrate our results across various volumetric data sets.

2 RELATED WORK

Interactive transfer function design has been addressed with many different approaches, ranging from simple (yet

intuitive) 1D transfer functions (e.g., [1], [2]) in which a scalar data value is mapped to color and opacity, to more complex multidimensional transfer functions in which color and opacity are mapped across multiple variables. Early work by Kindlmann and Durkin [6] and Kniss et al. [3] applied the idea of a multidimensional transfer function [7] to volume rendering. This work identified key problems in transfer function design, noting that many interactive transfer function widgets lack the information needed to guide users to appropriate selections, making the creation of an appropriate transfer function essentially trial-and-error which is further complicated by the large degrees of freedom available in transfer function editing. While many volume rendering systems have adopted multidimensional transfer function editing tools, the creation of an appropriate transfer function is still difficult as the user must understand the dimensionalities of the attribute space that they are interacting with.

Recent work on transfer function design has proposed higher dimensional transfer functions based on mathematical properties of the volume. Examples include the Contour Spectrum by Bajaj et al. [8] which proposed a user interface for displaying computed contour attributes using the surface area, volume, and the gradient integral of the contour. Work by Kindlmann et al. [9] employed the use of curvature information to enhance multidimensional transfer functions, and Tzeng et al. [10] focused on higher dimensional transfer functions which use a voxel's scalar value, gradient magnitude, neighborhood information, and the voxel's spatial location. Work by Potts and Möller [11] suggested visualizing transfer functions on a log scale in order to better enhance attribute visibility. Lundström et al. introduced the sorted histogram [12], the partial range histogram [13], and the α-histogram [14] as means for incorporating spatial relations into the transfer function design. Correa and Ma introduced size-based transfer functions [15] which incorporate the magnitude of spatial extents of volume attributes into the color and opacity channels and visibility-based transfer functions [16] where the opacity transfer function is modified to provide better visibility of attributes. Maciejewski et al. [17] proposed a method to structure attribute space in order to guide users to regions of interest within the transfer function histogram, and Bruckner and Möller [18] depicted similarities of isosurfaces through the use of mutual information theory. While such extensions enhance the volume rendering and provide a larger separability of volumetric attributes, they still fail to provide users with information about the structures within a given attribute space. In fact, the addition of more dimensionality into the transfer function is often automatically incorporated into the rendering parameters, obscuring the relationship between the volumetric properties and the volume rendering.

Other work has focused on means of better displaying the data dimensionality to users, aiding the attribute space exploration. Shamir [19] applied attribute-space cluster analysis to unstructured meshes in order to automatically incorporate spatial information for identify structures within the volume. This clustering is performed across a 5D space (the x, y, and z components of the volume and the value versus value gradient magnitude attribute space), where as we only perform this on the attribute space.

However, the goal of Shamir's work was volume segmentation as opposed to transfer function design and interactive attribute extraction. Work by Roettger et al. [20] also generated transfer functions through attribute space analysis. They enable the automatic setup of multidimensional transfer functions by adding spatial information to the histogram of the underlying data set. Work by Henze [21] developed a system in which users can interact with a variety of attribute space views, interactively brushing and linking data, thereby allowing the user to define a set of data conditions over a variety of attributes. Woodring and Shen [22] proposed a method in which the user is presented with several different volumetric renderings and is able to compare values from these data sets over space and time, and combine various rendering attributes into one volumetric rendering. Akiba et al. [23], [24] utilized parallel coordinate plots to create a volume rendering interface for exploring multivariate time-varying data sets. By means of a prediction-correction process, Muelder and Ma [25] proposed to predict the attribute regions in the previous frame, making the attribute tracking coherent and easy to extract the actual attribute of interest.

In all of this related work, one can note that various statistical properties of the volumes are being used in order to extract attributes of interest and segment properties of the volume. Unfortunately, as the number of dimensions increases, interaction in n-dimensional space becomes cumbersome to the point that few systems exceed 2D transfer functions; instead, the extra dimensionality is incorporated automatically, somewhat limiting the user's control. In order to enhance the information provided in the transfer function histogram widget, our work focuses on incorporating statistical properties of user selected attributes into the projected attribute space domain. Recent work on incorporating statistical and information metrics into time-varying volumetric data includes work by Fang et al. [26] on the time activity curve, Jänicke et al. on local statistical complexity analysis [27], and Haidacher et al. [28] on utilizing information theory for fusing traditional transfer function space with information enhanced transfer function spaces.

In exploring volumetric attribute space, one can think of the 2D histogram widget as a special form of a scatterplot where the points are binned as opposed to plotted individually. Scatterplots have long been recognized as a useful tool for multidimensional visualization due to their relative simplicity and high visual clarity [29], [30], and have been incorporated in a number of multidimensional visualization toolkits (e.g., Tableau/Polaris [31], GGobi [32], XmdvTool [33]). While recent work explored methods to enhance user interactions across the entire attribute space of a multidimensional data set using scatterplot matrices [34], little work has been done in regards to incorporating attribute relationship visualizations into scatterplots. Some work in that area includes methods by Wang et al. [35] that introduces an importance-driven time-varying data approach, work in brushing moments by Kehrer et al. [36], and the statistical transfer-function space work by Haidacher et al. [37]. In Wang et al. [35], a user is presented with importance curves based on the temporal component of the data. Other work on analyzing statistical properties within
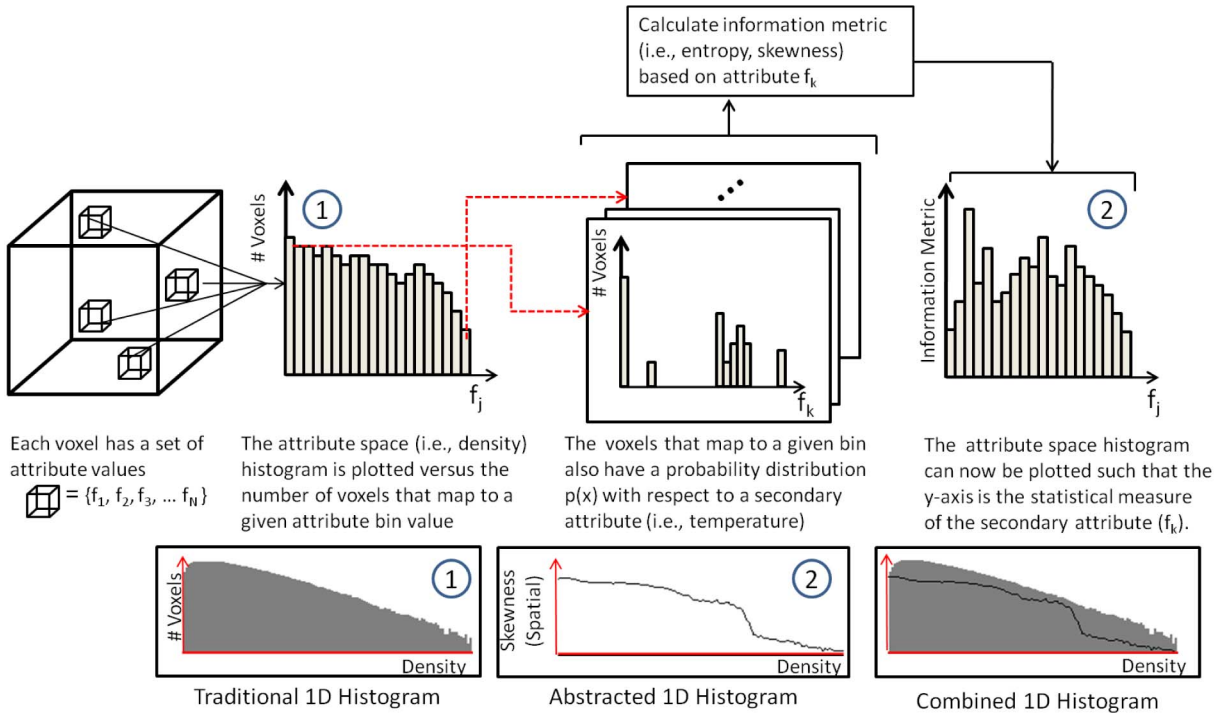
Fig. 2. The attribute space abstraction pipeline for the 1D histogram case. The algorithm is the same for the 2D histogram case.

transfer function space includes the work by Bachthaler and Weiskopf. Their work includes a method to create a continuous scatterplot [38] relating traditional discrete histograms to the histograms of isosurface statistics and they extend this work [39] to utilize subdivision within the spatial domain for creating continuous scatterplots and show their applications in volumetric rendering. In Kehrer et al. [36] the authors estimate statistical moments of the data including the mean, standard deviation, skewness, and kurtosis and apply this to the analysis of multivariate scientific data. In Haidacher et al. [37] the authors adaptively estimate statistical properties including the mean and standard deviation of the data values within a given neighborhood. The authors perform similarity tests within regions to determine if a material within a neighborhood is homogenous and create a new statistical transfer function space. Our work is similar to previous work where the user is presented with a variety of information metrics based on user defined specifications; however, work by Haidacher et al. creates a new transfer function space to enable the distinction of new materials. Our method does not enable the distinction of new materials. Instead, our method overlays the relationship of the current attribute being visualized in the histogram with another user selected attribute, thus providing fundamentally different views and functionality to the user.

3 ABSTRACT ATTRIBUTE SPACE GENERATION

Current transfer function design widgets typically present users with a low-dimensional projection of the volumetric attribute space onto an interactive one or 2D histogram widget. These projections represent the magnitude of a given attribute (i.e., the number of voxels that map to a given attribute). Attribute selection is then accomplished

through traditional brushing of the histogram, and the voxels that map to a given attribute are assigned a color and opacity by the user.

In this work, we propose a modification to the traditional histogram transfer function design widgets. Instead of visualizing the attribute space as a function of the magnitude of the number of voxels, our work utilizes a novel pipeline that presents users with information metrics about a given attribute set. We refer to this information enhanced attribute space as an *abstract attribute space*. In our transfer function widget, a user is presented with information about the interactions between attributes within their data set as opposed to the magnitude of the volumetric attributes. Previous work has shown that there are links between the geometric properties of isosurfaces and the statistical properties of data [40], [41], and our information enhanced transfer function widgets help aid scientists in understanding these (and other) links. These tools aid the users by promoting new insights into volumetric data and enhance the knowledge discovery process.

Fig. 2 illustrates our abstraction pipeline in the 1D histogram attribute space case (note that the 2D abstraction process is exactly the same). Beginning with the volumetric data, each voxel consists of a set of N measured or derived attributes, $\{f_1, \ldots, f_N\}$. The user chooses any one of these volumetric attributes, $\{f_j\}$ in Fig. 2 (for example, f_j could be density), and a histogram distribution of this attribute is created. This histogram is typically found in the 1D transfer function editor widget common in many volume rendering systems.

We modify the histogram by taking the voxels that map to a given bin in the $\{f_j\}$ attribute space histogram and derive an information metric about that attribute space with respect to one or more of the remaining attributes in the

attribute space set, $\{f_k\}$ (for example, f_k could be temperature). Since the voxels corresponding to each bin in the $\{f_j\}$ histogram are known, we are able to map these voxels to a 1D distribution with respect to the user defined attribute of interest, f_k. This mapping is never shown to the user; instead, it is used to derive a set of information metrics, this information metric can then be presented to the user as a 1D plot, as shown in Fig. 2. Thus, the user is presented not with the magnitude of an attribute, but with a derived set of information about the attribute with respect to a secondary volumetric feature.

We have also extended our process to the 2D histogram widget as well. In the case of the 2D widget, the traditional 2D transfer function editor widget utilizes two attributes of the volume data $\{f_i, f_j\}$. The original histogram is plotted such that each bin is colored based on the number of voxels in each bin. The application of our algorithm will utilize a third attribute of the volume data $\{f_k\}$ and calculate the statistical properties of this feature for all voxels found in a given bin in the 2D histogram. The original histogram is then redrawn, where the color is now mapped to the derived statistical property. Examples of both the 1D and 2D abstract attribute space generation are shown throughout this paper.

In this work, we survey the use of four information metrics: mean, standard deviation, skewness, and entropy. This section details the information metric computations used, thereby formalizing the final (calculate information metric) step in our abstraction pipeline (Fig. 2). This final step in our pipeline is extensible, and future work will explore the use of other information metrics for transfer function design.

All information metrics are calculated as a precomputation step during volume loading and the resulting attribute space plots can be toggled between in real time. Precomputation time for a $512 \times 256 \times 128$ floating point data set takes approximately 9,093 ms for the combined 2D skewness calculations, 513 ms for the combined 2D entropy calculations, 97 ms for the combined 1D skewness calculations, and 20 ms for the combined 1D entropy calculations. Note that the mean and standard deviation calculations are incorporated as part of the skewness calculation. Results are from a $256 \times 256 \times 256$ data set on an Intel Xeon(R) CPU E5335 2.00 Ghz machine with 4 GB of RAM. Timings scale linearly with respect to the number of voxels.

3.1 Mean

Given an entry m in the attribute space $\{f_j\}$, we define $S_V = \{V_1, \ldots, V_Z\}$ to be the set of voxels that map to this given volume attribute. There are Z voxels that will map to this space. Within S_V, we calculate the mean, $\mu(S_V)$, with respect to the user chosen secondary attribute $\{f_k\}$

$$\mu(S_V) = \frac{1}{|S_V|} \sum_{x \in S_V} f_k(x). \qquad (1)$$

The calculated mean thus depicts the average value of the f_k attribute with respect to a value range for f_j. Furthermore, this mapping provides users with more information about their data by showing them how a certain value of attribute f_j is related to values in f_k.

3.2 Standard Deviation

Once the mean is calculated, we can then determine the standard deviation, $\sigma(S_V)$, with respect to the secondary attribute $\{f_k\}$.

$$\sigma(S_V) = \sqrt{\left(\frac{1}{|S_V|} \sum_{x \in S_V} (f_k(x) - \mu(S_V))^2 \right)}. \qquad (2)$$

The standard deviation of the set, S_V with respect to the f_k attribute of the voxels is then mapped back to the corresponding bin in the $\{f_j\}$ histogram attribute space.

The standard deviation thus depicts the variance of the f_k attribute values corresponding to a particular value of f_j. Areas of low standard deviation typically indicate a high correlation between the bin range of f_j and the average value of f_k found with respect to that bin. Such information can reveal relationships between data. For example, a scientist may wish to explore temperature ranges of their data in which the pressure is highly fluctuating. To do this, they would interactively select the temperature ranges where the standard deviation of the pressure is found to be large. By using this attribute mapping, such ranges can be immediately found, whereas in the traditional 1D histogram users would only be unable to see such relationships.

3.3 Skewness

Skewness is a measure of the asymmetry of the distribution of the underlying data. The skewness of a random variable is the third standardized moment about the mean and standard deviation of an underlying data distribution. Given the set of voxels S_V that map to the attribute space $\{f_j\}$ at m in the attribute space histogram, the skewness, $\gamma(S_V)$, of the voxels at that location with respect to an attribute $\{f_k\}$ (e.g., velocity, enstrophy) is calculated.

$$\gamma(S_V) = \frac{\frac{1}{|S_V|} \sum_{x \in S_V} (f_k(x) - \mu(S_V))^3}{\left(\frac{1}{|S_V|} \sum_{x \in S_V} (f_k(x) - \mu(S_V))^2 \right)^{\frac{3}{2}}}. \qquad (3)$$

Our choice of utilizing skewness as an attribute space abstraction metric is influenced by the work of Patel et al. [42] in which the authors demonstrate the effectiveness of using high order statistical moments for transfer function generation. Skewness may be either positive or negative. In the 1D transfer function case, we add a secondary axis to the plots to show positive and negative skewness. In the 2D transfer function case, we utilize a blue to red color map with blue values representing negative skewness and red being positive. Areas with high skewness typically represent inhomogeneity in the data, while areas of low skewness typically represent regions where the data are normally distributed implying some underlying structural homogeneity.

3.4 Entropy

For the entropy calculation, given the set of voxels S_V that map to the attribute space $\{f_j\}$ at m in the attribute space histogram, the entropy, $H(S_V)$, of the voxels at that location with respect to a user chosen secondary attribute $\{f_k\}$ can also be calculated
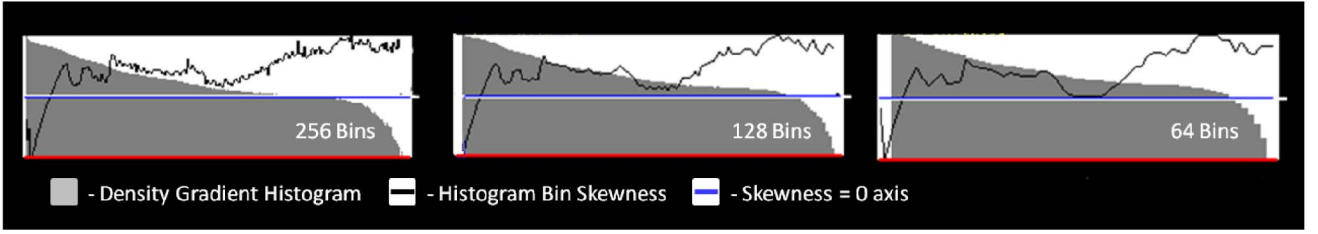
Fig. 3. Exploring the effect of the histogram bin size on abstracted attribute spaces. Here, we show the 1D density gradient histogram of the Bonsai CT data (in gray) and the skewness of the data (as a black line). As the number of bins increases, the statistical measures are smoothed.

$$H(S_V) = - \sum_{x \in S_V} p(f_k(x)) \ln(p(f_k(x))). \qquad (4)$$

Here, $p(f_k(x))$ is the probability that a voxel in the set S_V has a value $f_k = x$. Since S_V is known, we can simply use the one-dimensional histogram of these voxels with respect to their f_k attribute values to calculate the probability distribution. For example, if three voxels within S_V map to $f_k = q$, then $p(f_k = q) = 3/Z$ where Z is the total number of voxels in S_V. Once $H_S(*)$ is calculated, the entropy value is mapped back to the $\{f_j\}$ histogram attribute space. The entropy depicted at each position in the abstract attribute space provides information about the extent of the corresponding set of voxels.

With regards to spatial attributes (x, y, and z), spatially isolated structures often define an attribute and low spatial entropy is well suited to indicate them. High entropy, on the other hand, is reached with a uniform distribution of voxels over the physical domain. In this case, the combination of values is likely to belong to background dynamics and corresponding positions are often irrelevant for the user.

3.5 Histogram Binning Effects

In creating the abstracted attribute space, the results are directly related to the voxels that map to a given histogram bin. Thus, the bin size (or number of bins used) will directly impact the resultant abstracted attribute space visualization. In order to understand and explore the impact that histogram binning will have on our proposed method, we have modified the number of bins used and compared the resultant abstracted attribute space plots, as shown in Fig. 3. As the number of bins decreases, the overall shape of the skewness curve remains intact; however, minor peaks are smoothed out and the bin skewness values change.

We have further explored the effects of bin sizes over all data sets presented in this paper. The shape of all abstracted attribute space curves remains consistent for all statistical measures used in our work (i.e., mean, standard deviation, entropy, and skewness). Overall, the changes to the abstracted attribute space curves seem to be minimal; however, we plan to explore adaptive histogram binning as a means of potentially extracting stronger statistical correlations.

As with all histograms, there is no such thing as a "best number" for bin size. Different choices in bin width can reveal different features about the data; however, the size of the bin will directly impact our underlying statistical measures. As the bin width gets smaller, the number of voxels that maps to each bin becomes sparser. Such sparsity can bias the data resulting in details that may appear to be

random noise. Conversely, as the bins become larger, more data are being placed into each bin. In this manner, more voxels are being grouped together, and features within the volume will be less separable. Thus, it is important to choose a reasonable bin width when employing the use of our abstracted feature space metrics. Preliminary work indicates that using equal interval bins, where the number of bins is approximately the square root of the number of voxels in the volume is a reasonable approach.

4 ATTRIBUTE SPACE EXPLORATION TOOLS

In order to better facilitate volume exploration through transfer function design, we have developed a small suite of attribute space exploration tools to complement the traditional interactive transfer function design metaphors. Our tools include an expanded 1D histogram view, linked 1D and 2D histogram views for explorations into manual attribute segmentation, and an opacity brushing tool for highlighting complex attribute space structures such as arcs.

4.1 Expanded 1D Histogram View

The traditional 1D histogram view consists of the frequency space of the voxel data with a particular volumetric attribute (value, value gradient magnitude, etc.) being assigned to the x-axis of the histogram and the number of voxels this attribute maps to is assigned to the y-axis. Our expanded 1D histogram view utilizes this convention and plots the frequency distribution of both the value and value gradient magnitude space in a 1D histogram widget tool. This frequency plot is the gray background plot seen in the histograms of Fig. 4(left). Note that the left set of histograms show the value versus frequency and the right set of histograms show the value gradient magnitude versus frequency. From top to bottom, the histograms are then overlaid with a statistical plot representing some spatial information (in this case the standard deviation) of the voxels with respect to their x, y, and z positions. User interaction in any of the three windows will modify the same transfer function applied to either the value or value gradient magnitude attribute space. However, only one transfer function (the value attribute space transfer function or the value gradient attribute space transfer function) is applied for volume rendering based on the most recently edited transfer function.

4.2 Linked 1D-2D Exploration View

While the user may create a 1D transfer function, work has shown that by increasing the dimensionality of the attribute space under exploration, the separability of more attributes
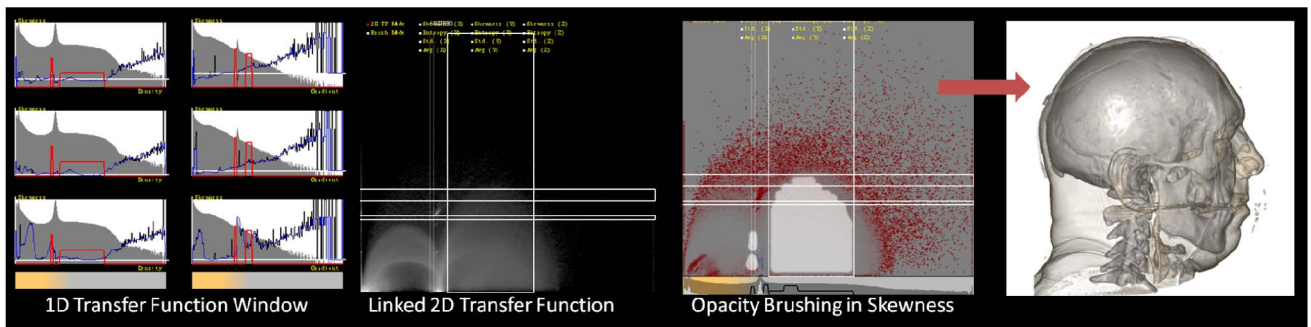
Fig. 4. Attribute space explorations tools. (Left) A hexa-window 1D histogram view showing the user the standard deviation spatial properties of the Visible Male Head data set for the x, y, and z directional data (the top, middle, and bottom views, respectively) in the density (left) and density gradient magnitude (right) attribute space. The gray area is the frequency distribution of the data. The black line is the abstracted attribute space property. The red line is the user defined transfer function representing the opacity. (Middle-Left) A 2D histogram view of the density versus density gradient magnitude histogram view with linked regions brushed based on the user defined 1D transfer function widget. (Middle-Right) A user defined transfer function in the 2D histogram view where opacity brushing was applied to explore the linked region. (Right) The resultant volume rendering from the opacity brushed transfer function. Note that the 1D transfer function was designed with respect to the z-directional skewness.

becomes possible (although this is highly influenced by the choice of dimensions). As the user modifies the transfer function space in the 1D histogram views, a linked 2D histogram view is created in which the corresponding bins are highlighted by rectangular bounding boxes so that the user may explore further details within the 2D attribute space. This linked view is shown in Fig. 4(Middle-Left). Here, the white rectangular regions depict the areas of nonzero opacity brushed by the user in the 1D histogram space.

4.3 Opacity Paintbrushing

Finally, to enhance user interaction, we utilize an opacity paintbrush tool for the 2D histogram widget similar to the painting tool described in Co et al. [43]. Traditional transfer function design widgets for 2D histograms include drawing a series of rectangular boxes and defining a color and opacity within the selected region. However, most structures that appear in the 2D attribute space are unable to easily fit within a rectangular widget. To overcome this, we utilize an opacity brushing tool in which the user has applied some underlying color map to the data, either automatically as done by Maciejewski et al. [17] or manually through rectangular widget creation or

through the linked coloring from the 1D transfer function design. Fig. 4(Middle-Right) demonstrates a user created transfer function through the use of our opacity brushing widget. The application of this transfer function for volume rendering is seen in Fig. 4(Right).

4.4 Scaling by View Direction

In multivariate volumetric data, there are many attributes that can be utilized in the attribute abstraction process. In computational fluid dynamic simulations, users may have variables such as spatial position, temperature, pressure, vorticity, etc. However, in CT data, the attributes are often limited to only the density and the spatial position. While each coordinate of the spatial attribute could be mapped to its own separate value, this then requires the user to mentally integrate the three vector components. To handle this case, our system provides the users with two options: each component of the spatial attribute (x, y, and z) are plotted in a separate graph and presented to the user (as shown in Fig. 4), or the view direction is used to scale the information vector and the scaled value is shown to the user (Fig. 5). The scaling of the spatial attribute vector based on the view direction is defined as follows:
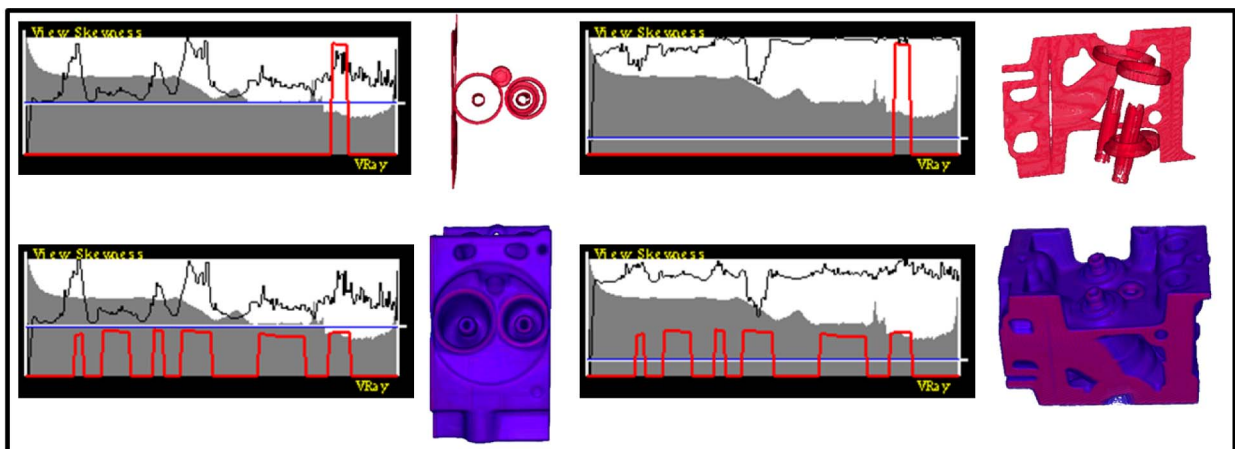
Fig. 5. Exploring CT data using abstracted attribute spaces. The histogram shows the 1D density attribute space of the Engine data. Here, the user has selected various regions of high skewness for analysis and explores the effect the view direction has on the data skewness. Note that the underlying histogram distribution is the same in both the right and left image transfer functions; however, the overlaid skewness plot varies based on the viewing angle. The red line in the histogram widget is the user defined opacity.
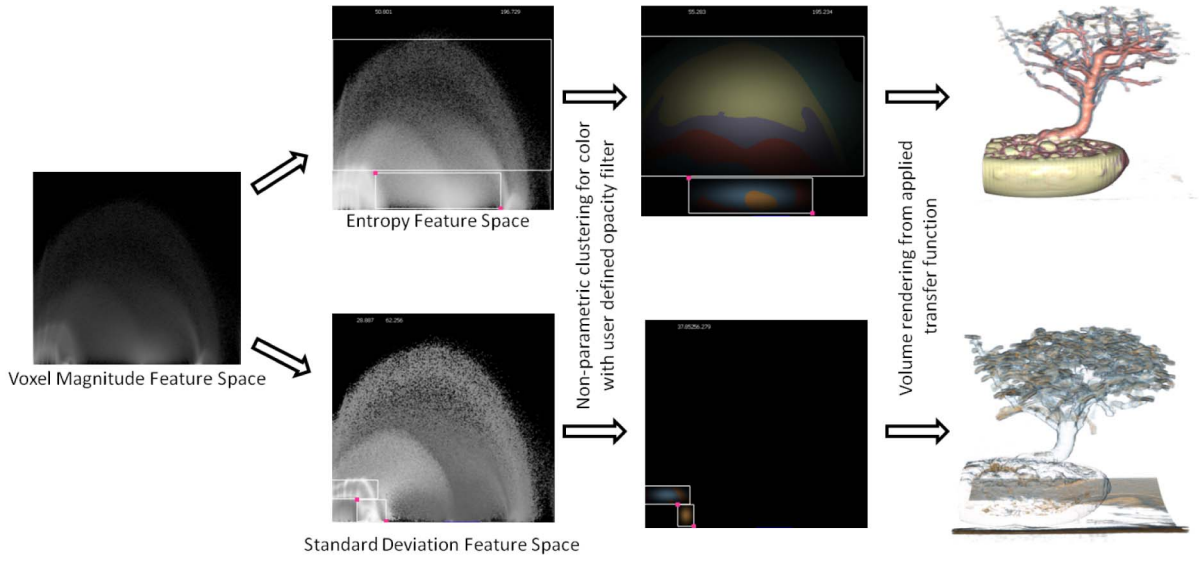
Fig. 6. Exploring CT data using abstracted attribute spaces. The density versus density gradient magnitude attribute space of the Bonsai data.

$$S = \frac{A_x V_x + A_y V_y + A_z V_z}{\sqrt{A_x^2 V_x^2 + A_y^2 V_y^2 + A_z^2 V_z^2}}, \tag{5}$$

where A_x, A_y, and A_z are the information metrics calculated with respect to the spatial attributes (x, y, and z), and V is defined as the view directional vector calculated as

$$V = M^{-1}[0\ 0\ -1]^T, \tag{6}$$

where M is the camera matrix.

We illustrate the effect that scaling by view direction has on the data exploration process in Fig. 5. Here, we explore skewness as an information metric in CT data utilizing the x, y, and z components of the voxels in the volumetric space and calculate the skewness with respect to each direction for each bin in the 1D histogram. The spatial skewness is then scaled and normalized based on the view directional vector of the volume and plotted as a black line on top of the density histogram plot. In the density 1D attribute space (Fig. 5(top)), we can see that the skewness provides new peaks in the data to explore that are not evident in the traditional frequency plot (the underlying gray histogram).

In Fig. 5(Top-Left) the user explores a peak of particular interest in the upper range of the density values. As the viewpoint changes, the user defined transfer function remains constant; however, the spatial skewness changes as is seen in Fig. 5(Top-Right). Using the same view point in Fig. 5(Bottom-Left), we highlight all the skewness plateaus and explore the resulting visualization. Again, we see that changing the viewpoint results in a change in the plotted spatial skewness Fig. 5(Bottom-Right).

Here, the user has interactively selected various areas in the plots based on either high skewness peaks or rates of change in the skewness. This process is similar to that shown in Fig. 1 where the user explores various peaks in the abstracted space. Our final rendering shows only a subset of the peaks after exploration was done. From this we can see that the information metric provides new cues as to which regions in a data set may be of interest for visualizing results. Depending on the data alignment,

different information about material boundaries and segmentation regions can be found. Future work will look at incorporating these various information metrics as a means of shading parameters.

5 ATTRIBUTE SPACE EXPLORATION EXAMPLES

In this section, we survey the meaning of our abstract attribute space representations in the context of the related volumetric properties. First, we illustrate the effect of using the abstract attribute space in a traditional CT data set in order to illustrate the meaning of the attribute space entropy. Second, we present a case study exploration of CFD data using the derived abstract attribute space representations. Finally, we explore the effect of noisy data on our derived abstract attribute space representations.

5.1 Spatial Abstract Attribute Spaces

5.1.1 Entropy and Standard Deviation

A key step in utilizing the abstract attribute space representation effectively for volume exploration is understanding the notion of entropy. As stated in Section 3.3, locations in the attribute space that consist of high spatial entropy are unlikely to represent structures within the data. In Fig. 6(Top), the user is comparing the traditional voxel magnitude-based attribute space representation to the entropy-based representation. Here, we can quickly see that the area of the highest entropy maps to that of the highest number of voxels. As this is likely the air surrounding the volumetric data, the user chooses to segment the data by utilizing a square opacity filter in the entropy attribute space. The underlying colors are based on the nonparametric transfer function generation of Maciejewski et al. [17]. Here, we see that the most spatially coherent volumetric structures are able to be quickly extracted from the data.

Given the spatial randomness of the leaves, it is likely that the spatial standard deviation will be able to guide the user in segmenting the leaves. In Fig. 6(bottom), the user switches to the spatial standard deviation attribute space. In the lower right corner, a particularly high band of standard
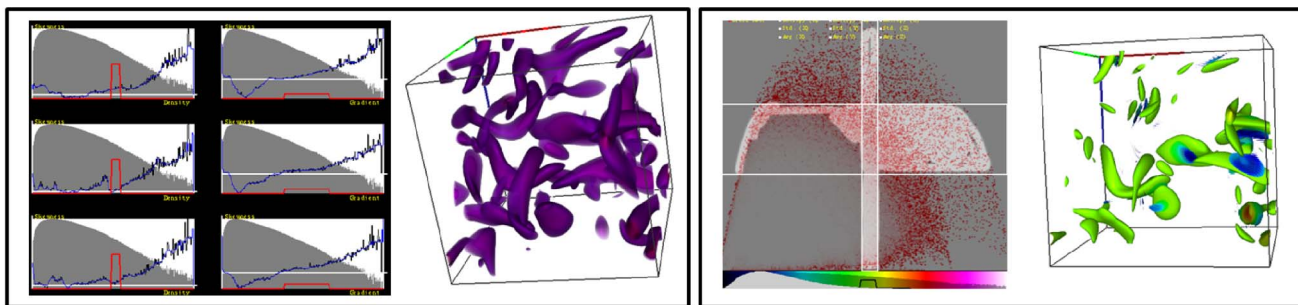
Fig. 7. Exploring the turbulent vortex data using skewness. (Left) The value attribute space of the vortex data where the user has extracted areas of low skewness in the z spatial direction and the resultant rendering. (Right) The value versus value gradient magnitude attribute space where the user has selected regions of high skewness within the previously selected 1D transfer function regions and the resultant rendering. The red line in each 1D histogram widget represents the user defined opacity.

deviation is evident to the user. Again, using a combination of square opacity filters, the user is able to directly segment the leaves. As such, one can see that by using a combination of abstract attribute space representations, a user is able to quickly and effectively explore their volumetric data.

While the use of abstract attribute spaces does allow the user to find very clean segmentation areas within the volume, in CT and MRI data, there are known properties of the density versus density gradient attribute space that allow users to create transfer functions by selecting the arc like structures within the attribute space. These arc-like structures correspond to material boundaries, and these cues have been shown to aid users in attribute extraction. However, CFD data do not have well-defined boundaries, and such methods are unable to extract structures such as shocks and vortices. Furthermore, when looking at an arbitrary attribute space, such as Pressure versus Velocity Magnitude, the relationships between data values may not be apparent.

5.1.2 Skewness

Fig. 7 illustrates the application of our abstract attribute space representation when applied to the turbulent vortex data set. The data set is sized at $128 \times 128 \times 128$ and is a fluid flow simulation that has been used in many studies [25], [44], [45]. Here, we explore regions of near zero skewness in the value gradient magnitude attribute space of

the data and the area of high skewness change in the value magnitude attribute space with respect to the y-directional component, see the transfer function designs in Fig. 7(left). Next, we change to the 2D attribute space of value versus value gradient magnitude and highlight the regions of high skewness within the bins chosen from the 1D attribute space interactions, Fig. 7(right). This method allows us to further refine the visualization that is of particular importance with turbulent flow simulations. Areas in turbulent flow where there is a high skew level represent potential attribute areas and should be more closely explored. This high skew level highlights change in the flow. Since turbulent flow attributes can be difficult or impossible to analytically describe, highlighting regions in the flow that have attributes of interest is a powerful tool for gaining insight to these data sets.

5.2 Nonspatial Abstract Attribute Spaces

5.2.1 Skewness

As previously stated, it is often that CFD data does not have well defined boundaries. Furthermore, scientists are often interested in relationships between not only position, but also measured properties, such as temperature and velocity. Fig. 8 demonstrates the abstraction of the attribute space with respect to nonspatial properties. In Fig. 8(top), the user is exploring the relationship between the velocity and the
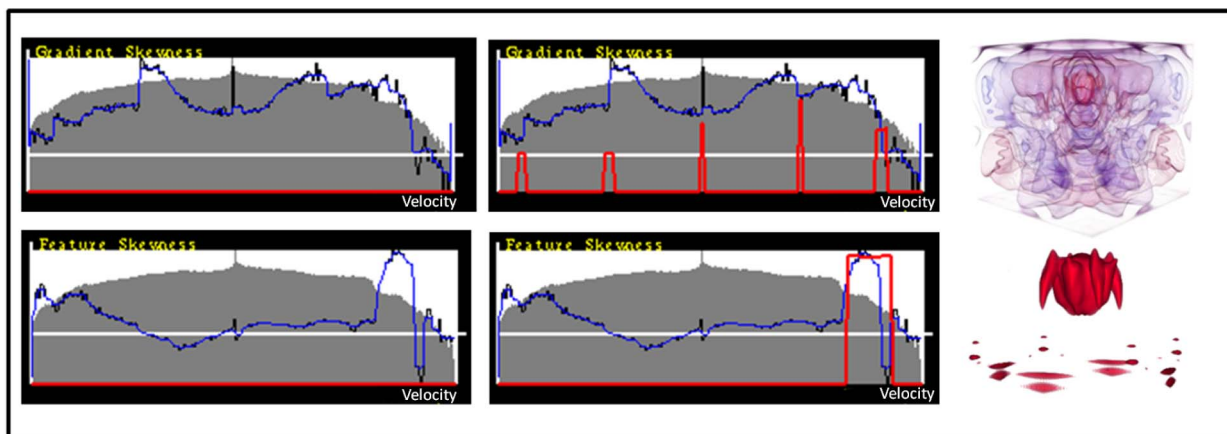
Fig. 8. Exploring the convection in a box data set with respect to nonspatial attributes (Time stamp 120). (Top) The velocity histogram overlaid with the velocity gradient skewness abstracted attribute space. (Bottom) The velocity histogram overlaid with the user defined attribute skewness (in this case the user is exploring the temperature when compared to the velocity). The red line in each 1D histogram widget is the user defined opacity.
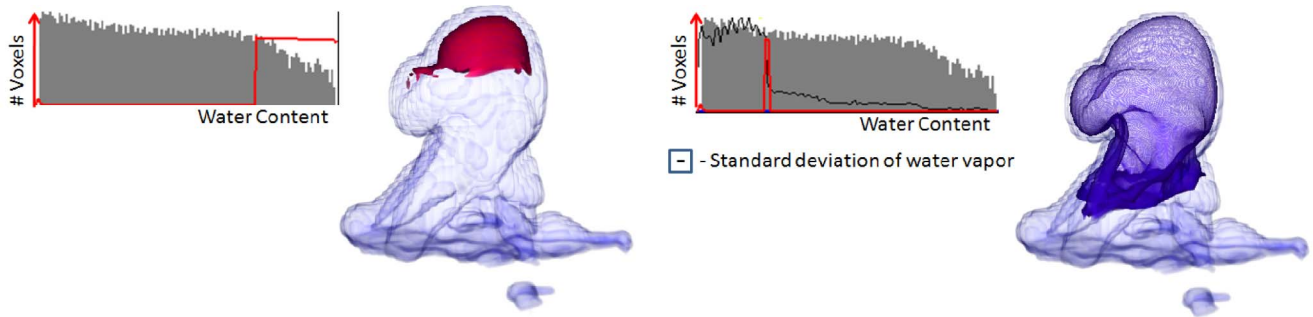
Fig. 9. Exploring the formation of a cumulus cloud with respect to the water content and water vapor values. The red line in each 1D histogram widget is the user defined opacity.

velocity gradient skewness. We create an oscillating transfer function, the peaks of which map to the skewness peaks in the gradient, creating an advection rendering similar to work by Svakhine et al. [46].

In Fig. 8(bottom), the user is exploring the relationship between the velocity and the temperature skewness. Here, the user has selected regions of highly variable temperature (both the highest and lowest range of skewness). The blue line in the bottom part of Fig. 8 is the skewness of temperature with respect to velocity. At high velocities, we can see that the temperature is nonconstant. Visualizing these values in abstract attribute space allows us to highlight areas of mixing. When discussing these plots with a domain scientist, they were able to point out that convective heat transfer takes place through both diffusion and advection, each of which can be highlighted in the abstract attribute space. The domain expert indicated that such plots were useful to her in exploring the data as it provided her with details about the underlying data relations as opposed to only a singular value distribution. In this case, the domain expert was a transfer function design expert as well. While the expert had certain value ranges in mind to explore as part of the transfer function design phase, the expert indicated that the enhanced information view presented was valuable for explaining phenomenon within the data. Furthermore, the expert indicated that such plots would be useful for detecting anomalies within simulations.

5.2.2 Standard Deviation

Our next data exploration example utilizes a cumulus cloud simulation. In Fig. 9, the user is presented with a histogram of the water content of the cloud. A conventional histogram (Left) does not provide the user with the information needed to quickly extract the boundary of the clouds evolution. However, if the user chooses to overlay this histogram with the abstracted attribute of the water vapor (Right), it is easier to extract the precise boundary of the cloud evolution.

The user analyzes the standard deviation of the change in water vapor with respect to the water content distribution. From this analysis, the user selects the discontinuity where the abrupt change in the standard deviation indicates a movement from regions of highly varying water vapor to regions of constant water vapor values. The resultant rendering allows the user to visualize the boundary of the cloud evolution, showing a boundary where there is

condensation from vapor to water droplets. The discontinuity would be somewhat near the edges of the visible boundary of the cloud and, as the cloud further evolves, this area may become quite turbulent. This helps show the initial entrainment of dry air into the cloud formation. Using the addition of the water vapor information, the user indicates that they are better able to match their mental model of the simulation to the rendering parameters. Here, the user has a concept of how the water content and water vapor will interact. By adding this information into the histogram visualization and transfer function design phase, the user can feel more comfortable in transfer function exploration and explore more properties of the data simultaneously.

Our final CFD example utilizes the X38 data set based on a tetrahedral finite element viscous calculation computed on geometry configured to emulate the X38 Crew Return Vehicle. The geometry and the simulation were computed at the Engineering Research Center at Mississippi State University by the Simulation and Design Center. This data set represents a single time step in the reentry process into the atmosphere. The simulation was computed on an unstructured grid containing 1,943,483 tetrahedra at a 30 degree angle of attack. however, for ease of testing in collaboration with the CFD researchers to guarantee accuracy, we resampled the data onto a $512 \times 256 \times 128$ regular grid.

Fig. 10 illustrates the application of our abstract attribute space representation when applied to the X38 data set. Here, the user compares the distribution of pressure values around the X38 to the underlying air velocity. In discussing with the domain scientist, we chose to view the standard deviation of the air velocity with respect to pressure. When looking at the plots, the domain scientist was interested in first exploring the high standard deviation of velocity in the low pressure area. This represented the body and the vortex cores (both rendered in blue). Next, the domain
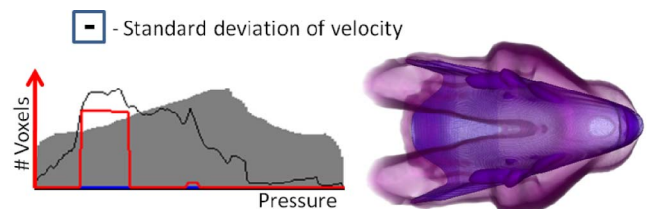
Fig. 10. Exploring the x38 data set with respect to pressure and velocity. The red line in each 1D histogram widget is the user defined opacity.
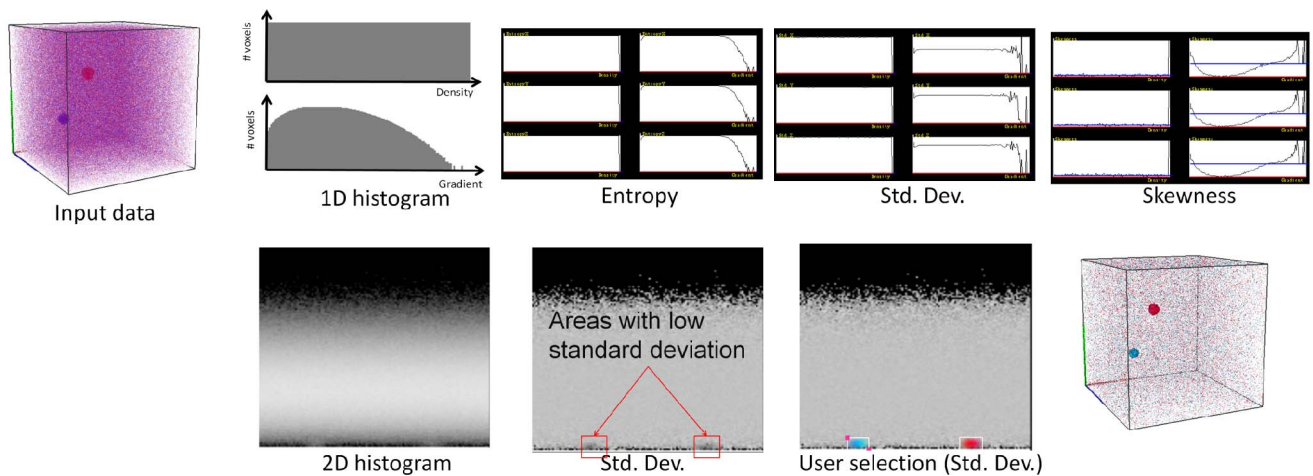
Fig. 11. Comparing the effects of noise on the abstract attribute space. Here, the user is exploring a volumetric data set with a density that approximates a uniform distribution; however, two small spheres are embedded in this distribution. As expected, the results for entropy, standard deviation and skewness match that of a uniform random distribution. However, in the 2D standard deviation abstraction, areas of low standard deviation value become visible indicating areas of interest.

expert also wanted to explore the sharp standard deviation peak in velocity which appears almost as a discontinuity. By interactively selecting these two regions, the scientist can visualize the bow shock, the vortex cores, and the secondary shocks.

While such properties could have been found while exploring the pressure histogram alone, the domain scientist indicated that they like the ability to overlay secondary attributes on the data. The derived properties allowed them to visualize the interaction between pressure and temperature. By seeing these properties matched together, the scientist is able to relate the transfer function design to the physical interaction of properties and found this to be a natural means of analysis.

5.3 Abstracting Attributes in Noisy Data

The final analysis we performed was to explore the effect that noisy data would have on our abstracted attribute derivations. For this, we created a $64 \times 64 \times 64$ cube volumetric data set. The data set contained two small spheres; however, the voxels were distributed within the data set such that the overall histogram distribution of the volume density would be a uniform random distribution. The synthetic data set is shown in Fig. 11.

Fig. 11 explores the derived entropy, standard deviation, and skewness measures of the x, y, and z spatial attributes of our synthetic data set. As expected, the resultant entropy calculations showed a constant high entropy value (the entropy of the uniform random distribution function will be the maximum value entropy can obtain). The standard deviation and skewness also showed properties associated with a uniform random distribution, with the standard deviation being a high constant and the skewness being zero (as the data are symmetrically distributed). Based on these results, the abstraction of the spatial volumetric attributes provides little insight into the data; however, this is as expected given that the underlying data should approximate a uniform random distribution.

Of interest, though, was the 2D histogram showing the density versus density gradient values shown in Fig. 11.

When using the standard deviation and the x, y, and z directional components scaled by the view direction, we are able to find small areas in the histogram with a very low standard deviation. From previous examples, areas of low standard deviation represent compact structure within the volume. By selecting these areas, a reasonable rendering of the data can be produced.

Clearly, the application of the abstracted attribute space is sensitive to noise; however, this example was provided to highlight the worst case scenario. As the amount of noise in the data set increases, the amount of information that can be extracted decreases. Assume you have a volumetric feature, with property $f_k = i$, where i would be the histogram bin to which the N voxels making up this feature of interest map to, for example, a velocity, density, etc., that is known to map to a volumetric feature. If there exist M noisy voxels placed randomly in the volume that also have the property $f_k = i$, then this strategy fails if M is greater than $.5N$.

While the results of our abstraction were unsurprising for the uniform random distribution, we also chose to explore noisy CT data in order to evaluate the potential effect of noise. In Fig. 12, we explore a noisy CT foot data set. From our results, we see that the noise found in this data does little to affect the resulting attribute statistics. Moreover, using the spatial skewness close to zero in the 2D histogram (bottom) allows us to avoid choosing noise in the data compared to the traditional transfer function (top).

From these experiments, it is clear that noise will affect the underlying statistical calculations in the abstract attribute space. However, the degree to which this will affect the calculations is based on how the noise is distributed through the entire data set. Our calculations are performed for each bin in the histogram. For any bin in which the distribution is approximately normal, our calculations will return the expected values. However, this is not to say that such a calculation is meaningless. Our attribute space abstraction provides insight into another attribute distribution with respect to value ranges in the primary attribute being analyzed. In this manner, insight
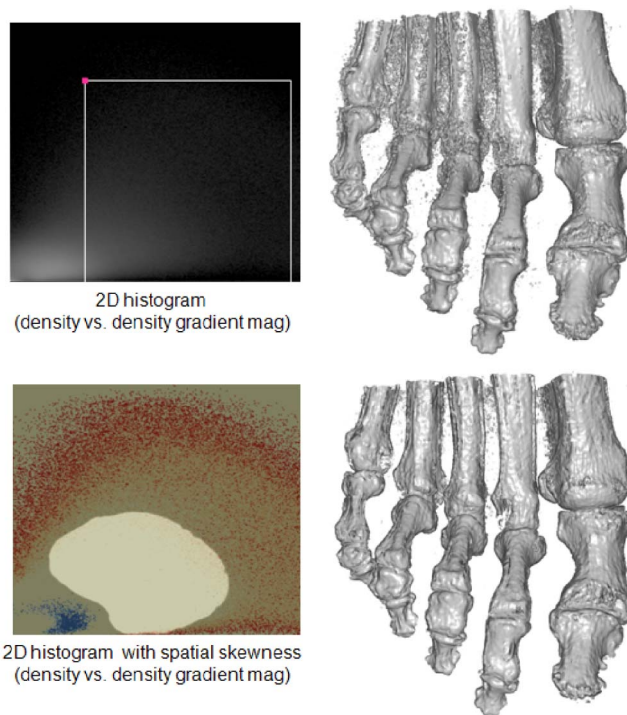
Fig. 12. Comparing explorations of the noisy CT feet data using a traditional 2D histogram widget and the spatial skewness abstracted 2D histogram widget. Here, the user has used a box transfer function on the traditional 2D transfer function (Top), whereas the user has brushed the zero skewness to avoid the noise in the data (Bottom).

can be gleaned from the data as shown in the previous sections.

6 CONCLUSIONS AND FUTURE WORK

Traditionally, the appropriate selection of attributes in multidimensional transfer functions is a difficult task, often requiring the user to have an underlying knowledge of the data set under exploration. By providing users with enhanced information about other attributes of their data in an abstracted attribute space (e.g., density, value versus value gradient magnitude) we are able to enhance the exploration, allowing users to better discover attributes within their data set. Users are able to quickly visualize and explore relationships within a given attribute space, enhancing their knowledge about interactions between the attribute space variables.

Our interactions with domain experts showed that these users tended to favor the mean and standard deviation plots as they were more comfortable with those metrics. Domain experts indicated that they liked the overlay of the abstracted attribute space, and that the added plot enabled them to think about transfer function design as more than a selection of value ranges, but also as a selection of variable interaction ranges. Experiences with the various data sets indicates that entropy was the least used statistical measure for exploration. Skeweness and standard deviation were useful for showing boundary areas and changes in data ranges, while the mean value was useful for direct comparison between two attributes. Often, the mean value calculation was used by domain experts in comparing velocity and pressure, water content and water vapor, etc. Future work will focus on deriving other information

metrics including interaction information and cross entropy for measuring attribute complexity. Furthermore, we plan to utilize abstracted attribute space measures for novel volume rendering parameters in order to reduce the burden of transfer function design on the user.

ACKNOWLEDGMENTS

REFERENCES

[1] T. He, L. Hong, A. Kaufman, and H. Pfister, "Generation of Transfer Functions with Stochastic Search Techniques," *Proc. IEEE Conf. Visualization,* pp. 227-234, 1996.

[2] J. Marks, B. Andalman, P.A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber, "Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation," *Proc. ACM SIGGRAPH Conf. Computer Graphics and Interactive Techniques,* pp. 389-400, 1997.

[3] J. Kniss, G. Kindlmann, and C. Hansen, "Interactive Volume Rendering Using Multi-Dimensional Transfer Functions and Direct Manipulation Widgets," *Proc. IEEE Conf. Visualization,* pp. 255-262, 2001.

[4] D.C. Banks and B.A. Singer, "Vortex Tubes in Turbulent Flows: Identification, Representation, Reconstruction," *Proc. IEEE Conf. Visualization (VIS '94),* pp. 132-139, 1994.

[5] K.-L. Ma, J. Van Rosendale, and W. Vermeer, "3D Shock Wave Visualization on Unstructured Grids," *Proc. Symp. vol. Visualization,* pp. 87-95, 1996.

[6] G. Kindlmann and J.W. Durkin, "Semi-Automatic Generation of Transfer Functions for Direct Volume Rendering," *Proc. IEEE Symp. Vol. Visualization,* pp. 79-86, 1998.

[7] M. Levoy, "Display of Surfaces from Volume Data," *IEEE Computer Graphics and Applications,* vol. CGA-8, no. 3, pp. 29-37, May 1988.

[8] C.L. Bajaj, V. Pascucci, and D.R. Schikore, "The Contour Spectrum," *Proc. IEEE Conf. Visualization,* pp. 167-173, 1997.

[9] G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Möller, "Curvature-Based Transfer Functions for Direct Volume Rendering: Methods and Applications," *Proc. IEEE Conf. Visualization,* pp. 513-520, 2003.

[10] F.-Y. Tzeng, E.B. Lum, and K.-L. Ma, "A Novel Interface for Higher-Dimensional Classification of Volume Data," *Proc. IEEE Conf. Visualization,* pp. 505-512, 2003.

[11] S. Potts and T. Möller, "Transfer Functions on a Logarithmic Scale for Volume Rendering," *Proc. Graphics Interface,* pp. 57-63, 2004.

[12] C. Lundström, P. Ljung, and A. Ynnerman, "Multi-Dimensional Transfer Function Design Using Sorted Histograms," *Proc. Eurographics/IEEE VGTC Workshop Vol. Graphics,* pp. 1-8, July 2006.

[13] C. Lundström, P. Ljung, and A. Ynnerman, "Local Histograms for Design of Transfer Functions in Direct Volume Rendering," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 6, pp. 1570-1579, Nov./Dec. 2006.

[14] C. Lundström, A. Ynnerman, P. Ljung, A. Persson, and H. Knutsson, "The Alpha-Histogram: Using Spatial Coherence to Enhance Histograms and Transfer Function Design," *Proc. Eurographics/IEEE-VGTC Symp. Visualization,* pp. 227-234, May 2006.

[15] C. Correa and K.-L. Ma, "Size-Based Transfer Functions: A New Volume Exploration Technique," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1380-1387, Oct. 2008.

[16] C. Correa and K.-L. Ma, "Visibility-Driven Transfer Functions," *Proc. IEEE-VGTC Pacific Visualization Symp.,* Apr. 2009.

[17] R. Maciejewski, I. Woo, W. Chen, and D.S. Ebert, "Structuring Feature Space: A Non-Parametric Method for Volumetric Transfer Function Generation," *IEEE Trans. Visualization and Computer Graphics,* vol. 15, no. 6, pp. 1473-1480, Nov./Dec. 2009.

[18] S. Bruckner and T. Möller, "Isosurface Similarity Maps," *Computer Graphics Forum,* vol. 29, no. 3, pp. 773-782, 2010.

[19] A. Shamir, "Feature-Space Analysis of Unstructured Meshes," *Proc. IEEE Conf. Visualization,* pp. 185-192, 2003.

[20] S. Roettger, M. Bauer, and M. Stamminger, "Spatialized Transfer Functions," *Proc. Eurographics/IEEE-VGTC Symp. Visualization,* pp. 271-278, 2005.

[21] C. Henze, "Feature Detection in Linked Derived Spaces," *Proc. Conf. Visualization (VIS '98),* pp. 87-94, 1998.

[22] J. Woodring and H.-W. Shen, "Multi-Variate, Time Varying, and Comparative Visualization with Contextual Cues," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 909-916, Sept./Oct. 2006.

[23] H. Akiba and K.-L. Ma, "A Tri-Space Visualization Interface for Analyzing Time-Varying Multivariate Volume Data," *Proc. Eurographics/IEEE VGTC Symp. Visualization,* pp. 115-122, May 2007.

[24] H. Akiba, K.-L. Ma, J.H. Chen, and E.R. Hawkes, "Visualizing Multivariate Volume Data from Turbulent Combustion Simulations," *Computing in Science and Eng.,* vol. 9, no. 2, pp. 76-83, 2007.

[25] C. Muelder and K.-L. Ma, "Interactive Feature Extraction and Tracking by Utilizing Region Coherency," *Proc. IEEE Pacific Visualization Symp.,* pp. 17-24, Apr. 2009.

[26] Z. Fang, T. Möller, G. Hamarneh, and A. Celler, "Visualization and Exploration of Time-Varying Medical Image Data Sets," *Proc. Graphics Interface (GI '07),* pp. 281-288, 2007.

[27] H. Jänicke, A. Wiebel, G. Scheuermann, and W. Kollmann, "Multifield Visualization Using Local Statistical Complexity," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1384-1391, Nov./Dec. 2007.

[28] M. Haidacher, S. Bruckner, A. Kanitsar, and M.E. Gröller, "Information-Based Transfer Functions for Multimodal Visualization," *Proc. Visual Computing for Biology and Medicine,* pp. 101-108, Oct. 2008.

[29] *Dynamic Graphics for Statistics,* Statistics/Probability Series, W.S. Cleveland and M.E. McGill, eds. Wadsworth & Brooks/ Cole, 1998.

[30] *The Visual Display of Quantitative Information,* E.R. Tufte, ed. Graphics Press, 1993.

[31] C. Stolte, D. Tang, and P. Hanrahan, "Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases," *IEEE Trans. Visualization and Computer Graphics,* vol. 8, no. 1, pp. 52-65, Jan.-Mar. 2002.

[32] D.F. Swayne, D.T. Lang, A. Buja, and D. Cook, "GGobi: Evolving from XGobi into an Extensible Framework for Interactive Data Visualization," *Computational Statistics and Data Analysis,* vol. 43, no. 4, pp. 423-444, 2003.

[33] M.O. Ward, "XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data," *Proc. IEEE Conf. Visualization (VIS '94),* pp. 326-333, 1994.

[34] N. Elmqvist, P. Dragicevic, and J.-D. Fekete, "Rolling the Dice: Multidimensional Visual Exploration Using Scatterplot Matrix Navigation," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, pp. 1141-1148, Nov./Dec. 2008.

[35] C. Wang, H. Yu, and K.-L. Ma, "Importance-Driven Time-Varying Data Visualization," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1547-1554, Nov./Dec. 2008.

[36] J. Kehrer, P. Filzmoser, and H. Hauser, "Brushing Moments in Interactive Visual Analysis," *Computer Graphics Forum,* vol. 29, no. 3, pp. 813-822, 2010.

[37] M. Haidacher, D. Patel, S. Bruckner, A. Kanitsar, and M.E. Gröller, "Volume Visualization Based on Statistical Transfer-Function Spaces," *Proc. IEEE Pacific Visualization Symp. '10,* pp. 17-24, 2010.

[38] S. Bachthaler and D. Weiskopf, "Continuous Scatterplots," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1428-1435, Nov./Dec. 2008.

[39] S. Bachthaler and D. Weiskopf, "Efficient and Adpative Rendering of 2-D Continuous Scatterplots," *Computer Graphics Forum,* vol. 28, no. 3, pp. 743-750, 2009.

[40] H. Carr, B. Duffy, and B. Denby, "On Histogram and Isosurface Statistics," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 1259-1265, Sept./Oct. 2006.

[41] C. Scheidegger, J. Schreiner, B. Duffy, H. Carr, and C. Silva, "Revisiting Histograms and Isosurface Statistics," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1659-1666, Nov./Dec. 2008.

[42] D. Patel, M. Haidacher, J.-P. Balabanian, and M.E. Gröller, "Moment Curves," *Proc. IEEE Pacific Visualization Symp. '09,* pp. 201-208, Apr. 2009.

[43] C.S. Co, A. Friedman, D.P. Grote, J.-L. Vay, E.W. Bethel, and K.I. Joy, "Interactive Methods for Exploring Particle Simulation," *Proc. Eurographics/IEEE-VGTC Symp. Visualization '05,* pp. 279-286, 2005.

[44] D. Silver and X. Wang, "Tracking and Visualizing Turbulent 3D Features," *IEEE Trans. Visualization and Computer Graphics,* vol. 3, no. 2, pp. 129-141, Apr.-June 1997.

[45] J. Woodring, C. Wang, and H.-W. Shen, "High Dimensional Direct Rendering of Time-Varying Volumetric Data," *Proc. IEEE 14th Visualization (VIS '03),* p. 55, 2003.

[46] N.A. Svakhine, Y. Jang, D.S. Ebert, and K.P. Gaither, "Illustration and Photography Inspired Visualization of Flows and Volumes," *Proc. IEEE Conf. Visualization,* pp. 687-694, 2005.

Ross Maciejewski received the PhD degree in electrical and computer engineering from Purdue University in December, 2009. He is currently an assistant professor at Arizona State University in the School of Computing, Informatics & Decision Systems Engineering. Prior to this, he served as a visiting assistant professor at Purdue University and worked at the Department of Homeland Security Center of Excellence for Command Control and Interoperability in the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) group. His research interests are geovisualization, visual analytics and nonphotorealistic rendering. He is a member of the IEEE and the IEEE Computer Society.

Yun Jang received the bachelor's degree in electrical engineering from Seoul National University, South Korea, in 2000, the master's and doctoral degrees in electrical and computer engineering from Purdue University in 2002 and 2007, respectively. He is an assistant professor of computer engineering at Sejong University, Seoul, South Korea. He was a postdoctoral researcher at CSCS and ETH Zurich, Switzerland from 2007-2011. His research interests include interactive visualization, volume rendering, and data representations with functions.

Insoo Woo received the BS degree in computer engineering in 1998 from Dong-A University in South Korea and working toward the PhD degree in the School of Electrical and Computer Engineering at Purdue University. He is a research assistant in Purdue University Rendering and Perception Lab. He was employed as a software engineer during 1997 to 2006. His research interest is GPU-aided Techniques for Computer Graphics and Visualization.

Heike Jänicke received the MS degree (Diplom) in computer science with a special focus on medical computer science in 2006 from Leipzig University. For her research of information-theoretic methods in visualization, she was awarded the PhD degree in 2009 from the same university. During 2009 and 2010, she worked as a postdoctoral researcher at Swansea University. Since March 2010, she is working as a junior professor for computer graphics and visualization at Heidelberg University, Germany. Her research interests include data analysis methods from statistics and information theory, visualization of unsteady, multivariate data with applications in biology, medicine, astronomy, and climate research. She is a member of the IEEE.

Kelly P. Gaither received the bachelor's and master's degrees in computer science from Texas A&M University in 1988, and 1992, respectively, and the doctoral degree in computational engineering from Mississippi State University in May, 2000. While obtaining the PhD degree, she worked full time at the Simulation and Design Center in the National Science Foundation Engineering Research Center as the leader of the visualization group. She is the director of Data & Information Analysis at the Texas Advanced Computing Center (TACC) and leads the scientific visualization, data management & collections, and data mining & statistics programs at TACC while conducting research in scientific visualization and data analysis. She, a research scientist, also serves as the area director for visualization in the National Science Foundation funded TeraGrid project. She has a number of refereed publications in fields ranging from Computational Mechanics to Supercomputing Applications to Scientific Visualization. She has given a number of invited talks. Over the past 10 years, she has actively participated in the IEEE Visualization conference, and served as the IEEE Visualization conference general chair in 2004. She is currently serving on the IEEE Visualization and Graphics Technical Committee. She is a member of the IEEE and the IEEE Computer Society.

David S. Ebert received the PhD degree in computer science from Ohio State University. He is a professor in the School of Electrical and Computer Engineering at Purdue University, a University Faculty scholar, director of the Purdue University Rendering and Perceptualization Lab, and director of the Purdue University Regional Visualization and Analytics Center. His research interests include novel visualization techniques, visual analytics, volume rendering, information visualization, perceptually based visualization, illustrative visualization, and procedural abstraction of complex, massive data. He is a fellow of the IEEE and the IEEE Computer Society, and a member of the IEEE Computer Society's Publications Board.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

Applied visual analytics for exploring the National Health and Nutrition Examination Survey

Silvia Oliveros Torres*, Heather Eicher-Miller+, Carol Boushey+, David Ebert*, Ross Maciejewski*

* Purdue University Regional Visualization and Analytics Center, + Dept. of Nutrition, Purdue University

Abstract

The National Health and Nutrition Examination Survey (NHANES) is a research program to assess the health and nutritional status of the population in the United States. In this work, we present a visual analytics system designed to help researchers explore patterns and form hypotheses within the NHANES dataset. The visualization component of the environment is an extension of traditional scatterplot matrices. Since the upper portion of the scatterplot matrix is a redundant encoding, we utilize this space, to show the projected N-dimensional clustering of points. The rows and columns of the matrix are automatically ordered using information about the cluster projection in each space as a means of showing the most meaningful dimensions. A comparison module has also been included that allows the user to compare groupings of people to the 2010 Dietary Guidelines for Americans. This tool enhances the analysis work by aiding discovery and hypothesis formation.

Keywords: visual analytics, scatterplot, NHANES, diet records.

1. Introduction

The United States health sector has deployed many survey programs that produce large datasets with increasing complexity and dimensionality. One such survey program is the National Health and Nutrition Examination Survey (NHANES) [1], which is a population-based survey designed to collect information on the health and nutrition of the U.S. household population. NHANES collects data from physical examinations along with surveys where they ask the responder to recall their ingestion of food for the past 48 hours. The 2-day food recollection survey includes demographic questions such as gender, age and race/ethnicity.

Visual data exploration techniques have been shown to be an effective tool in aiding analysts in exploring and understanding these types of large, multivariate datasets. Our main contribution in this paper is the development of a visual analytics system for NHANES. The system will help researchers explore patterns, form hypotheses, understand the underlying structure of the dataset, and will also provide the researchers with means of presenting their findings. The visual analytics system relies on an interactive scatterplot matrix to visualize the different dimensions of the data set. The scatterplot matrix was chosen because it provides the user with an easy way to interpret relationships between different pairs of dimensions within the data. Distinguishing correlation between variables allows the user to understand how one variable affects the other. Additionally, we have incorporated advanced analytics tools for exploring this scatterplot matrix, including clustering and dimensional ordering that provides a more guided exploration of this large dataset. Our primary analytic tool is clustering. Cluster analysis is a powerful tool for the exploration of high dimensional data. Clustering can be used to discover hidden associations without a prior hypothesis, therefore, we included automatic clustering via k-means [11]. We have also included two different types of automatic dimension ordering based on cluster density and dimension similarity. Dimensional ordering has proven to reduce the clutter that obscures the underlying structures in high dimensional data [5]. By automatically ordering the way the scatterplots are presented, we can enhance the process of exploring the data for the users by ordering the data by its information content.

Along with automatic analysis, the user has the ability to filter the data depending on the problem being analyzed. The system provides users with the ability to filter the data based on age, gender, and ethnicity of the participants. The user also has full control over the number of dimensions displayed in the scatterplot matrix, providing both global and local analysis options. The size of the matrix can be

increased or decreased, and users can modify which dimensions are being shown. All these features provide the user with a fully customizable experience while navigating and analyzing the NHANES dataset.

Finally, the system also includes a familiar graphing/comparison model to aid the user in understanding the overall dataset while focusing on a specific variable. The system transforms the data points presented in the scatterplot matrix into bar graphs that are similar in style to the ones presented in publications by the U.S. Department of Health and Human Services [10]. These bar graphs can also be used to present the findings of the data exploration. Feedback from nutrition experts indicated that researchers would be familiar with these graphs, and find them to be easily understood.

The remainder of this paper is organized as follows. Section 2 provides a summary of related work. Section 3 provides an overview of the structure of the NHANES dataset. Section 4 discusses the visual analytics environment and its principal components such as the scatterplot matrix, clustering and dimension ordering. Section 5 introduces the graphing/comparison model being used with this dataset. Section 6 provides the reader with a usage scenario. Finally, conclusions and future work are discussed in Section 7.

2. Related work

Many multi-dimensional visualization tools exist that utilize scatterplots including XmdvTool [2], which supports many interaction modes and tools, and Polaris [3], which takes a database and projects the data into a scatterplot matrix. Other tools, such as ScatterDice[13], have also included interactive techniques for navigation within the scatterplot matrix and methods for dimension reordering designed to show correlation and differences between individual dimensions. Our work integrates some of the features seen in these previous tools, such as, brushing, linking, zooming, panning, and reordering of dimensions. We extend these features through the addition of clustering into a new tool to visualize this specific multidimensional dataset.

Furthermore our work also expands approaches for ordering and filtering the dimensions of multi-dimensional datasets. Ankerst et al. [4] explored a variety of clutter reduction metrics, along with some work in dimension reduction. Ankerst et al. proposed a method for arranging dimensions using pair wise similarity measures that are used to calculate a global optimization method. The similarity measure in this work was based on the Euclidean distance function

proposed by Ankerst et al. For one of the dimension reordering methods used in our system, the notion of a pair wise correlation is used to compute final scatterplot arrangement.

Recent work by Peng et al. [5], shows that by reordering the dimensions, clutter in a representation can be reduced without reducing the information content. Clutter is considered to be anything that interferes with the process of finding structures. For scatterplot displays Peng et al., proposed arranging the matrix based on scatterplot cardinality. For the high cardinality dimensions, the Pearson correlation coefficient is used to calculate the clutter measure and re-arrange accordingly. [We incorporate Peng et al.'s concept of using the Pearson correlation coefficient as one way to re-arrange the scatterplots in our matrix.] Tatu et al. [8] presented a ranking method for scatterplots that uses rotating variance, class density, and histogram density measures. We use their notion of density and apply it to the clustering of data points within the scatterplots.

Yang et al. [6,7] presented a framework that provides a multi-resolution view of the data via hierarchical clustering where the user can interactively explore the desired focus region at different levels of detail. The framework works upon hierarchical cluster trees of data sets, and makes use of proximity-based color to help identify and link relationships between clusters. Our system uses some of the clustering concepts and interactive exploration without the dimension reduction that Yang presented.

Sips et al. [9] presented the concept of class consistency in classes of n-Data shown in 2D scatterplots. Their quantitative methods of consistency are based on clusters center of gravity and on entropies of the spatial distributions of classes. We did not compute any measure of class consistency, nor do we have a method to qualify the placement of the scatterplots. However it is a topic that we are interested in and will leave this as part of the system's future work.

3. Structure of NHANES Data

The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by medical personnel. In this study, we focus exclusively on the dietary survey information collected.

NHANES collects data from a 2-day food recollection survey. The survey collects the types and
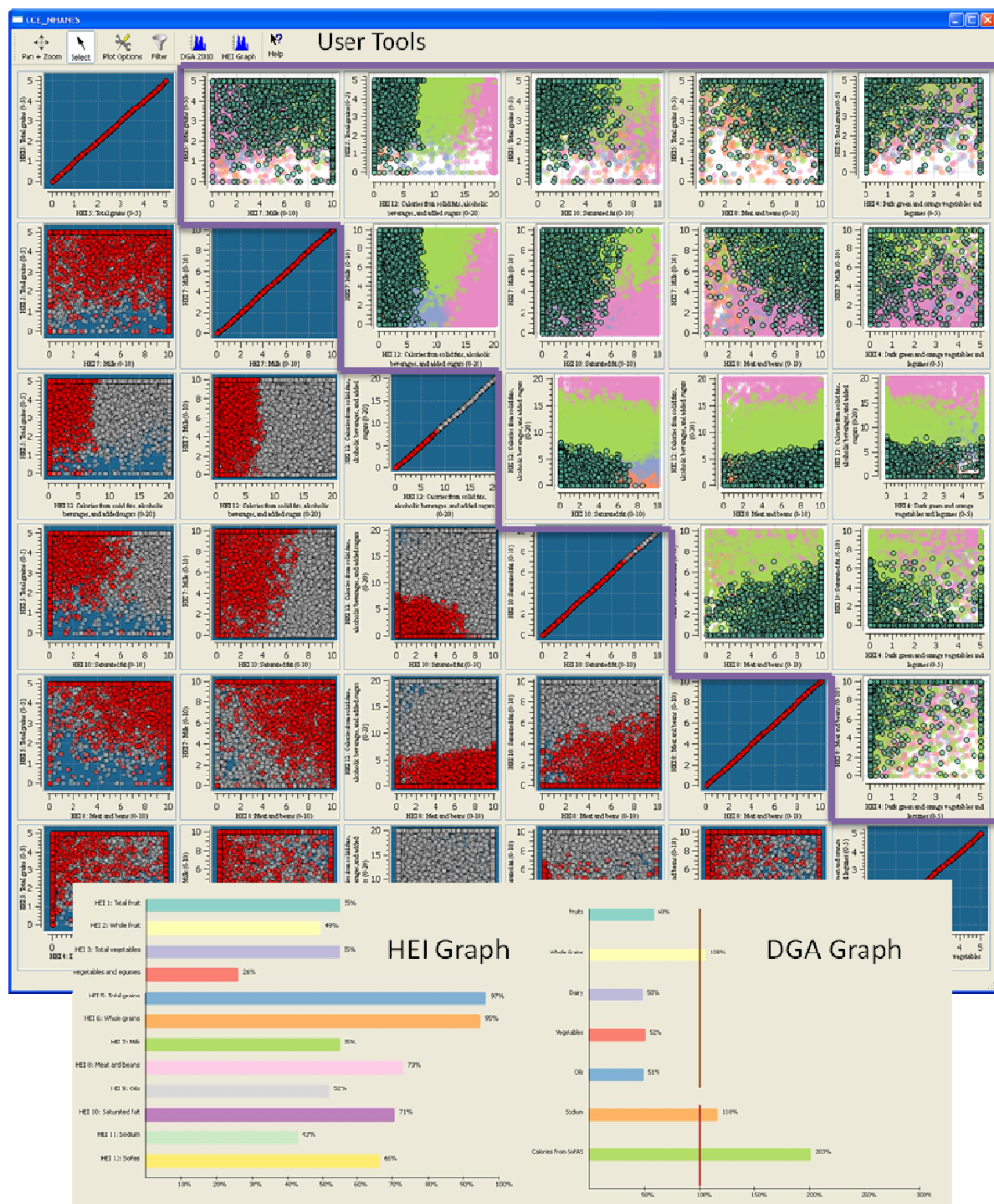
Upper portion shows k-means clustering

Figure 1. General overview of the visual analytics system for NHANES exploration. One of the clusters has been selected in the upper portion and all the participants that belong to that cluster have been highlighted across the matrix.

quantities of food ingested during 48 hours. After the survey, health experts separate the food into different groups and calculate their caloric and nutrient components.

The nutritional content is classified into 12 different Healthy Eating Indexes (HEI) as described

in Table 1. Each component of the index has different maximum scores and a minimum score of zero. High component scores indicate intakes close to the recommend ranges or amounts; low component scores indicate less compliance with the recommended ranges or amounts. Low compliance

means a HEI score of zero. The HEI scores indicate levels, thus the scales are not normalized. Finally, all the components can be compounded and an overall HEI score is calculated and viewed as a measure of diet quality that assesses conformance to the federal dietary guidance [10].

HEI	Description	Range
1	Total Fruit	0-5
2	Whole Fruit	0-5
3	Total Vegetables	0-5
4	Dark green and orange vegetables and legumes	0-5
5	Total grains	0-5
6	Whole grains	0-5
7	Milk	0-10
8	Meat and beans	0-10
9	Oils	0-10
10	Saturated fat	0-10
11	Sodium	0-10
12	Calories from solid fats, alcoholic beverages, and added sugars	0-20

Table 1. HEI Components

4. Visual analytics environment

The visual analytics environment consists of a scatterplot matrix of selected variables that can be modified at any time. The upper diagonal portion of the matrix displays the same data as the lower diagonal but in a k-means clustered form as shown in Figure 1. The initial dimensions chosen are the result of dimension ordering. This ordering can be implemented using two different methods: one based on cluster density and the other on dimension correlation. The different user tools implemented in the system are described at the end of this Section.

4.1 Basic scatterplot matrix

The main component of the application consists of a traditional scatterplot matrix set up initially as a 6x6 grid of default variables as shown in Figure 1. These default variables are determined based on two different methods of dimension ordering that are discussed in Section 4.3. In Figure 1, the variables are HEI 5, 7, 12, 10, 8, and 4. One of the main concerns while designing the system was the user. We do not want the user to become overwhelmed by the high dimensionality of the dataset. Instead, we want the user to be able to focus on a given scatterplot and explore it freely and completely. With

this in mind, we set the grid to be 6x6 which is half of the maximum number of HEI variables. The initial grid size was determined based on aesthetics. The chosen size displays all the necessary labels and fits multiple monitor sizes. As mentioned before, the user can choose to increase or decrease the number of dimensions shown at any time as he/she sees appropriate.

The matrix is divided diagonally into an upper and lower portion as shown in Figure 1. The lower diagonal portion shows the regular scatterplots of the components plotted against each other. The upper diagonal portion shows the clusters projected into each of the plots. In Figure 1, one of the clusters has been selected and all of the participants who belong to that cluster are highlighted across the entire matrix. We discuss the upper diagonal of the scatterplot matrix below.

4.2 Upper diagonal clustering

Clustering is one of the most frequently used methods to improve the perception and recognition of patterns in multivariate datasets. In our system the k-means [11] clustering algorithm is applied and a Euclidean distance metric between the points in the N-dimensional space is used.

In k-means, we are given a set of n data points in a d-dimensional space (in the NHANES case d=12) R^d , and an integer k. The problem is to determine a set of k points in R^d, called centers, which will minimize the mean square distance from each point to its nearest center. Our k-means clustering was computed using an enhanced form of a kd tree as specified by Kanugo et al. [11].

The clusters computed at the beginning of the program are projected into each of the 2D plots on the upper diagonal matrix. Each cluster is differentiated with the aid of different colors and occlusion is addressed by using transparency when we plot the different points. We revised the work of Chiang and Mirkin [14] about an intelligent choice of the number of clusters which described Hartigan's rule as one of the best methods to calculate the initial number of clusters. Hartigan's rule suggests using 25 different clusters for this specific dataset, however the initial number was reduced in order to not overwhelm the users and distract them from seeing general trends. A large number of clusters can compromise the legibility of the scatterplots. We leave the decision of the number of clusters to the user, who can modify it at any time as part of the exploration process.

Although there is no guarantee that the 2D projections will show a clear distinction between the
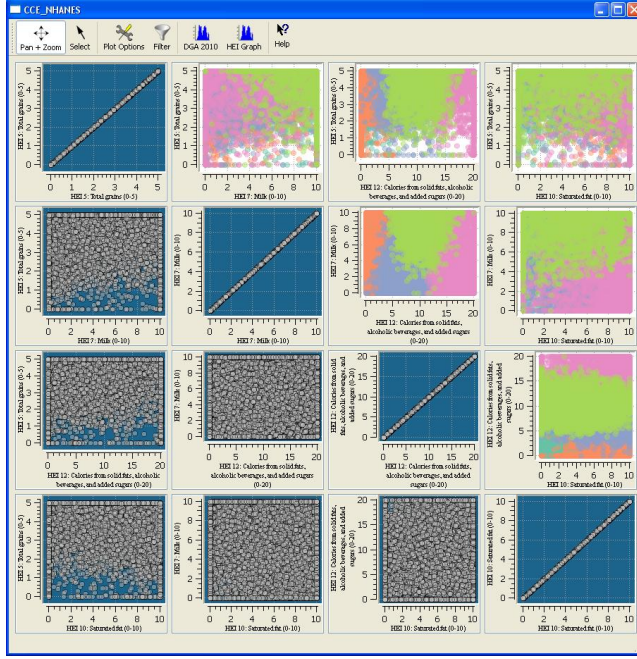
(a)
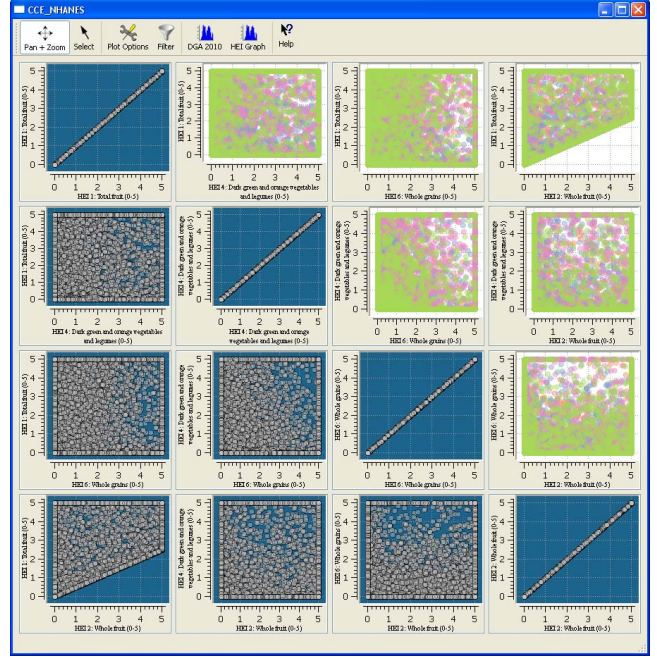
(b)

Figure 2. (a) Cluster density ordering. (b) Pearson's correlation ordering.

clusters, the dimension ordering described in the following subsection can help in further separating the clusters. These computed clusters also provide a reference group of participants, which can be further explored with the aid of brushing and linking.

4.3 Dimension ordering

Our system automatically generates a default ordering scheme that the user can modify later. There are two different measures employed for reordering the scatterplot matrix.

The first measure arranges the dimensions according to their cluster density. This measure simply allows the user to view the scatterplots that present the most defined clusters calculated with k-means as explained in the previous section. Cluster density may be defined on a local or global scale within the N-dimensional space.

Global cluster density is a comparison across all the N-dimensions. Local cluster density compares only the projected dimensions for any given scatterplot. The cluster density, the variance, and standard deviation for each cluster are computed using Welford's method [12].

Traditionally, the mathematical formula for computing the sample variance is:

$$s^2 = \frac{1}{n(n-1)}\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)$$

where s represents the variance, n is the size of the sample, and x is the data point being used.

Welford's method computes the variance as the x's arrive, one at a time, and gives more arithmetic precision. Once the algorithm is initialized with $M1 = x1 \; and \; S1 = 0$ the subsequent x's use the recurrence formulas:

$$Mk = Mk - 1 + (xk - Mk - 1)/k$$
$$Sk = Sk - 1 + (xk - Mk - 1) * (xk - Mk).$$

Welford's method is more efficient because it requires looping through all the data points only once. Welford's method is also more accurate since it does not use abnormally large numbers in calculation, compared to the sum of squares method.

Using these global metrics, we order the scatterplots such that the dimensions with the tightest clusters will be displayed to the left of the dimensional ordering lineup. These plots should potentially contain the most relative information as the projected clusters are the most compact in this space. The local scale is then used during the rendering of the clusters within a given 2D plot. Clusters are rendered in order of their local density, from smallest to largest. By rendering the clusters in

this manner, we also reduce some of the issues in over-plotting.

The second measure is based on Pearson's correlation coefficient, which is defined to be the measure of the correlation (linear dependence) between two variables. The formula used to calculate the coefficient is the following:

$$r = \frac{\sum_{i-1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y}$$

In the equation above the variables are represented as X and Y, with means $\bar{X}$ and $\bar{Y}$ respectively, and standard deviations, S_X and S_Y respectively.

The coefficient is calculated for each plot individually and plots are arranged based on their coefficients. This measure also allows the scatterplots with the most similar features to be plotted close to each other.

In Figure 2, we show how the two different methods reordered the given dimensions in a 4x4 matrix. Figure 2(a) shows the re-ordering based on cluster density. Since it presents the scatterplots with the most compact clusters, the user can visually identify the different clusters in some of the scatterplots as clearly defined color areas. Figure 2(b) shows the re-arrangement based on Pearson's correlation. In Figure 2(b), the clusters do not play an important role when the correlation is being computed. In Figure 2(a) cluster density shows defined clusters in the upper diagonal portion, whereas in Figure 2(b), Pearson's correlation shows more defined areas in the bottom diagonal portion.

4.4 User tools and interaction

Besides the normal interactive capabilities such as zooming and panning, our system also includes a selection tool that allows the user to choose a subset of specific participants in the dataset that will be highlighted across the scatterplot matrix. The selection tool allows for regular brushing and linking.
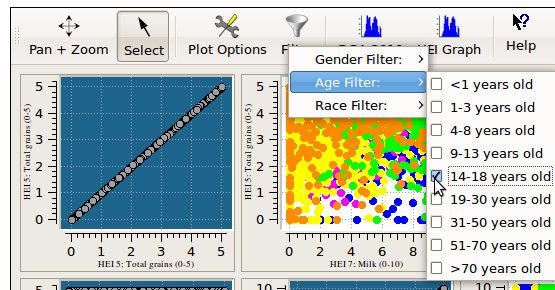
Figure 3. Filtering tool.

Highlighted participants in a scatterplot are shown as red points and this provides an easy method to observe these participants behavior across all the other categories. The highlighting ability across the matrix allows for a quick discovery of relationships between different components. Finally, the user has the ability of further narrowing down the dataset by filtering the population being displayed in the scatterplots. There are filters available based on age, gender, race and ethnicity and calculated body mass index (BMI).

5. Comparison graphs

Every five years, the Dietary Guidelines for Americans (DGA) are published jointly by the U.S. Department of Health and Human Services (HHS) and the U.S. Department of Agriculture (USDA) [10] with the last one being published in 2010. These guidelines provide authoritative advice for Americans, ages two and older, about consuming fewer calories, making informed food choices, and being physically active to attain and maintain a healthy weight, reduce risk of chronic disease, and promote overall health. Since these guidelines are followed and understood by most of the health sector in the country, our system has the ability to generate two different graphs to transform the NHANES data into familiar graphs and standards presented in the DGA.
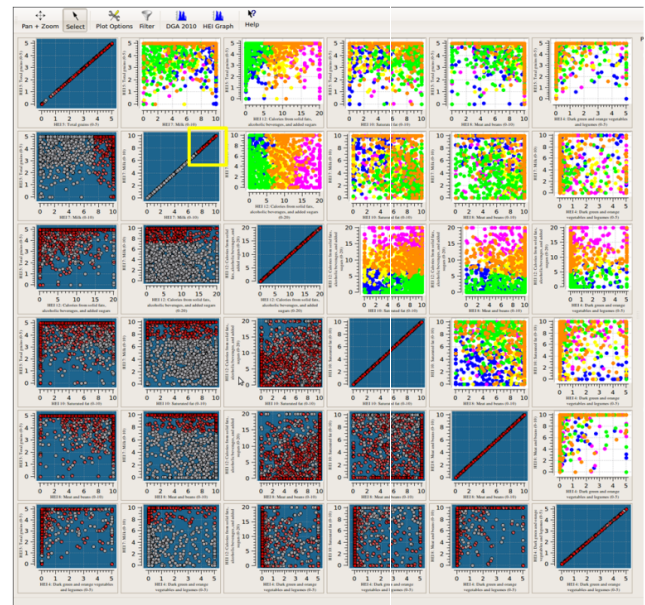
Figure 4. Scatterplot matrix highlights the participants with high dairy consumption in red.
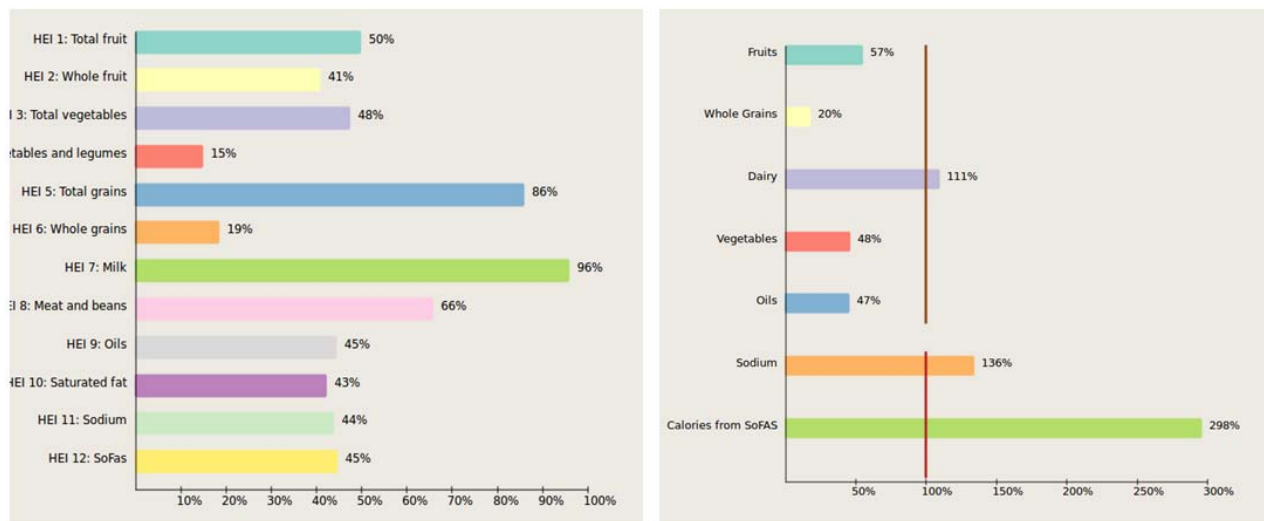
Figure 5. Comparison graphs for participants with high milk consumption (HEI 7). Hei graph on the left and DGA on the right.

Figure 6. Comparison graphs for participants with high vegetable and legume consumption. (HEI 4)

Using the selection tool described in the previous section, the user can select a region, or any given cluster, to update the two given graphs. The first graph transforms the selected points into the most representative DGA graph. The graph adds up all the selected participants data, averages it and then converts it into a percentage based on the guideline for consumption. The bottom of Figure 1 presents the newly created graphs at the bottom. The DGA graph on the right includes two separate markers, the top one indicates the amount of foods that the Americans needs to consume in order to fulfill the 100% recommended daily intake. The bottom marker indicates the foods that need to be consumed in moderation, showing a red line when the participants have consumed more than the recommended intake.

The left side graph at the bottom of Figure 1 provides the user with an overview of all the HEI within the selected group. Since the user might only be looking at fewer scatterplots, the summary graph provides insight into which HEI should be observed next. A 100% score in this graph indicates the maximum score given in Table 1 for any given HEI. For example, if a participant's HEI 5 score is 4.5 the percentage shown in the graph for HEI 5 will be 90%. The percentage HEI graph was implemented for the purpose of representing all HEI scores in a common scale.
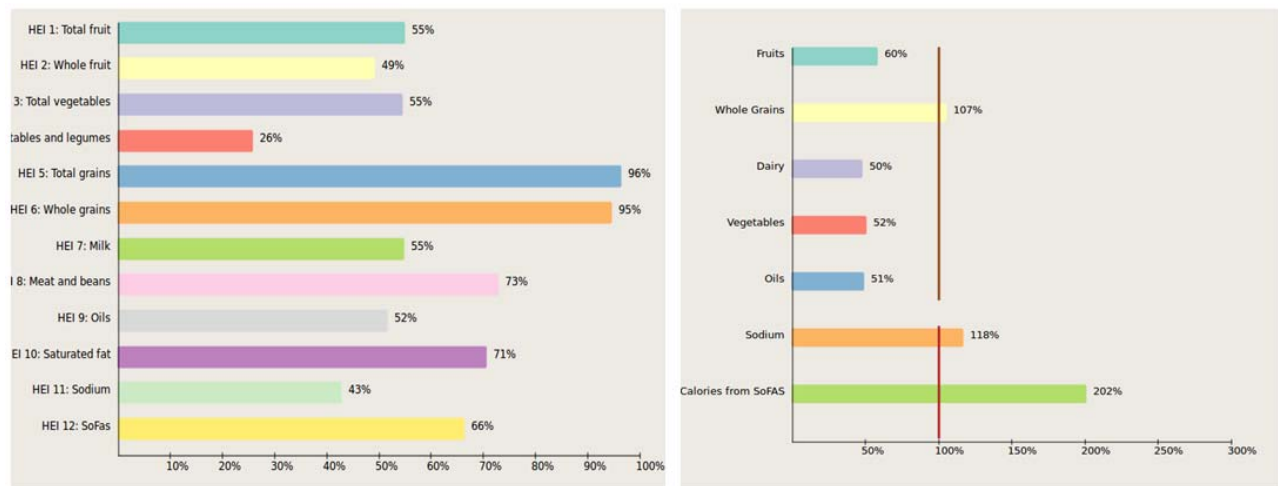
Figure 7. Comparison graphs for participants with whole grains consumption. (HEI 6)

6. Usage scenario

We begin our usage scenario with a 6x6 scatterplot matrix that shows the plots arranged based on higher cluster density. First, we select only the population that we are interested in working with (i.e., males in the age group 14-18). After we have added this filter, the clusters are recomputed and the scatterplots are re-arranged (see Figures 3-4). Our goal is to determine which single HEI category might prove to be the best indicator of healthy eating habits.

Since HEI 7 (Milk) was one of the first two dimensions shown in the matrix, we make use of the scatterplots in the diagonal of the matrix and select the participants with a high HEI 7. We are able to select those participants by selecting the upper right corner of the diagonal line of HEI 7. After performing the selection we can see the adjacent plots being highlighted with the same participants as shown in Figure 4.

Most of the plots do not show any correlation with HEI 7 since all the points highlighted appear across all the scatterplots. However, we can also see the two updated graphs. With the aid of the HEI graph (upper graph in Figure 5) we can visualize the other dimensions that we do not see in the scatterplot matrix. We can see that teenagers who consume an adequate amount of milk do not necessarily consume an adequate amount of other foods; however, they do relatively well on consuming low quantities of sodium and. We also notice that participants who show a high dairy intake do not do perform well in HEI 4 (vegetables and legumes). Since we know that participants with high milk consumption have a low vegetable consumption, we now select patients with high consumption in HEI 4 by selecting the upper

diagonal on the HEI 4 vs HEI 4 scatterplot. The resulting graphs for this selection can be seen in Figure 6.

Based on the DGA generated graph on the left of Figure 6 we notice that these participants perform better in other categories except for whole grains, which seems to remain the same. This hypothesis is confirmed by the HEI graph that shows an increase in most of the categories. The remaining question is to examine the whole grain category since it seemed to remain constant during the last two examples.

This time around, we select participants with a high whole grain intake that is represented by HEI 6. After looking at the generated graphs (Figure 7), we can observe that these participants maintain a more balanced diet since their graphs demonstrate an overall increase as well as being more leveled. The participants also show a lower consumption of sodium and lower consumption of any extra added calories. Based on the past three examples, we hypothesize that the best indicator of a good diet in teenagers comes from their whole grains intake. Teenagers who consume a high quantity of whole grains may lead a more balanced diet compared to those who consume high quantities of milk or vegetables and legumes.

7. Conclusions and future Work

We have described a visual analytics system developed to explore the multidimensional NHANES dataset. The system makes use of an interactive scatterplot with automatic clustering that is rearranged to present the user with the variables that show the most correlation from the beginning. The system also displays familiar graphs that condense

the information presented to the user. The system aids the user in developing and testing new hypotheses. We showed a usage scenario that demonstrates the use of the comparison/graphing system implemented to navigate the data.

We recognize there is still work that could be done to improve the system. We would like to experiment with new layouts to present the dietary elements that seem to have the most impact in the participants diet, possibly a combination of the methods used in this paper, as well as an extension exploring dimensionality reduction. In order to further differentiate the clusters, we are planning on adding an interactive exploration of the scatterplot matrix. This visualization system could also be used to analyze other types of complex data sets. Finally, future work will include modifying the system to fit similar datasets.

8. References

[1] Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey: http://www.cdc.gov/nchs/nhanes.htm May 27, 2011 [May 30, 2011]

[2] Matthew O. Ward, "XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data," IEEE Conf. on Visualization '94, pp 326 - 333, Oct. 1994.

[3] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. IEEE Transactions on Visualization and Computer Graphics, 8(1):52–65, 2002.

[4] M. Ankerst, S. Berchtold, and D.A. Keim. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In Information Visualization, 1998. Proceedings. IEEE Symposium on, pages 52 –60, 153, October 1998.

[5] Wei Peng, Matthew O. Ward, and Elke A. Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In Proceedings of the IEEE Symposium on Information Visualization, pages 89–96, Washington, DC, USA, 2004. IEEE Computer Society.

[6] J. Yang, M. O. Ward, E. A. Rundensteiner, and S. Huang. Visual hierarchical dimension reduction for exploration of high dimensional datasets Eurographics / IEEE TCVG Symposium on Visualization, pages 19–28, May 2003.

[7] Jing Yang, Matthew O. Ward, and Elke A. Rundensteiner. Interactive hierarchical displays: a general framework for visualization and exploration of large multivariate data sets. Computers & Graphics, 27(2):265–283, April 2003.

[8] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, D. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST), Atlantic City, New Jersey, USA (2009)

[9] Mike Sips, Boris Neubert, John P. Lewis, and Pat Hanrahan. Selecting good views of high-dimensional data using class consistency. Computer Graphics Forum, 28(3):831–838, June 2009.

[10] U.S. Department of Agriculture and U.S. Department of Health and Human Services. Dietary Guidelines for Americans, 2010. 7th Edition, Washington, DC: U.S. Government Printing Office, December 2010.

[11] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. A local search approximation algorithm for k-means clustering. Comput. Geom. Theory Appl., 28:89–112, June 2004.

[12] Donald E. Knuth. The art of computer programming, volume 2 (3rd ed.): seminumerical algorithms. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.

[13] N. Elmqvist, P. Dragicevic, J.-D. Fekete. Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation. In IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2008), 14(6):1141-1148, 2008.

[14] Mark Ming-Tso Chiang and Boris Mirkin. 2010. Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. J. Classif. 27, 1 (March 2010), 3-40.

Automated Box-Cox Transformations for Improved Visual Encoding

Ross Maciejewski, *Member*, *IEEE*, Avin Pattath, Sungahn Ko,
Ryan Hafen, William S. Cleveland, and David S. Ebert, *Fellow*, *IEEE*

Abstract—The concept of preconditioning data (utilizing a power transformation as an initial step) for analysis and visualization is well established within the statistical community and is employed as part of statistical modeling and analysis. Such transformations condition the data to various inherent assumptions of statistical inference procedures, as well as making the data more symmetric and easier to visualize and interpret. In this paper, we explore the use of the Box-Cox family of power transformations to semiautomatically adjust visual parameters. We focus on time-series scaling, axis transformations, and color binning for choropleth maps. We illustrate the usage of this transformation through various examples, and discuss the value and some issues in semiautomatically using these transformations for more effective data visualization.

Index Terms—Data transformation, color mapping, statistical analysis, Box-Cox, normal distribution

✦

---

1 INTRODUCTION

IN the visual analysis of data, the appropriate choice of display parameters for variable comparison is a complex issue. Under the assumptions of data normality, choices of data binning and axes scaling for visual analysis have been well studied [6], [7], [9], [10], [17], [23], [31]. Unfortunately, real-world data often fail to meet any approximation of a normality assumption. One of the most effective ways of transforming data to a suitable approximation of normality is to utilize a power transformation. The power transformation was introduced by Tukey [29], [30] and further discussed as a means of visualizing data by Cleveland [8].

This concept of preconditioning data (utilizing a power transformation as an initial step) for analysis and visualization is well established within the statistical community and is employed as part of statistical modeling and analysis. However, within the visualization community, the application of appropriate power transformations for data visualizations is largely ignored in favor of interactive explorations (e.g., [16]) or default applications of logarithmic or square root transforms (e.g., [19]). Yet, transformation is a critical

tool for data visualization as it can substantially simplify the structure of a data set.

Traditionally, statisticians have applied data transformations to reduce the effects of random noise, skewness, monotone spread, etc., [8], all of which can affect the resulting data visualizations. For example, reducing random noise can help show global trends in the data, changing the range of values can help fit the data on displays with small screens, and reducing the variance can help improve comparative analysis between multiple series of data. In approximately normal data, methods of data fitting and probabilistic inference are typically simple and often more powerful. Furthermore, the description of the data is less complex, leading to a better understanding of the data itself. As such, by choosing an appropriate power transformation, data can often be transformed to a normal approximation, lending itself to more powerful visual and analytical methods.

Thus, it is clear that there is a strong need to emphasize and explore the preconditioning of data using a power transformation for visualization and analysis. In this paper, we explore the use of the Box-Cox transformation [5] as a means for automatically determining an appropriate power transformation coefficient. Automatic and semiautomatic analyses of the data are performed, and the power transform coefficient that best normalizes the data is calculated and applied. We demonstrate the usefulness of such data preprocessing using examples in time-series visualization, geographical visualization, and histogram binning. This preprocessing step is directly applicable to positively or negatively skewed data; however, bimodal distributions or other irregular data distributions will require different preprocessing steps. Contributions of this work include the following:

1. An approach for applying Box-Cox transformations to scale multiple time series at once.
2. A novel use of Box-Cox transformations for simplifying time series.

- R. Maciejewski is with the School of Computing, Informatics and Decision Systems Engineering, Arizona State University, PO Box 878809, Tempe, AZ 85287-8809. E-mail: rmacieje@asu.edu.
- A. Pattah is with the Microsoft Corporation, Potter Engineering Center, 500 Central Drive, Suite 226, West Lafayette, IN 47907. E-mail: avin.pattah@gmail.com.
- S. Ko, W.S. Cleveland, and D.S. Ebert are with the Purdue University, Potter Engineering Center, 500 Central Drive, Suite 226, West Lafayette, IN 47907. E-mail: ko@purdue.edu, wsc@stat.purdue.edu, ebertd@ecn.purdue.edu.
- R. Hafen is with the Pacific Northwest National Laboratory, Potter Engineering Center, 500 Central Drive, Suite 226, West Lafayette, IN 47907. E-mail: ryan.hafen@pnnl.gov.

3. Methods for computing color bin widths for choropleth maps that can incorporate user analytic interest.

2 RELATED WORK AND TECHNICAL CONCEPTS

The choice of an appropriate power parameter is the most important aspect of the application of the power transform. Power transformations help to achieve approximate symmetry, stabilize variance across multiple distributions, promote a straight line relationship between variables and simplify the structure of a two-way or higher dimensional table [5], [13], [28], [29]. The power transformation [29] is a class of rank-preserving data transformations parameterized by λ (the power) defined as

$$x^{(\lambda)} = \begin{cases} x^{\lambda} & (\lambda \neq 0) \\ \ln(x) & (\lambda = 0), \end{cases} \quad (1)$$

where x is the observed or recorded data.

Under this transformation, for $\lambda = 1$, the data remain untransformed, for $\lambda = -1$, the data are inverted, etc. For data skewed toward large values, powers in the range of $[-1, 1]$ are generally explored. Powers above 1 are not typically used if the data have large positive values because they increases the skewness. It is also commonly observed that as the power is reduced from 1 to -1, the data are transformed until they are nearly symmetric and upon further reduction they become asymmetric again [8]. This is important for visualization as skewed data tend to result in overly large graphs to represent the full dynamic range, or squished graphs where outliers are visible but data near the mean of the distribution are bunched together.

Statistically, we want to find a suitable power for the most appropriate transformation such that the variance in the data is stabilized. Such a value helps in conditioning the data, enabling easier data analysis in subsequent stages. At the same time, it also leads to desirable changes in the data that helps to improve visualizations in 1D and 2D. Traditionally, an appropriate power for the power transformation is chosen through trial and error, by plotting the mean of each data series versus its standard deviation for different powers from a finite set of possible powers determined empirically. Typical choices that are used by statisticians for the power are $\{-1, -\frac{1}{2}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}, 1\}$ since they provide a representative collection of the power transformation [8]. Based on this statistical observation, an appropriate power can be chosen to make the distribution symmetric. Statistically, this means that the data distribution is rid of spread variation, thus leaving us with only location variations which are easier to model. In many cases, the power chosen using the above method also brings the data closer to normality which is always a desired effect in data modeling. While this method of interactively selecting the power transformation provides more control over the choice of the power for each data set, it is cumbersome and may not always result in the best possible power as one cannot examine all the possible choices. Therefore, we utilize an alternative to this manual procedure using the Box-Cox family of power transformations [5].

2.1 The Box-Cox Power Transformation

The transformation, introduced by Box and Cox [5], is a particular family of power transformations with advantageous properties such as conversion of data to an approximately normal distribution and stabilization of variance. Given a vector of n observations $x = \{x_1, \ldots, x_n\}$, the data are transformed using the Box-Cox transformation given by

$$x^{(\lambda)} = \begin{cases} \frac{x^{\lambda} - 1}{\lambda} & (\lambda \neq 0) \\ \ln(x) & (\lambda = 0), \end{cases} \quad (2)$$

where x is the vector of observed or recorded data and the parameter λ is the power. Note that both (1) and (2) are defined only for positive data. However, any nonpositive data can be converted to this form by adding a constant.

Given this initial transformation, Box and Cox [5] then assumed that for some unknown λ, the transformed observations $x_i^{(\lambda)} (i = 1, \ldots, n)$ are independently normally distributed with constant variance σ^2 and with expectations that the transformed responses $x^{(\lambda)}$ will be approximately normal such that

$$x^{(\lambda)} \sim N(A\beta, \sigma^2 I_n), \quad (3)$$

where $N(0, \sigma^2 I_n)$ denotes the multivariate-normal distribution with a mean vector 0. Furthermore, $x^{(\lambda)}$ is an $(n \times 1)$ matrix, A is an $(n \times k + 1)$ matrix and β is a $(k + 1 \times 1)$ matrix.

The likelihood in relation to the original observations, x, is obtained by multiplying the normal density by the Jacobian of the transformation, thus

$$\frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left\{ -\frac{(x^{(\lambda)} - A\beta)^T (x^{(\lambda)} - A\beta)}{2\sigma^2} \right\} J(\lambda; x), \quad (4)$$

where

$$J(\lambda; x) = \prod_{i=1}^{n} \left| \frac{dx_i^{(\lambda)}}{dx_i} \right|.$$

One can then maximize the logarithm of the likelihood function. Readers of this work should refer to the work of Box and Cox [5] for details and derivations. The final derivation for the maximum likelihood estimator yields,

$$L_{\max}(\lambda) = -\frac{1}{2} \log S(\lambda; y)/n, \quad (5)$$

where

$$S(\lambda; y) = y^{(\lambda)^T} a_r y^{(\lambda)}, \quad (6)$$

$$A_r = I - A(A^T A)^{-1} A^T, \quad (7)$$

and

$$y^{(\lambda)} = x^{(\lambda)}/J^{\frac{1}{n}}. \quad (8)$$

Finally, λ can be maximized by taking the derivative of $L_{\max}$ with respect to λ and finding the critical points. In the special case of the one parameter power transformation, $x^{(\lambda)} = (x^{\lambda} - 1)/\lambda$ (which is the focus of our work),
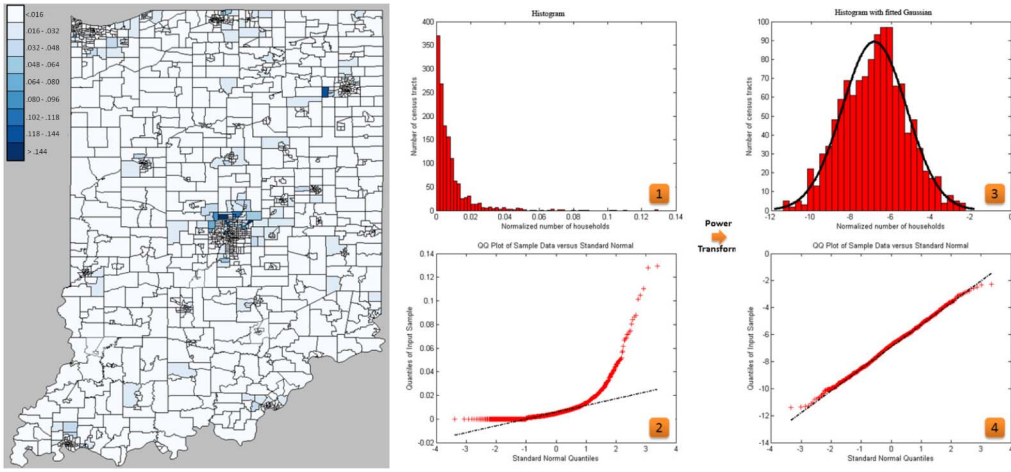
Fig. 1. Skewness in data visualization. (Left) Visualization of the normalized number of households with income greater than $150,000 grouped by census tracts in Indiana (using an equal interval colormap). (Middle) Statistical characteristics of the original data—1: Histogram. 2: Normal Q-Q plot showing skewness. (Right) Statistical characteristics of power transformed data—3: Histogram with a fitted normal distribution. 4: Normal Q-Q plot showing closeness to normality.

$$\frac{d}{d\lambda} L_{\max}(\lambda) = -n \frac{x^{(\lambda)^T} a_r u^{(\lambda)}}{x^{(\lambda)^T} a_r x^{(\lambda)}} + \frac{n}{\lambda} + \sum \log x_i, \qquad (9)$$

where $u^{(\lambda)}$ is the vector of components $\{\lambda^{-1} x_i^{\lambda} \log x_i\}$. One can use Newton's Method for this maximization, and the local maximum can be used (as is done in Matlab).

2.2 Axis Transformations for Visualization

Once an appropriate power transformation is chosen, the data are transformed, which, in turn, means the axis on which the data are plotted is also transformed. Such transformations are of key significance when data are skewed, and, despite the guidelines from a statistical visualization viewpoint [8], few visualizations address the issue of statistically preconditioning skewed data in practice.

Skewed data are data in which the majority of the samples lie near the median values with outliers stretching the data domain to large (or small) values thus increasing the range needed for a given display axis. The plotting of this skewed data compresses values into small regions of the graph resulting in a lower fidelity of visual assessment of the data [8]. One option to improve data assessment would be to remove the outliers and focus on the range of data near the median requiring an interactive technique such as zooming, or users may select the data they are interested in (by brushing) to create a new plot that focuses on the subset of interest. Another option is to apply an appropriate choice of power transformation as a preprocessing step and use this power transformation to transform the axis. This transformation reduces some of the need for interaction and massages the data into a form that is statistically more suitable for advanced analytical techniques.

We illustrate this skewness phenomenon using an example in Fig. 1. The map on the left shows a plot of the normalized number of households with income greater than $150,000, grouped by census tracts, in the state of Indiana. The data are normalized using the total population of the corresponding census tract and displayed using an equal interval colormap. A quick look at the visualization shows high values in the center (around Indianapolis), and low values in most of the rest of the map. However, this

map hides details of the variation found within the lower range of the data as most of the values fall into the lower valued color bins. Plot 1 in Fig. 1 shows the histogram with frequency counts of the data. The x-axis represents the normalized household count and the y-axis is the number of census tracts that fall into the corresponding histogram bin. A quick look at the histogram shows the skewness toward smaller values which results in the unbalanced coloring on the map. This unwanted data characteristic not only makes the data harder to visualize but also makes it difficult to apply standard statistical techniques.

In Fig. 1, we illustrate the effect of applying the Box-Cox power transformation to skewed data. The choropleth map of Fig. 1 shows the number of households by census tract across the state of Indiana whose income exceeds $150,000. In order to compare distributions, we utilize the Q-Q plot (or quantile-quantile plot). Q-Q plots can be used to compare two data distributions by plotting their quantiles on both axes [8]. Graphically, two distributions similar to each other will lie close to the line $y = x$. Linearly related distributions will lie along a straight line, but not necessarily along $y = x$. Normal Q-Q plots show quantiles of given data with standard normal quantiles. In Plot 1 of Fig. 1, we see the original, skewed distribution of the data and the Q-Q plot of this distribution in Plot 2. The quantiles here show a significant departure from the straight line especially on the right, indicating high skewness of the corresponding map data. Plot 3 in Fig. 1 shows the frequency histogram after application of the Box-Cox transformation with automatic power estimation for the map on the left. Additionally, a normal distribution is fit to the transformed histogram data using the expectation maximization algorithm [14] in order to illustrate the transformation to an approximately normal distribution. Plot 4 in Fig. 1, however, shows the Q-Q plot of the transformed data that align closely with a straight line indicating that the power transformation converted the underlying data to a near-normal distribution. As such, in our paper, normal Q-Q plots are used to illustrate skewness in data distributions, since the normal distribution

is symmetrical about its mean and any skewed data cannot produce a linear plot.

Previous work has looked at utilizing power transformations for axis transformations. For example, Cook and Weisberg's Arc system [11], has utilized interactive interfaces in which the user can drag sliders to change the Box-Cox transformation, or simply click a button to set the transformation to the log-likelihood-maximizing value. Unfortunately, many current visual analysis tools still fail to consider the underlying data distribution and instead rely on user intuition. For example, Tableau incorporates frequency plots and histograms and groups the data into bins of equal width; however, the frequency plots and binning used often results in suboptimal visual displays for comparison and analysis and users often will resort to interactive techniques to zoom into the data or manually adjusting bin sizes to remove the effects of outliers [1]. Such procedures can become tedious and often inaccurate, especially when skewed data are involved. Thus, there is a need for the continued exploration and application of power transformations for enhancing both the visual representation and underlying analytical processes.

2.3 Data Binning/Classification

While power transformations are a well studied means of transforming data for analysis and visualization, another important application of such statistical preconditioning of data is the determination of appropriate color intervals for colormaps as seen in prior research in visualization and cartography. Monmonier [23] states that poorly chosen intervals may convey distorted or inaccurate impressions of data distribution and do not capture the essence of the quantitative spatial distribution. As such several simple class interval selection/binning methods (such as quantile, equal interval, and standard deviation) and more complex methods (natural breaks [18], minimum boundary error [12], and genetic binning scheme [2]) have been used traditionally [23]. Several researchers have reported the comparative utility of these methods. Smith [26] reported that quantile and standard deviation methods were most effective with normally distributed data and were most inaccurate with asymmetrical and/or peaked distributions. Moreover, equal interval and natural breaks methods were inconsistent for various data distributions. Frequency-based plots have been used to delineate class intervals [22], particularly for data sets with a standard normal distribution with the curve split into equal bins based on mean and standard deviation [3]. These observations point to the fact that normality, and hence the power transformation, can be useful in determining an effective colormap.

Visual analysis software such as Tableau [1] provide interactive techniques for data binning. Kidwell et al. [19] applied power transformation-based colormaps to visualize incomplete data. However, in all these cases, users need to be familiar with the underlying data distribution to obtain an effective colormap. Therefore, an automatic classification method is favorable when data distributions change frequently, as is the case in interactive visual analysis environments. An automatic color binning/classification method based on extracted statistical properties, including skewness, was described by Schulze-wollgast et al. [24]. However, they

limited the choice of classification to just the logarithmic and exponential mappings, which may not be the best choice for every data set. As such, the power transformation, with an appropriate power value that is best able to reduce skewness and condition the data to near normality, is beneficial in interactive environments to provide an automatic initial visualization based on the transformed data.

Furthermore, in the case of skewed data, research has shown that traditional methods, such as equal interval classification, is ineffective at aiding users in identifying clusters and rates of change in choropleth maps due to inaccurate binning [26]. However, research showed [6], [26] that equal interval classification is as effective as the more sophisticated binning/classification schemes (e.g., Jenks natural breaks [18] and minimum boundary error [12]) when the data fall under a normal distribution.

3 RANGE-SCALING AND AUTOMATIC BINNING OF TIME-SERIES PLOTS USING POWER TRANSFORMATION

In this section, we illustrate an application of the automated Box-Cox transformation for improved visual encoding in time-series plots. Time-series plots are often used for visual analysis and are likely to be skewed due to unusually large data values occurring as spikes in the plot. Representing such highly varying plots causes issues in scaling and simply changing the aspect ratio cannot solve the problem. This issue is compounded when representing multiple time-series plots in a single graph as some of the plots may become compressed (e.g., Fig. 2b (left)). We use a time-series data set representing patient visits to a hospital within the INPC (Indiana Network for Patient Care) [4] as an example. The recorded observations are the total number of visits as well as visits categorized into three groups (constitutional, respiratory and gastrointestinal complaints) over a period of five years. The y-axis of the transformed graphs is then in the newly transformed space; however, the values of the labels are transformed back to the original space in order to allow for the analysts to work with the values in their original representation.

Results of applying the Box-Cox power transformation to the time plot are shown in Fig. 2. The leftmost column shows the data plotted with time on the x-axis and the number of visits on the y-axis. The rightmost column shows transformed data plotted with time on x-axis and the transformed data value on y-axis. For comparison and analysis, the middle column shows normal Q-Q plots of the original data (top) and transformed data (bottom), allowing us to visualize the skewness of the corresponding data.

3.1 Transformation of a Single Plot

The first column in Fig. 2a shows a plot of total patient visits to a hospital and one can clearly see that most of the data are compressed to the bottom of the graph as they need to accommodate both high and low values. Its corresponding Q-Q plot, in the middle column, confirms the skewness of the data. However, the Box-Cox transformation can be used to find a suitable power to transform the data that better utilizes the space, allowing us to simultaneously see the spike as well as the detail in the previously compressed
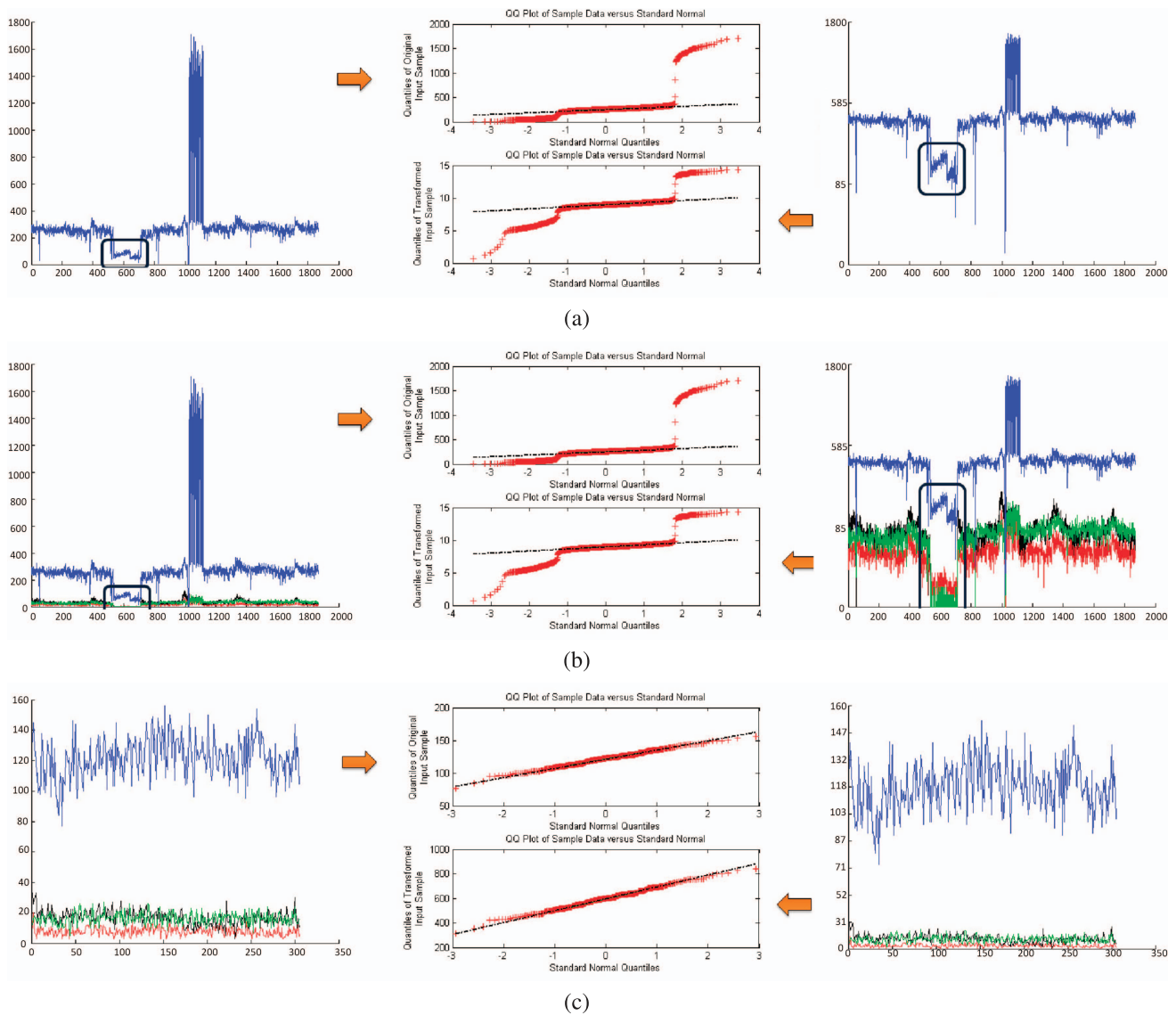
Fig. 2. Visual and statistical conditioning in time-series plots using the Box-Cox power transformation applied to daily INPC hospital visit data. (Left) Original data plots. (Middle) Normal Q-Q plots of original data (top) and transformed data (bottom). (Right) Transformed data plots. The x-axis represents the day count and the y-axis represents number of hospital visits. The graph legend is as follows: blue (total number of hospital visits), red (constitutional complaint visits), black (respiratory complaint visits) and green (gastrointestinal complaint visits). Figures show the power transformation applied to: (a) a single skewed plot ($\lambda = .3613$), (b) multiple skewed plots simultaneously ($\lambda = .3613$), and (c) ($\lambda = 1.4011$). Plots without significant skewness leading to compression. Highlighted windows indicate improvement in plot depiction after the Box-Cox transformation leading to better data interpretation.

region. For example, in the transformed plot, the dip in the graph to the left of the spike allows the user to visualize details during the dip period which are compressed in the original plot. The corresponding normal Q-Q plot shows that the transformed data are more symmetric with a smaller data range, indicating the statistical conditioning is closer to normality.

3.2 Simultaneous Transformation of Multiple Plots

During visual analysis, analysts often compare multiple plots of related data simultaneously to get a quick overview of potential correlations or temporal trends. During such comparisons, a spike in one of the data sets can cause other data plots to be compressed. An example is shown in Fig. 2b second row. The actual plot in the left column shows the total number of visits in blue and the numbers of constitutional,

respiratory and gastrointestinal complaints in red, black, and green, respectively. From the original plot, we cannot see the details about individual complaint categories. In cases with multiple data, we use the data set with the largest range to compute an appropriate power transformation using the Box-Cox transformation. This power is then used to transform all other data sets in order to maintain uniformity while simultaneously maximizing the use of the display space (since we normalize the data set with the highest variance). The normal Q-Q plots shown in this case correspond to the data set with the largest range (i.e., in this case, the blue plot showing total number of visits). From the transformed plot, we can clearly see that during the dip in total visits, the number of constitutional complaints were greater than gastrointestinal complaints, where as this was not the case for the rest of the duration. Future work will focus on more
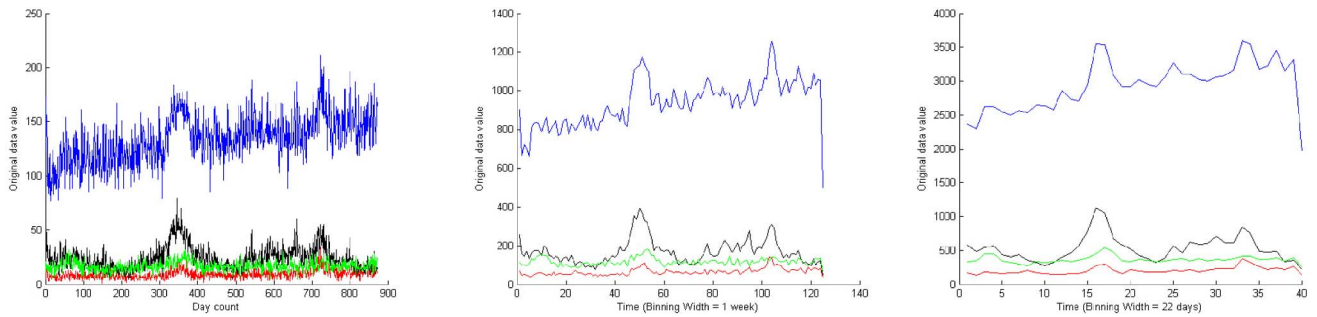
Fig. 3. Power transformation for automatic time-series binning. (Left) Original INPC hospital time-series data. (Middle) Original data with a predetermined bin width of one week that still retains some noise. (Right) Original data with bin width ($= 22$ days) determined from a normal distribution fit over the power transformed data. Most of the noise is removed, while the overall temporal trend are easily perceived.

advanced schemes where properties of all data sets being plotted for comparison could be analyzed and incorporated in a more sophisticated transformation.

3.3 Limitations

While the power transformation is a powerful tool, there are a limited number of cases in which it is not appropriate to use, particularly, in cases where the power is outside the range of $[-1, 1]$. In these cases, the data may be overly exaggerated or inverted depending on the sign of the power. Fig. 2c shows an example of this case. From the raw data, we can see there are no clearly visible spikes in the data, and the corresponding normal Q-Q plot suggests that the data are already close to normal. The power computed by the log-likelihood function maximization in this case is 1.4011. Data transformation with this power causes the plot to be further compressed while in the Q-Q plot, there is no significant change in the normality of the data. Therefore, the application of this procedure should be limited to data plots that contain at least one skewed plot that is significantly nonnormal. Normality of the given data plot can be measured automatically by computing the correlation coefficient of its normal probability plot and thresholding the coefficient value. Moreover, the automatically computed power should be checked if it is in the range $[-1, 1]$ before application of this procedure. Generally, transformations will fall within this range; however, this limitation is significant and transforming data that are already a reasonable approximation of normality add another layer of complication to the analysis process that is unnecessary.

3.4 Automatic Time-Series Binning for Global Trend Display

Time-series plots are typically noisy. Traditionally, either interactive methods or bin widths based on a prior knowledge of the data distribution are utilized to highlight temporal trends in the data. Automatic binning is beneficial in an interactive visual analytic environment to provide a good initial time-series display showing global trends. As such, using the Box-Cox power transformation, we can convert data to an approximately normal distribution and use the properties of the normal distribution (standard deviation) to determine an approximate bin width. We determine the maximum-likelihood estimates of the parameters (mean and standard deviation) of a normal distribution fitted to the data using the expectation maximization algorithm [14]. The standard deviation is

then used as a binning factor to smooth out local variances while retaining global trends in the data.

Fig. 3 shows an example of applying this procedure to the INPC hospital data shown on the left. In this case, we obtain a result of 22 days as the bin width. The rightmost figure shows the original plots binned by 22 days and the middle figure shows the original plots binned by seven days for comparison. As can be in Fig. 3, the middle figure still retains some of the noise from the original plot where as the rightmost figure smoothes out most of the noise while presenting trends in all four plots of Fig. 3(Right). Furthermore, the standard deviation of the transformed data may provide analysts with cues as to cyclical behavior while preserving other trends.

Other methods may also be used to show global trends and smooth the data. If only slightly larger bins are used (say 31 days for a bin as opposed to 22 days) the results will be similar to the point that almost no differences would be observable. However, what this automatic binning provides is a means of automatically approximating an appropriate bin width by bringing the data in line with an assumption of normality through the power transformation. In other methods, such as weighted moving averages, or exponential smoothing, parameter choices need to be made about the smoothing parameter, which can affect the result depending on if the data are normal or nonnormal. By approximating the data as normal through the power transformation, other methods can be applied for such smoothing. Thus, the application of the power transformation can enhance both the analytic and visualization tools of a system.

4 RANGE-SCALING AND SEMIAUTOMATIC COLOR BINNING/CLASSIFICATION BASED ON INTENT

Colormaps in interactive visual analysis environments are typically generated either manually (by adjustment of color bins) or using preexisting methods such as quantile, equal interval, or standard deviation-based binning. However, each of these methods is most appropriate for specific data distributions and not applicable in general [6], [20], [26], [27]. Moreover, in interactive environments, users perform various operations such as zooming, panning, filtering, and selection, and temporal browsing that constantly change the underlying data distribution. Therefore, there is a need to generate automatic colormaps that can provide a good initial visualization that highlights important features in the
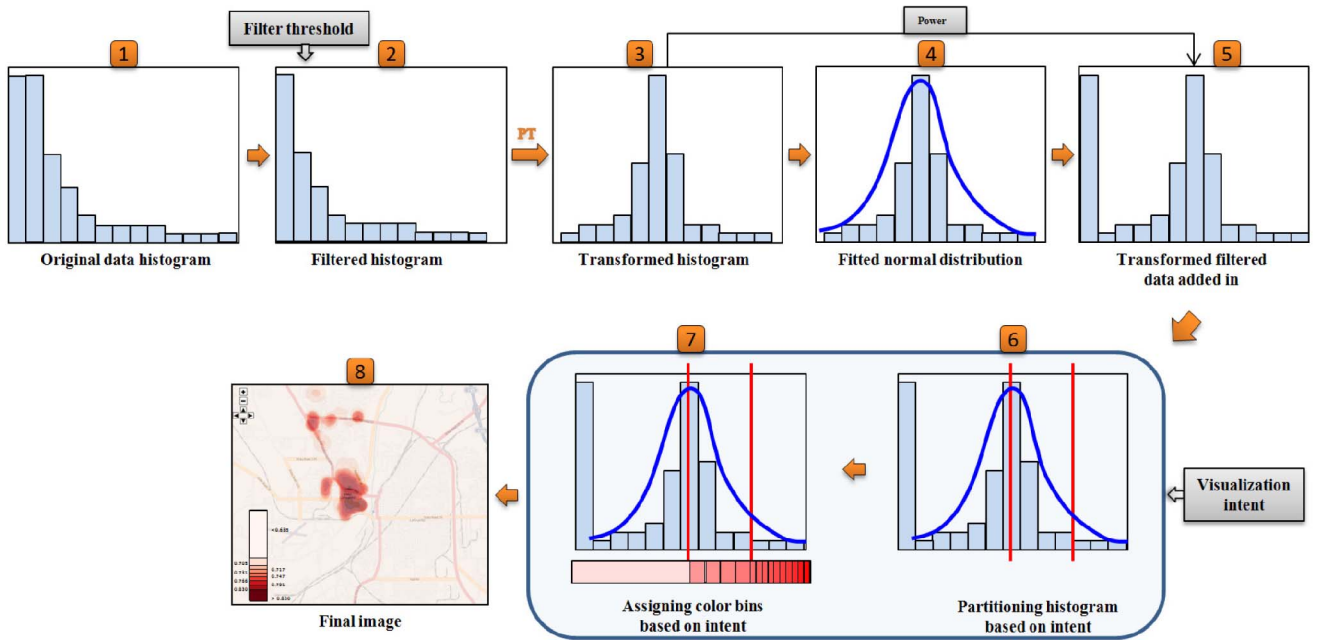
Fig. 4. Range scaling and our intent-based semiautomatic color binning procedure using the power transformation. This power transformation is applied after step 2. Steps 6 and 7 in the shaded rectangle are executed based on the users' visualization intent.

underlying data. In the next section, we describe a method for applying the power transform to the data and using the transformed data space to determine the bin widths for colormapping. This is a semiautomatic technique as the user needs to provide two initial parameter values—a threshold to filter out uninteresting data and specification of their visualization intent. Here, the filtering step is not necessary; however, it can be used to remove data that is of no interest to the analyst (low ranges when looking for hotspots or outliers when looking for trends) prior to visualization.

4.1 Procedure

Fig. 4 shows the procedure that starts with a default histogram representing the data as illustrated in step 1. The histogram is constructed such that the x-axis represents the data values that are mapped to colors and the y-axis, or the histogram counts, represents the number of entities with a specified data value or a range of data values depending on the histogram bin resolution. In many visualizations, a majority of the display entities represent default values that may not be of interest to the user. These values can interfere with the data transformation. In step 2, we filter out such values using a filter threshold. At the same time, we add a constant value to the data set to eliminate the zero values since the Box-Cox power transform requires positive data values. The filter threshold is data dependent and is determined based on the default values in the data set or the range of data values that do not significantly contribute toward understanding the visualization. Currently, this value needs to be specified by the user before starting an interactive visualization session.

The Box-Cox power transformation is applied, in step 3, to the filtered histogram to convert the data into an approximately normal distribution. The Box-Cox transformation automatically determines the power to be applied by maximizing the log-likelihood function. In step 4, the expectation maximization algorithm [14] is used to estimate the parameters of a normal distribution (mean and standard deviation) that best fit the transformed data. In step 5, we add in all the data that was filtered out in step 2 after applying the power transformation determined in step 3. Steps 6 and 7 are dependent on users' visualization intent. In step 6, we determine the positions of histogram divisions based on the normal distribution parameters estimated in step 4. The estimated normal distribution curve is shown overlaid on the histogram in red. In step 7, we determine the number of colors to be assigned to each division based on the visualization intent. Both steps 6 and 7 are described and illustrated in the subsequent sections using two visualization intents and three data sets. In step 8, we obtain the final visualization by applying this colormap.

Note that the user intent is actually used in two phases of this process. Initially, the user filters the data by selecting regions that they deem as uninteresting. While still in this space, the user can see only the untransformed space. After the data are transformed, the intent is that more details within the data will emerge, and the user can further refine their intent in the transformed data space.

4.2 Displaying Complete Range of Skewed Data

Skewed data can cause issues when a user is trying to visualize the entire range of data using predetermined colormaps as shown in Fig. 5(left). The left map in Fig. 5 shows the census tracts in Indiana colored by the number of households (normalized by the census tract population) with annual income greater than \$150,000. An equal interval color binning method, in this case, makes it difficult to see the variation of data in the lower range (all the white tracts) as most of the data are skewed toward tracts with low incomes. Our goal is to visualize the entire range of values simultaneously, without having to manually adjust the colormap each time the data change as a result of user interaction.
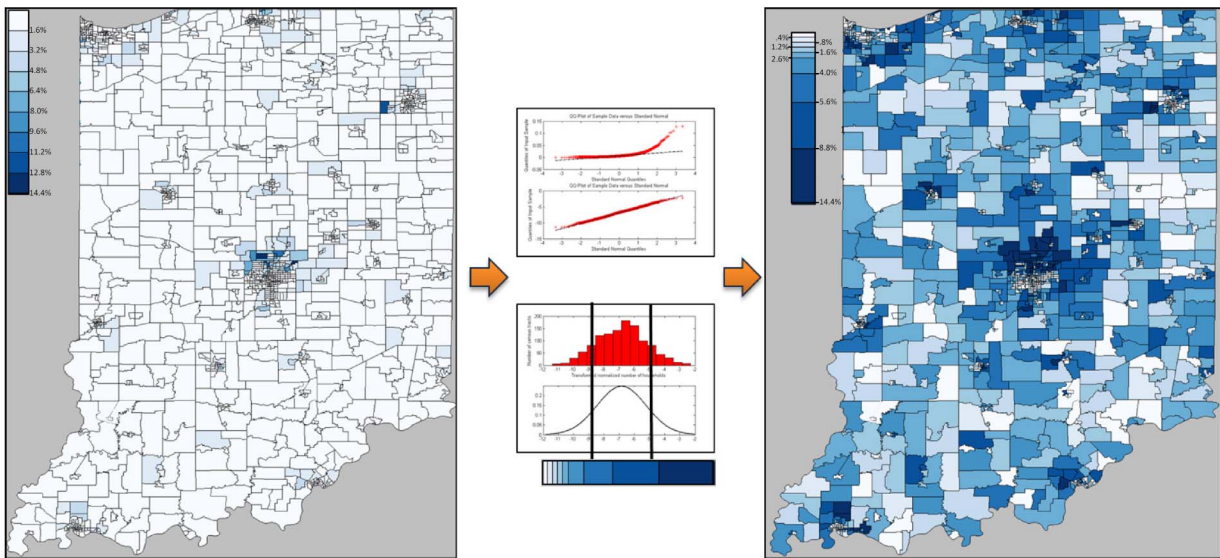
Fig. 5. Color bin boundary determination and color assignment for visualizing the census data. (Left) Visualizing the original data with an equal interval colormap showing the number of households (normalized by the census tract population) with an annual income greater than $150,000. (Middle) Q-Q plots of the original and transformed data (top) and the transformed histogram with a fitted normal curve and color bin boundary assignment (bottom). (Right) Visualization using the new colormap showing better variation in color throughout the data range.

4.2.1 Color Bin Boundary Determination and Assignment

We construct a histogram for this data with the number of households on the x-axis and the corresponding number of census tracts on the y-axis. We do not apply a filter threshold in this example as our intent is to visualize the entire data range. The first five steps of our procedure are applied to transform the filtered data. In step 6 of our procedure, we divide the histogram into three regions at one standard deviation from the mean on either side. The middle top plot in Fig. 5 shows the Q-Q plots for the original and transformed histogram. The middle bottom plot shows the transformed data histogram, fitted normal curve and histogram divisions. Following our goal of showing the entire range of data, we assign colors that follow the fitted normal distribution based on the standard deviation method. We assign more colors near the mean of the distribution and fewer colors beyond the standard deviation on either side of the mean in step 7 of our procedure. As 95 percent of the transformed data will fall within two standard deviations of the mean (as the transformed data should be an approximately normal distribution), this method robustly covers the data. The colormap below shows the color assignment for this data set with two, five, and two colors using ColorBrewer's sequential blue colormap. The map on the right of Fig. 5 shows the result after applying this colormap and one can now clearly see the entire range of the data, where as, in the map on the left side of Fig. 5, the majority of the data is mapped to two bins (the lightest two colors in our chosen scheme).

4.2.2 Results

Further power transform results that visualize the entire data range on the map are shown in Fig. 6. These figures show the number of households in Indiana earning an annual income between $45,000 and $60,000 grouped by census tracts and normalized by the corresponding census tract population. Here, we compare colormaps obtained using the traditional quantile binning method (Fig. 6(Left)) and a colormap determined using the commonly used logarithmic transformation (Fig. 6(Middle)), with our procedure using the power (Box-Cox) transformation (Fig. 6(Right)). The normal curve fitted after the log transform follows the same division as the Box-Cox transform color scale division described in Section 4.2.2. Note that the Box-Cox color mapping is able to bring out better variation when compared to the quantile mapping and log-based mapping.

4.2.3 Limitations

Occasionally, data can still be significantly nonnormal after the Box-Cox transformation. For example, bimodal distributions will fail to approximate normality even after the application of a power transformation. Although such data occur infrequently in practice, our procedure may not generate the best classification in these cases. However, in these situations, a goodness-of-fit value of the normal distribution can be computed using the normal probability plot of the transformed data. The applicability of our procedure can be assessed by computing the correlation coefficient of this plot and thresholding the value. Moreover, as with the time-series data, the application of this procedure to determine colormaps is limited to power values in the range $[-1, 1]$, as other powers can significantly alter the data values.

4.3 Visualizing Hotspots in Detail Using an Adaptive Nonlinear Colormap

In geospatial visualization, density-scaled heatmaps are often used to convey relative data densities on a map. One particular method of density estimation uses a variable kernel width to determine density estimates for each pixel on the map [21], [25]. Fig. 7(left) shows hotspots indicating criminal incidents during 2006 in the city of West Lafayette, Indiana. An equal interval colormap ranging between the minimum and maximum values of the density estimates is used. The actual colors are drawn with an alpha value less
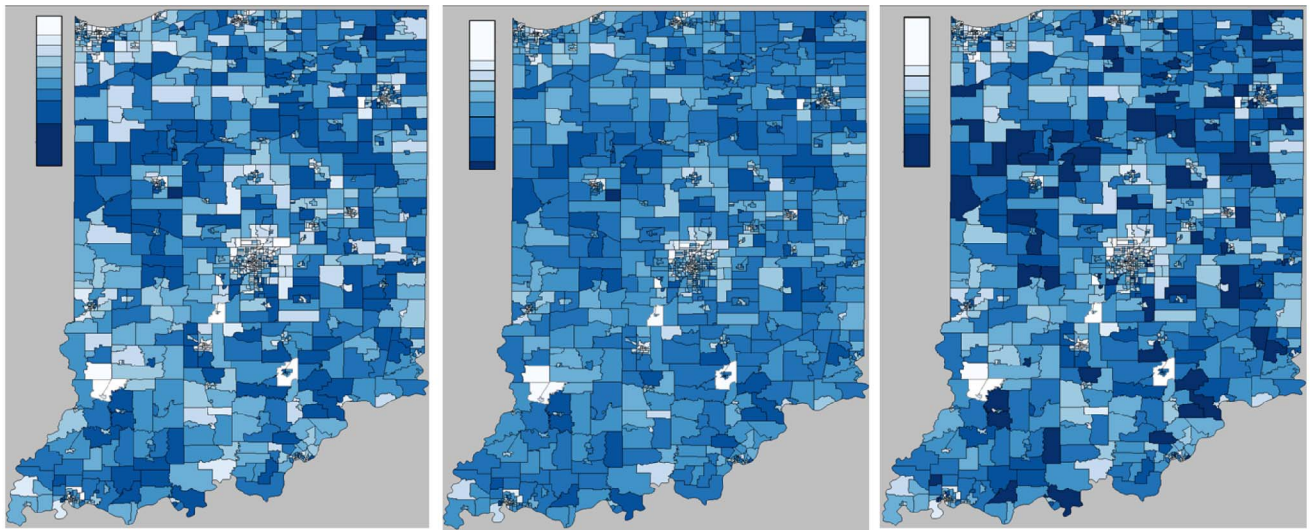
Fig. 6. Comparison of various colormaps in the visualization of income data by census tracts in Indiana showing the number of households (normalized by the corresponding population) with an annual income between $45,000 and $60,000 using Quantile binning (Left), Logarithm transformed binning (Middle), and Power (Box-Cox) transformed binning (Right). The power transformation mapping shows better global and local variation than both the quantile binning and the logarithmic transformation. For example, areas near the center of the state are now able to show more local variation, whereas before, these census tracts were all binned to approximately the same color.

than 1 so as to show the underlying area map. In an interactive visualization environment, these density maps, and hence the underlying density histograms, change frequently as a result of user actions (such as zooming/ panning, temporal browsing and data selection and filtering). In this situation, a predetermined colormap defined without regard to the underlying data distribution may wash out large areas of the map due to lack of sufficient color resolution as shown in the highlighted boxes in the left figure. However, our procedure from Section 4.1 can provide a more effective initial colormap that adapts itself based on changing density estimates. The visualization intent in this example is to view details within high density hotspots by allocating more colors to such areas. Using the power transformation, we modify the histogram into a structure with certain assumptions of normality, and then assign a colormap based on the normal distribution parameters as described below.

4.3.1 Color Bin Boundary Determination and Assignment

Based on our visualization intent of finding details within hotspots, our focus of interest in the histogram is around the values that determine these hotspots, i.e., based on high density estimates occurring toward the right end of the histogram. We use a filter threshold value of 0.0005 in step 2, for this data set, which helps to remove the default zero values, while retaining most of the significant data values. After performing steps 1 to 5 of our procedure, in step 6, we divide the histogram into three unequal parts using divisions at the mean and one standard deviation beyond the mean of the fitted normal curve. The middle top graphs of Fig. 7, shows Q-Q plots of the original and transformed data (top and bottom, respectively). The middle bottom plot shows the transformed data histogram along with the fitted normal curve and the histogram
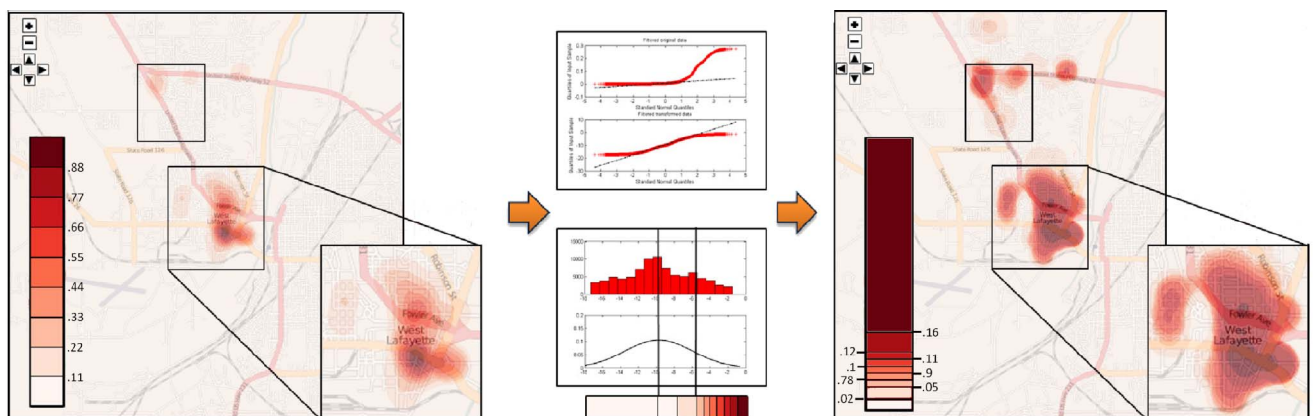
Fig. 7. Color bin boundary determination and color assignment for visualizing hotspots within hotspots of criminal activity density estimates in the year 2006 in West Lafayette. (Left) Original density estimate hotspots using an equal interval color map. (Middle) Q-Q plots of the original and power transformed data (top) and the transformed histogram, fitted normal curve and color bin boundary and assignment (bottom). (Right) Hotspot visualization using the new color map. Notice the inner hotspots (darkest red regions) and newly visible hotspots in the highlighted rectangular regions.
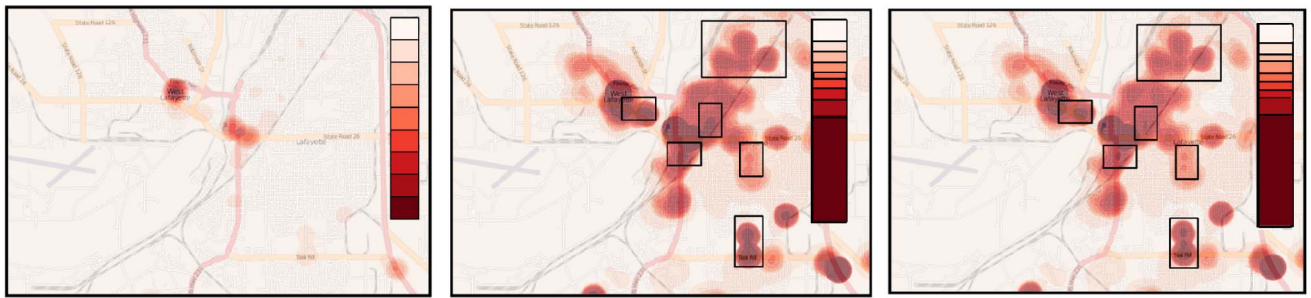
Fig. 8. Comparison of various colormaps in visualizing hotspots of arrests in the year 2008 in Lafayette using (Left) Equal interval, (Middle) Logarithm transformed, and (Right) Power (Box-Cox) transformed colormaps. Both (Middle) and (Right) show more accurate hotspots. Note, however, that in (Right), one can better see the hotspot differentiation within the highlighted rectangular regions.

divisions. The colormap below shows the color assignment in step 7 of our procedure. Using ColorBrewer's sequential red colormap [15] with nine colors, we assign colors exactly as shown in this colormap. Color bin lengths are assigned such that they divide the normal distribution into equal quantiles within each division. Note that while quantiles do not change due to power transformations, the user's selection divides the data into user-defined chunks, which are then divided by quantiles. Thus, this is not the same as assigning the entire data range to a set of quantiles, which would wash out the effects of the transformation.

The lightest color is assigned to the first histogram division as well as to one of the bins in second division. The next lightest color is assigned to the other bin of the second division and the rest of the colors to the seven bins in the third division with larger density values. The corresponding visualization is shown on the right. Note that in the zoomed region on the bottom right we can clearly see hotspots within hotspots as compared to the washed out region in the left map. Moreover, we can also simultaneously see more hotspots in the rectangular highlighted region on the top part of the right map which were missing in the left map using equal interval color bins. The choices are all designed by the user guided intent, where the user chooses the number of bins to be assigned to each partition.

4.3.2 Results

Additional results of the density hotspot visualization, using our procedure, are shown in Fig. 8, which represents arrest data in Lafayette, Indiana in the year 2008. Here, we compare traditionally used equal interval color binning (left) and a colormap determined using the commonly used logarithmic transformation (middle), with our procedure using the power (Box-Cox) transformation (right). The normal curve is fitted after the log transform using the same quantile division as the Box-Cox transform. While the equal interval colormap in Fig. 8(Left) hardly produces any hotspots, logarithmic (Fig. 8(Middle)) and Box-Cox (Fig. 8(Right)) colormaps show a more separated representation of the data. However, the Box-Cox colormap is better able to differentiate regions within the hotspots in the highlighted rectangular areas. These examples illustrate that our procedure can automatically generate an appropriate colormap for different data sets with varying data distributions without any manual parametric modifications.

5 CONCLUSIONS AND FUTURE WORK

We have demonstrated the utility of the power transformation in conditioning data for better visualization. Using the Box-Cox class of power transformation with automatic power estimation, we demonstrated automatic and semi-automatic procedures to determine visualization parameters that yield good initial visualizations. We used examples of time-series plots to illustrate the usage of this transformation to visualize significantly skewed data as well as to automatically bin time-series data to highlight global temporal trends. Further, we described a procedure to use the normality conversion property of the power transformation to determine an effective colormap based on users' visualization intents. We presented results of this procedure using geospatial data, compared them with commonly used transformations and binning methods in visualization, and discussed some limitations.

In the future, we will employ better data fitting models, such as a normal mixture model, to more accurately fit the transformed data, and explore histogram binning techniques based on multiple normal curves. We also plan to automate our color binning procedure by determining an appropriate filter threshold automatically. Further, motivated by Schulze-Wollgast et al. [24], we plan to develop visualization methods to better represent and interpret new color legends with significantly nonuniform bin lengths obtained after transformation.

REFERENCES

[1] Tableau Binning Measures, http://www.tableausoftware.com/public/knowledgebase/binning-measures, 2012.
[2] M.P. Armstrong, N.C. Xiao, and D.A. Bennett, "Using Genetic Algorithms to Create Multicriteria Class Intervals for Choropleth Maps," *Annals Assoc. of Am. Geographers,* vol. 93, no. 3, pp. 595-623, 2003.
[3] R.W. Armstrong, "Standardized Class Intervals and Rate Computation in Statistical Maps of Mortality," *Annals Assoc. Am. Geographers,* vol. 59, no. 2, pp. 382-390, 1969.
[4] P.G. Biondich and S.J. Grannis, "The Indiana Network for Patient Care: An Integrated Clinical Information System Informed by over Thirty Years of Experience," *Public Health Management Practices,* vol. 10, pp. 81-86, Nov. 2004.

[5] G. Box and D. Cox, "An Analysis of Transformations," *J. Royal Statistical Soc. Series B (Methodological),* vol. 26, no. 2, pp. 211-252, 1964.

[6] C.A. Brewer and L. Pickle, "Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series," *Annals Assoc. Am. Geographers,* vol. 92, no. 4, pp. 662-681, 2002.

[7] W.S. Cleveland, *The Elements of Graphing Data.* Wadsworth Publ. Co., 1985.

[8] W.S. Cleveland, *Visualizing Data.* Hobart Press, 1993.

[9] W.S. Cleveland and R. McGill, "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *J. Am. Statistical Assoc.,* vol. 79, no. 387, pp. 531-554, 1984.

[10] W.S. Cleveland and R. McGill, "Graphical Perception and Graphical Methods for Analyzing Scientific Data," *Science,* vol. 30, pp. 828-833, 1985.

[11] R.D. Cook and S. Weisberg, *Applied Regression Including Computing and Graphics.* John Wiley, 1999.

[12] E.K. Cromley and R.G. Cromley, "An Analysis of Alternative Classification Schemes for Medical Atlas Mapping," *European J. Cancer,* vol. 32A, no. 9, pp. 1551-1559, 1996.

[13] C. Daniel, F. Wood, and J. Gorman, *Fitting Equations to Data.* Wiley, 1980.

[14] A. Dempster et al., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc. Series B (Methodological),* vol. 39, no. 1, pp. 1-38, 1977.

[15] M. Harrower and C. Brewer, "Colorbrewer. org: An Online Tool for Selecting Colour Schemes for Maps," *The Cartographic J.,* vol. 40, no. 1, pp. 27-37, 2003.

[16] H. Hauser, F. Ledermann, and H. Doleisch, "Angular Brushing of Extended Parallel Coordinates," *Proc. IEEE Symp. Information Visualization,* pp. 127-130, 2002.

[17] J. Heer, N. Kong, and M. Agrawala, "Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time Series Visualizations," *Proc. 27th Int'l Conf. Human Factors in Computing Systems,* pp. 1303-1312, 2009.

[18] G.F. Jenks, "The Data Model Concept in Statistical Mapping," *Int'l Yearbook of Cartography,* vol. 7, pp. 186-190, 1967.

[19] P. Kidwell, G. Lebanon, and W. Cleveland, "Visualizing Incomplete and Partially Ranked Data," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1356-1363, Nov./Dec. 2008.

[20] A. MacEachren, *Some Truth with Maps: A Primer on Symbolization and Design.* Assoc. Am. Geographers, 1994.

[21] R. Maciejewski, S. Rudolph, R. Hafen, A.M. Abusalah, M. Yakout, M. Ouzzani, W.S. Cleveland, S.J. Grannis, and D.S. Ebert, "A Visual Analytics Approach to Understanding Spatiotemporal Hotspots," *IEEE Trans. Visualization and Computer Graphics,* vol. 16, no. 2, pp. 205-220, Mar./Apr. 2010.

[22] J.R. MacKay, "An Analysis of Isopleth and Choropleth Class Intervals," *Economic Geography,* vol. 31, pp. 71-81, 1955.

[23] M.S. Monmonier, "Contiguity-Biased Class-Interval Selection: A Method for Simplifying Patterns on Statistical Maps," *Geographical Rev.,* vol. 62, no. 2, pp. 203-228, 1972.

[24] P. Schulze-wollgast, C. Tominski, and H. Schumann, "Enhancing Visual Exploration by Appropriate Color Coding," *Proc. Int'l Conf. Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG),* pp. 203-210, 2005.

[25] B.W. Silverman, *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, 1986.

[26] R.M. Smith, "Comparing Traditional Methods for Selecting Class Intervals on Choropleth Maps," *Professional Geographer,* vol. 38, no. 1, pp. 62-67, 1986.

[27] L. Stegena and F. Csillag, "Statistical Determination of Class Intervals for Maps," *The Cartographic J.,* vol. 24, no. 2, pp. 142-146, 1987.

[28] M.A. Stoto and J.D. Emerson, "Power Transformations for Data Analysis," *Sociological Methodology,* vol. 14, pp. 126-168, 1983.

[29] J.W. Tukey, "On the Comparative Anatomy of Transformations," *Annals Math. Statistics,* vol. 28, pp. 602-632, 1955.

[30] J.W. Tukey, *Exploratory Data Analysis.* Univ. Microfilms Int'l, 1988.

[31] L. Wilkinson, "Algorithms for Choosing the Domain and Range when Plotting a Function," *Computing and Graphics in Statistics,* pp. 231-237, Springer-Verlag, 1991.

Ross Maciejewski received the PhD degree in electrical and computer engineering from Purdue University in December 2009. He is currently an assistant professor at Arizona State University in the School of Computing, Informatics & Decision Systems Engineering. Prior to this, he served as a visiting assistant professor at Purdue University and worked at the Department of Homeland Security Center of Excellence for Command Control and Interoperability in the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) group. His research interests are geovisualization, visual analytics, and nonphotorealistic rendering. He is a member of the IEEE and the IEEE Computer Society.

Avin Pattath received the PhD degree in computer engineering from Purdue University. He is a computer scientist with Microsoft. His research interests include mobile visualization.

Sungahn Ko is currently working toward the PhD degree in electrical and computer engineering from Purdue University. His research interests include visual analytics and information visualization.

Ryan Hafen received the PhD degree in statistics at Purdue University. His research interests include exploratory data analysis and visualization, massive data, computational statistics, time series, modeling, and nonparametric statistics.

William S. Cleveland received the PhD degree in statistics from Yale University. He is the Shanti S. Gupta distinguished professor of statistics and courtesy professor of computer science at Purdue University. His research interests include statistics, machine learning, and data visualization. He is the author of *The Elements of Graphing Data* (Hobart Press, 1994) and *Visualizing Data* (Hobart Press, 1993).

David S. Ebert received the PhD degree in computer science from Ohio State University. He is a professor in the School of Electrical and Computer Engineering at Purdue University, a University faculty scholar, director of the Purdue University Rendering and Perceptualization Lab, and director of the Purdue University Regional Visualization and Analytics Center. His research interests include novel visualization techniques, visual analytics, volume rendering, information visualization, perceptually based visualization, illustrative visualization, and procedural abstraction of complex, massive data. He is a fellow of the IEEE and the IEEE Computer Society, and a member of the IEEE Computer Society's Publications Board.

Bristle Maps: A Multivariate Abstraction Technique for Geovisualization

SungYe Kim, Ross Maciejewski, *Member*, *IEEE*, Abish Malik, Yun Jang, *Member*, *IEEE*,
David S. Ebert, *Fellow*, *IEEE*, and Tobias Isenberg, *Member*, *IEEE*

Abstract—We present Bristle Maps, a novel method for the aggregation, abstraction, and stylization of spatiotemporal data that enables multiattribute visualization, exploration, and analysis. This visualization technique supports the display of multidimensional data by providing users with a multiparameter encoding scheme within a single visual encoding paradigm. Given a set of geographically located spatiotemporal events, we approximate the data as a continuous function using kernel density estimation. The density estimation encodes the probability that an event will occur within the space over a given temporal aggregation. These probability values, for one or more set of events, are then encoded into a bristle map. A bristle map consists of a series of straight lines that extend from, and are connected to, linear map elements such as roads, train, subway lines, and so on. These lines vary in length, density, color, orientation, and transparency—creating the multivariate attribute encoding scheme where event magnitude, change, and uncertainty can be mapped as various bristle parameters. This approach increases the amount of information displayed in a single plot and allows for unique designs for various information schemes. We show the application of our bristle map encoding scheme using categorical spatiotemporal police reports. Our examples demonstrate the use of our technique for visualizing data magnitude, variable comparisons, and a variety of multivariate attribute combinations. To evaluate the effectiveness of our bristle map, we have conducted quantitative and qualitative evaluations in which we compare our bristle map to conventional geovisualization techniques. Our results show that bristle maps are competitive in completion time and accuracy of tasks with various levels of complexity.

Index Terms—Data transformation and representation, data abstraction, illustrative visualization, geovisualization

✦

1 INTRODUCTION

As data dimensionality increases, the encoding of variables and their relationships is often abstracted down to a representative subset for analysis in a single display, or dispersed across a series of coordinated multiple views [1], [2], [3]. Moreover, many techniques have been developed to visually encode multiple data attributes/ variables for each data sample to enable interactive analysis, ranging from discrete glyph attribute encoding [4] to more spatially continuous color, transparency, and shading encodings [5], [6], [7]. As the number of visualized variables increases, the amount of information that can be effectively displayed becomes limited due to overplotting and cluttering [8]. This is especially a problem in geographical visualization as a key attribute of the data is the location within the two-dimensional map space.

In geographical visualization, data can be described at any given location on a map. The data being described can come from an aggregated measurement, a direct event

occurrence, or various other means. In dense data sets, plotting events as symbols on the map (e.g., Fig. 1a) leads to cluttering and is often unable to convey a meaningful sense of event magnitude within the data. Aggregation of the data by defined boundaries, such as county or census tract boundaries (e.g., Fig. 1b), leads to a loss of specificity in data location and runs afoul of the Modifiable Areal Unit Problem [9]. Furthermore, it is known that the level of data aggregation can affect aspects of task complexity such as information load and the user's ability to recognize patterns within the data [10]. To combat problems associated with areal aggregation, dasymetric mapping focuses on using zonal boundaries that are based on sharp changes in the statistical surface being mapped [11]. However, even when grouping data into small spatial quadrats, data can either be overaggregated or underaggregated. A third option is to estimate the discrete event points as a continuous function (e.g., Fig. 1c); such a mapping, however, only allows for the use of color as a means of representing data variables. As an encoding based on underlying network data, Fig. 1d shows a traditional line map. However, its representation is still restrained by the color and thickness of the lines.

To increase the amount of information that can be visualized within the constraints of a thematic map, this paper explores a novel method of multivariate encoding. Inspired by ideas of symbolic encoding from Spence [12] and choices of visual encodings by Wilkinson [13], we have developed the bristle map (Fig. 1e), a novel method for the aggregation, abstraction, and stylization of geographically located spatiotemporal data. The bristle map consists of a series of straight lines extended from and connected to linear map elements (roads, train lines, subway lines, etc.) that have some contextual relationship with the data being

---

- *S. Kim, A. Malik, and D.S. Ebert are with the School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Avenue, West Lafayette, IN 47907. E-mail: {inside, amalik, ebertd}@purdue.edu.*
- *R. Maciejewski is with School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Mail Code 8809, Tempe, AZ 85287-8809. E-mail: rmacieje@asu.edu.*
- *Y. Jang is with Department of Computer Engineering, Sejong University, 98 Gunja-dong Gwangjin-gu, Seoul, 143-747, South Korea. E-mail: jangy@sejong.edu.*
- *T. Isenberg is with Team Aviz, INRIA-Saclay, Bât 650, Université Paris-Sud, 91405 Orsay Cedex, France. E-mail: tobias.isenberg@inria.fr.*
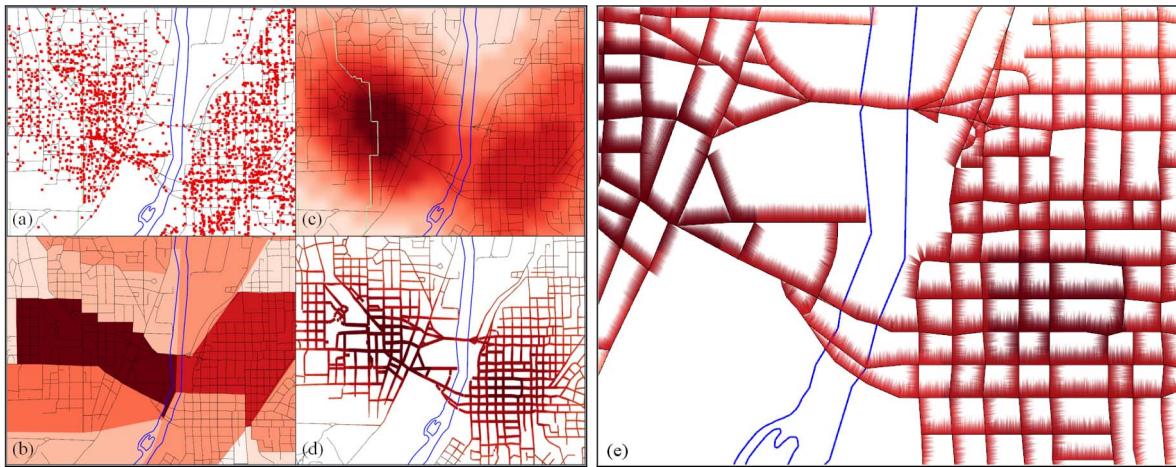
Fig. 1. Data abstraction in geovisualization. In this image, we show crimes in West Lafayette and Lafayette, Indiana, where the blue line represents the Wabash River. (a) Plotting events as points. (b) Aggregation of points by areal units. (c) Approximation of a continuous domain from point sampling. (d) Approximation of a continuous domain using solid lines applied to roads. (e) Our abstraction using a series of bristle lines applied to roads.

visualized. We vary these lines with respect to their color, length, density, and orientation to allow for a unique encoding scheme that can be used to create informative maps. With respect to the other representations shown in Fig. 1, our technique utilizes the underlying geographical context as a part of its symbology, thereby directly incorporating geographical elements within its encoding scheme. One of the major advantages of the bristle map technique is that the basis domain of the data (e.g., street network) remains highly visible regardless of the color scale being used. If one compares Figs. 1c and 1e, the street network in Fig. 1e is clearly visible because the lines only "bristle off" to one side, whereas in Fig. 1c some streets are hardly discernible due to the dark colors.

To demonstrate our technique, we focus on categorical spatiotemporal event data (e.g., emergency department logs, crime reports). In such data, events consist of locations in time and space where each event fits into a hierarchical categorization structure. These categories are typically processed as time series and snapshots of time are aggregated and typically visualized on a choropleth map [14]. Past work [6], [15] has shown that the use of kernel density estimation (KDE) [16] is highly suitable in the spatial analysis of such data. Thus, our approach incorporates kernel density estimation as a means of estimating the underlying distribution of spatiotemporal events. Using the estimated distribution in an area for a given category (or categories) and temporal unit, we incorporate the underlying geographical network structure into the visual encoding. Bristles are extended from this underlying structure, and the color, length, density, transparency, and orientation of each bristle are mapped to a particular variable (or set of variables). Schemes presented in this paper include combinations of the following mappings:

- length, density, and color as data magnitude,
- orientation and coloring for bivariate mapping,
- color and length for bivariate mapping,
- color and density for bivariate mapping, and
- length and transparency for temporal variance.

Given the available parameters for visual encoding within the bristle map, other encodings also exist, which illustrate the flexibility and power of our technique. Our work focuses on showing how bristle maps can be used to show spatial and temporal correlations between variables, encode uncertainty in a unique and aesthetically informative way, and maintain geographical context through linking our visual encoding directly to geographical components. As such, the bristle map is a powerful multivariate encoding scheme that is adaptable to various attribute encodings to create richly informative visualizations.

2 RELATED WORK

Many techniques in multivariate data visualization focus on a means of reducing clutter and highlighting information through a variety of approaches including filtering (e.g., [17]), clustering (e.g., [18]), and sampling (e.g., [19]). In this section, we focus particularly on techniques within geographical visualization for improving the understanding of thematic/statistical maps, as Wilkinson [13] noted that the problem of multivariate thematic symbology for maps is that they are not only challenging to make, but also challenging to read.

In geographical visualization, the most common means of data representation is the choropleth map in which areas are shaded or patterned in proportion to a measured variable. Such maps are typically used to display only one variable, which is mapped to a given color scale. Other research has focused on encoding multivariate information into choropleth maps (such as uncertainty) with textures and patterns [20], creating bivariate color schemes for visualizing interactions between two variables [21], [22], or animating choropleth maps to enhance the exploration of temporal patterns and changes [23]. We present bristle maps as a robust alternative to these schemes in which multivariate attributes are instead mapped to a variety of graphical properties of a line (length, density, color, and orientation), as opposed to utilizing a bivariate color scheme, texture overlays, or animation.

More recent geographical visualization techniques have included extensions to choropleth mapping ideas. Hagh-Shenas et al. [24] compared the effectiveness of visualizing geographically referenced data through the use color blending (in which a single composite color conveys the values of multiple color encoded quantities) and color weaving methods (in which colors of multiple variables are separately woven to form a fine grained texture pattern). The results from their study indicate color weaving to be more effective than color blending for conveying individual distributions in a multivariate setting. Saito et al. [25] proposed a two-tone pseudo coloring method for visualizing precise details in an overview display. Under this scheme, each scalar value is represented by two discrete colors. Sips et al. [26] focused on revealing clusters and other relationships between geospatial data points by their statistical values through the overplotting of points. This work was later extended [27] to combine a cartogram-based layout to provide users with insight to the relative geospatial positioning of the data set while preserving cluster information and avoiding overplotting. Other cartogram techniques include the WorldMapper Project [28] which is used to represent social and economic data of the countries of the world. In each of these, novel data visualization techniques are created; however, the distortion of spatial features (country boundaries, roads) is often undesirable. While these techniques focus on displaying large amounts of aggregate data on small screens, our technique focuses on enhancing details of geographical context within the data. A similar concept of preserving data context is found in Wong et al.'s [29] GreenGrid in which they visualize both the physics of the power grids in conjunction with the geographical relationships using graph-based techniques.

Along with the previously described map schemes and cartogram distortions, there has been work in the use of heatmaps based on spatial data. Fisher [30] applied heatmaps to visualize the trends of the interactions of users with interactive maps that are based on their view of the geographic areas. Maciejewski et al. [6] used heatmaps as one of the tools to find aberrations or hotspots that facilitate the exploration of geo-spatial temporal data sets. Work by Chainey et al. [15] illustrated a number of different mapping techniques for identifying hotspots of crime and demonstrated that kernel density estimation provides analysts with an excellent means of predicting future criminal activities.

In conjunction with previous visualizations, other research has focused on expanding the dimensionality of the data being displayed by utilizing three-dimensional visuals. Van Wijk and Telea [7] utilized color and heightfields to visualize scalar functions of two variables. Tominski et al. [31] explored embedding 3D icons into a map display as a means of representing spatiotemporal data. In contrast, our work focuses on a two-dimensional encoding scheme that incorporates a variety of the visual variables described by Bertin [32] and Wilkinson [13] as a means of representing multivariate data.

Finally, it is important to note that our technique is akin to traditional traffic flow maps (e.g., Fig. 1d) seen in a variety of atlases; however, provides more generalized schemes. In traffic flow maps, the amount of data that can be displayed is restrained by the color and the width of the line representing linear elements (i.e., roads) on the map. Our work is similar to that of the traffic flow maps in that we utilize width (specifically, matched to the length in our bristle maps) and color as underlying visual variables of our encoding. However, our work also incorporates bristle density as a means of further encoding parameters. In the following sections, we compare our encodings to a variety of methods including the point, color, and flow line maps.

3 BRISTLE MAP GENERATION

In Fig. 1, we developed our motivation for the need to directly incorporate geographic features to the underlying data to better preserve contextual information. It is clear that the aggregation of data into arbitrary geographical areas obscures data, while the continuous approximation of an underlying data source can lead to incongruent mappings with respect to geographic features. Furthermore, both these mappings are limited in the fact that only color and texture are available for variable encoding, limiting the amount of data that can be displayed to either a single variable or possibly two variables in the case of a bivariate color map. The goal of this work is to create visual encodings for higher order structures.

The bristle map was inspired by the Substrate simulation of Tarbell [33] and abstract renderings of map scenes in work by Isenberg [34]. Given these images, our work focuses on using the underlying visual properties to intelligently encode information for display. In *The Grammar of Graphics* [13], Wilkinson discusses the combination of several perceptual scales into a single display. Here, he notes the idea of separable dimensions of the data is a key issue, where discriminations between stimuli are of key importance in the visualization. The Substrate aesthetic directly lends itself to this approach as color, line length, and orientation are distinct classes within Wilkinson's table of aesthetic attributes and each of these visual parameters directly contributes to the substrate aesthetic.

Fig. 2 illustrates the bristle map generation pipeline. Given underlying data events, we compute a continuous distribution. We also create a topology graph from given geographically relevant linear content for clutter reduction described in Section 5. As an example of geographical content, if the underlying data was water pollution we could use a city sewage map for the geographic components, for our crime data examples we use roadways. Each linear geographic component consists of a series of line segments, and we extend *bristle lines* from these line segments. These bristle lines emerge perpendicularly from the underlying geographical line segment and are allowed to vary in length, density, color, transparency, and orientation, to facilitate multivariate data encoding. The third stage of the bristle map generation pipeline (Fig. 2) illustrates the bristle line concept for each geographical line segment, $\overline{SE}$, and $\overline{P_1P_2}$ defines our generated bristle line. Each bristle line is created using the vector equation of a line as shown in

$$P_2 = P_1 + \vec{V_l}L_l = P_1 + \vec{V_l}(t \times L_{lmax}). \qquad (1)$$
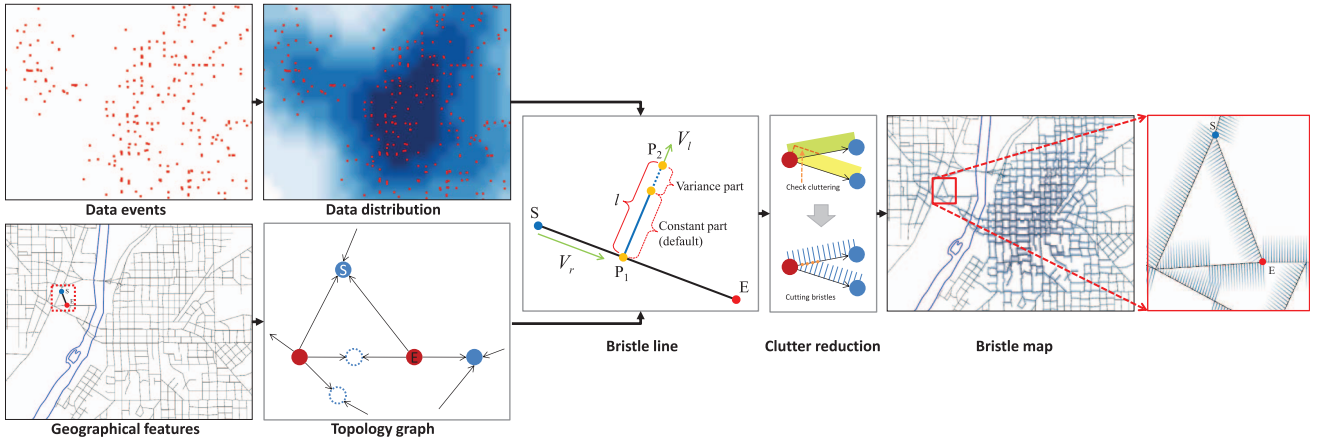
Fig. 2. The bristle map generation pipeline. Beginning with data events, a continuous abstraction is created. We also create a topology graph from contextually important linear features (in this case roads). Next, bristles are extended from these features based on the continuous abstraction and the topology. Clutter reduction is performed when generating each bristle, and finally the resultant bristle map is generated.

Here, P_1 is a point on the contextually relevant geographic line segment, $\overline{SE}$, $\vec{V_l}$ is a unit vector perpendicular to the line $\overline{SE}$, and $L_{l_{max}}$ is the maximum length of the $\overline{P_1P_2}$. L_l is the length of the $\overline{P_1P_2}$ determined by a parameter t.

Each line from P_1 to P_2 is drawn in such a manner that it will either encode different properties of a multivariate data set, or use a data reinforcement technique where properties are encoded to the same variable to provide redundant cues. We utilize three encoding properties for each bristle: length, color, and orientation. The length of a line $\overline{P_1P_2}$ is separated into two portions: a constant component, which is proportional to the magnitude of the variable being encoded, and a variance component. It can also capture other properties such as level of certainty. The color of a bristle $\overline{P_1P_2}$ is proportional to the underlying variable distribution to be encoded at point P_1. When the variance component is used, its transparency is adjusted as a means of visually distinguishing it from the constant component. Orientation of the bristle line is always perpendicular to $\overline{SE}$ and is utilized for bivariate comparison (i.e., day/night, two data types) and/or clutter reduction. To summarize, length and color represent a local data magnitude property at point P_1. We also choose to encode redundant information into the density of the number of bristles placed on a given line segment, where the density of the bristles along $\overline{SE}$ is decided by an average data value on a line segment $\overline{SE}$.

For each visual encoding, the underlying data is assumed to be continuous over a given geographical segment, such that for all points between any two nodes on the underlying contextual geographic structure, a data distribution value is associated with the point. In the case of a discrete data set (e.g., crime locations), the choice of an appropriate means of data interpolation with regards to the underlying geographic information is dependent on the data analysis being performed. Based on the recommendations of Chainey et al. [15], we apply a kernel density estimation [16] to approximate the underlying distribution of crimes over the geographic features. The kernel density estimation procedure used is defined by the following equation:

$$\hat{f}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{\mathbf{x}-X_i}{h}\right). \quad (2)$$

Here, the window width of the kernel placed on point x is proportional to a window bandwidth, h, and the total number of samples, N. We utilize the Epanechnikov kernel [16]:

$$K(\mathbf{u}) = \frac{3}{4}(1-\mathbf{u}^2)1_{(\|\mathbf{u}\|\leq 1)}, \quad (3)$$

where the function $1_{(\|\mathbf{u}\|\leq 1)}$ evaluates to 1 if the inequality is true and zero for all other cases.

Thus, given a multivariate data set where locations in space and time correspond to a series of categorized events, we can create bristle maps that encode various properties of the data. Note that this technique relies on the data being contextually relevant to an underlying geographical network. For example, crime event data with its 2D geographical coordinates is recorded and hence defined by addresses on streets; thus, it is contextually relevant to a street network. Data sets in which this contextual relationship does not exist should utilize other visual encoding schemes. Table 1 shows the parameters in our bristle map and their corresponding potential variables being encoded to each parameter. In the following section, we present a series of potential parameter combinations for various bristle map encodings and discuss the various results.

TABLE 1
Parameters, Corresponding Variables, and Ranges

Parameters	Potential variables	Range
Base position (P_1)	Geographic location	(Double, Double)
Length 1 (constant portion)	Data magnitude	Double
Length 2 (variance portion)	Temporal variance, accuracy	e.g., monthly/yearly
Color	Data magnitude	Discrete, Continous
Transparency	Temporal variance, accuracy	Double [0.0, 1.0]
Orientation ($\vec{V_l}$)	Temporal difference, data type	Clock-wise, Counter clock-wise
Density	Average data magnitude on an area ($\overline{SE}$)	Double

4 ENCODING SCHEMES

The bristle map is a powerful visual encoding scheme that lends itself to a variety of data encodings, examples of which we present next. For demonstration purposes, we employ categorical spatiotemporal police reports collected in Tippecanoe County (specifically West Lafayette and Lafayette, IN, USA), from 1999 to 2010. The data set contains the date, time, crime type (e.g., armed/unarmed aggravated assault, armed robbery, burglary, homicide, noise, other assaults, rape, rape attempted, residential entry, robbery, theft, vandalism, and vehicle theft), and the address of each recorded criminal event. Note that other data sets can be easily encoded with bristle maps, and our choice of data was only made to illustrate the technique.

Utilizing this multivariate crime data set, we discuss potential encoding schemes for multivariate spatiotemporal data. We then provide illustrations of each described encoding scheme with respect to our crime data set. Encoding schemes presented in this section include the use of bristle color, length, and density to encode data magnitude, the use of bristle orientation to inform temporal comparison, and the encoding of temporal variance in the bristle lengths.

4.1 Color, Length, and Density as Data Magnitude

Here, we discuss our technique for encoding the color, length, and density of the bristles into two separate variable groups. As both color and length (size) fall into two distinct categories of aesthetics according to Wilkinson [13], the use of separate variables for both categories allows for a distinguishable visual data encoding. In both cases, we assign data magnitude to both a color scale and a length scale. We note that such an encoding scheme has the potential to portray data more effectively than visualizations that map each data variable to a single display parameter. As noted in the arguments for the use of redundant color scales by Rheingans [35], the use of different display parameters is able to convey different types of information. Furthermore, by combining encodings in a redundant manner, it is possible to reinforce the encoding scheme. The utility of redundant color scales was confirmed by Ware [36].

In our encoding scheme, each bristle line's length, L_l, is calculated using (4) based on a parameter, t, and the maximum length, L_{lmax}:

$$L_l = t \times L_{lmax} = (\alpha \times \kappa_{P_1} + \beta \times v_{P_1}) \times L_{lmax}. \quad (4)$$

For this visual encoding of the bristles, the parameter t is defined by the ratio of the data value at P_1, which we call κ_{P_1}, the ratio of the temporal variance at P_1, v_{P_1}, and a set of tuning parameters (α and β) that provide weights to the constant and variance components as shown in Fig. 2. In this work, we use $\alpha = 1.0$ and $\beta = 0.3$. Note that the choice of encoding the variance at a 30 percent value was chosen through trial and error by generating visualizations that the authors found to be the most useful and aesthetically pleasing. For problems where determining exact data values from the visual encoding is required (as opposed to approximating high and low rates), the variance portion is removed from the equation entirely by using $\beta = 0.0$. As such, by creating the encoding scheme with diverse parameters, we are able to generate

more aesthetic choices and visualizations. It is important to note that not all encodings will be appropriate and are most likely task dependent.

The L_{lmax} portion of (4) is defined in

$$L_{lmax} = \rho \times log_b \left(\frac{1}{N_r} \sum_{i=0}^{N_r-1} L_{\overline{SE}} \right). \quad (5)$$

In this equation, we take the average length of all line segments (where N_r is the total number of line segments in the map) and calculate L_{lmax} using a nonlinear function such that the length of bristle lines does not grow in an unbounded manner when zooming in. Moreover, L_{lmax} is modified by the parameter ρ, where ρ is the ratio of the current zoom level to the initial zoom level, to decouple our technique with the zoom level. In this work, we use $b = 15$ for the base of a log function.

Next, we determine the number (or density) of bristles, N_l, to be drawn on each line segment $\overline{SE}$ using

$$N_l = \rho \left(\frac{\zeta}{\lambda} L_{\overline{SE}} \right) \kappa_{\overline{SE}}. \quad (6)$$

Here, N_l is calculated using two user-defined constants λ and ζ, where λ is the unit geographical length (distance) and ζ is the number of bristle lines per unit geographical length. We use $\lambda = 0.0009$ and $\zeta = 3$-15 in our current visualization. As the bristle density may also be used to encode data magnitude parameters in bristle map generation, N_l should be proportional to the ratio of average data value on $\overline{SE}$, $\kappa_{\overline{SE}}$. Moreover, we also apply ρ such that N_l will be independent of the zoom level to preserve the extent of density.

For color, we allow users to choose either a continuous or a sequential color scheme from Color Brewer [37]. Then, data are linearly mapped to a probability that a crime of type A will occur at geographic point B, where the probability is estimated from the underlying data distribution using kernel density estimation as described in Section 3.

Fig. 3 illustrates our length, density, and color encoding using the previously described crime data set. Burglary is encoded with the red color scheme, and color is proportional to the probability (calculated from the underlying point distribution using kernel density estimation) that a burglary occurred at a given location. Fig. 3(left) shows our bristle map encoding for burglary rates with a color scheme and bristle density, and Fig. 3(right) shows a line map encoding the same information with a color scheme and line thickness for comparison to our bristle map. Compared to this line map, our bristle map provides the advantages of additional dimensionality through the density of bristle lines. In this scheme, one is able to easily encode two variables in different combinations of bristle map parameters (i.e., color and density with a constant length, color and length with a constant density), and provide users with distinguishable visual parameters that seem to focus attention to various details.

4.2 Multivariate Encoding: Separating Length, Density and Color, and Using Orientation

In the previous section, we illustrated how our method can be utilized for univariate encoding by using a redundant encoding scheme. However, a major benefit of bristle maps
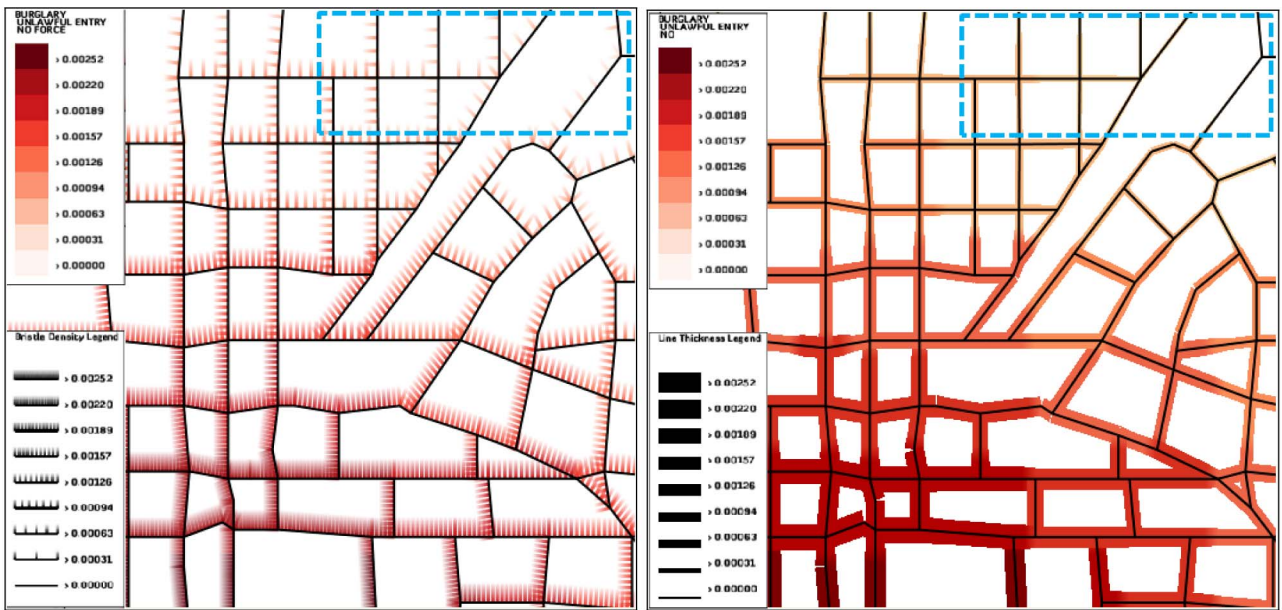
Fig. 3. (Left) Our bristle map encodes burglary rates with both bristle color and bristle density. (Right) A line map encoding burglary rates as both line color and line thickness. Compared to the line map, our bristle map provides a distinguishable visualization by incorporating bristle density. For example, bristle lines on the right top area are easily identified, whereas thickness in the line map on the same area is too small to clearly be perceived.
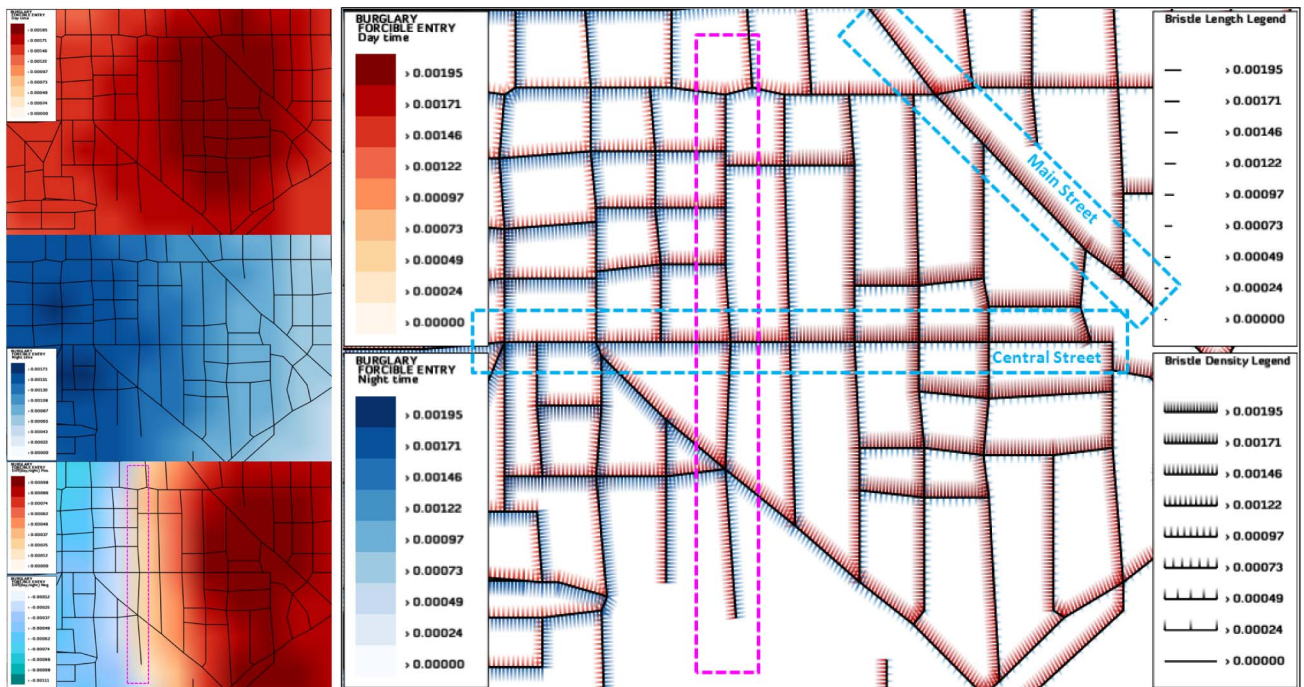

Fig. 4. Encoding daytime versus nighttime variations. (Left column) From top to bottom, color maps showing day, night, and the difference of day and night burglary rates. (Right) Our bristle map separating the burglary rates into their day and night components with opposite orientation along roads. Note that a color map cannot present two components (i.e., both daytime and nighttime burglary rates) at the same location, hence three color maps are needed to see day and night variations simultaneously. Our bristle map can present such information within one bristle map by using different orientations of bristle lines.

is the ability to encode multivariate attributes. One example of this is seen in day versus night time comparison.

Here, one can utilize the orientation to separate two temporal components of a single variable by mapping the temporal components to different orientations of the geographic feature. For instance, it is likely that the rates of data variable will be different with respect to day and night occurrences. We illustrate this visual encoding in Fig. 4. We separate the events into day (6:00 am-6:00 pm)

and night (6:01 pm-5:59 am) and map the daytime rates to red and one orientation, and nighttime rates to blue and the other orientation. In Fig. 4(right), we illustrate a bristle map encoding of one variable (burglary) during 2009 where length, density, and color represent the magnitude of the burglary as well as the encoding of day and night parameters is explored as line orientation.

In Fig. 4(right), we show areas of high/low nighttime crime, high/low daytime crime, and combinations there

within. In contrast, a traditional heatmap using a univariate color scheme can only show either daytime crime (Fig. 4(left top)), or nighttime crime (Fig. 4(left middle)). Hence, several heatmaps are needed to see day and night variations as shown in Fig. 4(left column). Viewers must mentally combine the images to locate regions of the map that have high crime levels at daytime and nighttime, thereby increasing their cognitive load.

Another means of reducing the cognitive burden would be to create a heatmap of the difference between night and day. Fig. 4(left bottom) shows the difference of day and night data, and the divergent color scheme shows where high daytime or high nighttime crimes occur. For instance, in Fig. 4(left bottom) the right area indicates higher rates during day, the left area shows higher rates during night, and the border area between the blue and red color schemes only indicates that day and night rates were approximately equal, regardless of them being low or high. Moreover, you need other color maps to explore areas, where one occurs similarly high or low during day and night time.

Bristle map encodings have benefits in this situation. When we explore a daytime versus nighttime bristle map in Fig. 4(right), we see that there exists distinct temporal profiles along the road lines, where we see exclusively dominant areas during either day or night. For instance, see the diagonal road from the top center to the right center (Main Street, Lafayette, IN) showing that daytime burglary dominates along this road. Another observation is made on the horizontal road at the center of the map (Central Street, Lafayette, IN). Along this road, daytime burglary rates increase from left (west) to right (east), whereas nighttime burglary rates decrease from left to right. For the center area in Fig. 4(left bottom), where the blue and red color schemes meet, we also see in Fig. 4(right) that it has relatively equally high rates during both day and night. Such a comparison allows people to understand the differences between the data; however, when subtracting, areas of nearly equal daytime and nighttime crimes will be colored the same. Thus, areas that are safe during both day and night, and areas that are highly dangerous during both day and night will appear the same in the difference color map. In contrast, bristle maps allow viewers to quickly observe trends related to both day and night.

Another example of multivariate encoding using our bristle map is done by separating and/or combining bristle parameters. For instance, bristle density (or length) encodes a variable A, and color encodes a variable B while being presented on one orientation. Similarly, another two variables (C and D) could be encoded and presented on the other orientation. However, this type of parameter combination should be determined carefully so as not to increase viewers' cognitive load. Its effectiveness would depend on several factors such as data type and analysis purpose. In Section 6, we conduct experiments to explore the effectiveness of different parameter combinations.

4.3   Encoding Data Variance

As introduced in Fig. 2, each bristle can include a portion generated for temporal variance of data, see (4). To present the temporal variance of the data over time, we compute both the monthly and yearly mean and variance values. For a given discrete data set during time periods N_T, we first calculate continuous distributions over time. Then, we determine mean and standard deviation values with respect to the underlying data distribution for the entire data set over a given temporal aggregation. Thus, we calculate the mean μ and variance σ values from time varying data K_i, where $i \in [0, N_T - 1]$. Note that μ and σ are computed only once as they represent constant values for a given data set. Mean and variance values for each grid point j are calculated using (7) and (8), respectively. Variance is then used to weight the parameter β in (4) such that given the data magnitude at the current time K_{cur}, we compute the ratio of variance at the current time, $\tilde{\sigma}$ as shown in (9). As such, the parameter t in (4) can be detailed as shown in (10) to represent the length of bristle lines with respect to temporal variance:

$$\mu[j] = \frac{1}{N_T} \sum_{i=0}^{N_T-1} K_i[j] \tag{7}$$

$$\sigma[j] = \sqrt{\frac{1}{N_T} \sum_{i=0}^{N_T-1} (\mu[j] - K_i[j])^2} \tag{8}$$

$$\tilde{\sigma}[j] = \frac{1}{\sigma[j]} |\mu[j] - K_{cur}[j]| \tag{9}$$

$$t = \alpha \times \kappa_{P_1} + \beta \times \upsilon_{P_1} = \alpha \times \kappa_{P_1} + \beta \times \left(\frac{\tilde{\sigma}_{P_1}}{\tilde{\sigma}_{max}}\right). \tag{10}$$

Furthermore, the variance term, υ_{P_1}, in parameter t in (4) can also be revised to encode an uncertainty factor by using randomness. We may also encode an uncertainty factor by using color and transparency to enhance the variance component. When using color and transparency, we use a highlight color for the variance component, and then fade out the variance component over the bristle length with a full alpha value for one end point and an alpha value weighted by the variance for the other end point. The constant portion of the bristle is assigned an alpha value of 1 to both end points as it represents an exact data value. Hence, according to the data type and analysis purpose, the encoding of parameter t and the use of the variance portion can be different and should be assessed with respect to the visual message trying to be conveyed. Fig. 5 illustrates the application of encoding the data variance of vandalism with the uncertainty factor. In Figs. 5a and 5c, we use the same color scheme for the constant and variance portions of bristle lines. To enhance the variance component in Figs. 5b and 5d, we highlighted the variance portion in a different color and assigned full alpha values for the constant portion of bristle lines. Figs. 5a and 5b show the same area. In this area, the bristle length shows large fluctuations, indicating a high yearly variance Figs. 5c and 5d show another area. In this area, the bristles are of a nearly constant length, indicating low yearly variance. When considering that the area in Figs. 5a and 5b includes residential areas, while the area in Figs. 5c and 5d includes the downtown Main street, an art theater, and the City Hall in Lafayette, IN, our bristle map shows that the residential areas have higher yearly
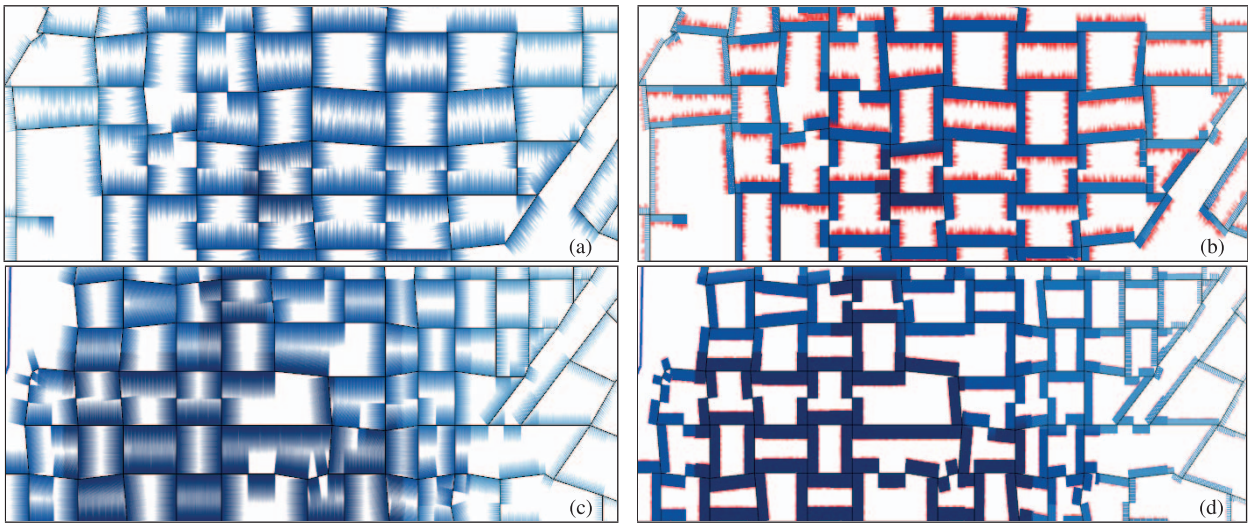
Fig. 5. Encoding data variance of vandalism-graffiti in Lafayette, IN, USA, in 2010 creating an uncertainty aesthetic. Yearly variance of vandalism-graffiti is represented in (a) a residential area and (c) a commercial area without distinguishing the variance component in the bristle length. Parts (b) and (d) show the results using a highlight color for the variance portion and full alpha values for the constant portion of bristle lines. Here, we clearly see that our bristle map can encode the temporal variance and create an uncertainty aesthetic using the variance component.

variance of vandalism (graffiti) when compared to commercial areas.

5 BRISTLE CLUTTER REDUCTION

Although our bristle map can encode various characteristics from multivariate data, it often suffers from clutter around the intersections of road lines. To minimize cluttering, we employ two strategies in our bristle map generation pipeline (Fig. 2): 1) using topology among road lines to determine bristle orientation to minimize clutter and 2) cutting bristle lines crossing neighbor road lines.

5.1 Using Topology

Each bristle map contains an underlying topology of the contextual geographic network that the data re mapped to. In the topology graph, each node is defined as either "outward" or "inward" as illustrated in Fig. 6. Using the topology graph, we choose each segment's bristle line orientation such that the overlap of the bristles at intersections will be minimized, thereby reducing the clutter. If the encoding scheme requires both sides of the edge to contain bristles, then clutter at each intersection is inevitable. However, in cases where bristles map to only one side of an edge, we use the right-hand rule to decide the orientation. Hence, bristle lines on edges connected to neighboring

outward and inward nodes are generated in a manner that provides a reasonable reduction in clutter (Fig. 6).

Choosing the orientation of bristle lines to minimize overlap can be considered as a 2-coloring problem in vertex coloring; one color presents "outward" while the other presents "inward." Vertex coloring is a well-known graph problem, where no two adjacent nodes share the same color. Moreover, coloring a general graph with the minimum number of colors is known to be an NP-complete problem. In our case, the minimum number of colors should always be 2 but such 2-colorability is not guaranteed for general road lines. While deciding the orientation of bristle lines, we often have undesirable topology generating inevitable overlap of bristle lines. Fig. 7 (upper row) shows such a bad topology example and our strategy to solve this issue. In Fig. 7a, we see two clutter areas caused by an undesirable configuration of neighbor nodes, which guarantee bristle overlap. To solve this, we consider the addition of a virtual node in a topology graph as shown in Fig. 7b, thereby allowing for an orientation switch midway across

Fig. 7. Two pairs of the cluttering cases and our methods to minimize clutter. Colored box areas on a side of each edge line indicate the orientation for bristle lines. (a) Case 1: bad topology, where two inward nodes (blue) share a line and two outward nodes (red) share a line, generates inevitable clutter. (b) Virtual nodes (dotted circles) are added to split an edge line. (c) Case 2: a small angle between edge lines causes a clutter area. (d) Bristle lines crossing a neighbor edge line are cut on the neighbor line.

Fig. 6. To minimize clutter, a topology graph consisting of directed edges as road lines and outward (red) and inward (blue) nodes on the intersection of lines is used to decide the bristle line orientation.

Fig. 8. Before/after image pairs of our clutter reduction. Each pair shows a case of (a) changing bristle orientation using topology, (b) cutting bristle lines crossing neighbor road lines, (c) circular roads, and (d) curved roads.

the edge and reducing the clutter. For neighboring two inward nodes (blue), we add a virtual outward node (red dotted circle) at the road line connecting two inward nodes resulting in splitting bristle lines on the road line. Similarly, a virtual inward node (blue dotted circle) is added for neighboring two outward nodes (red).

5.2 Avoid Crossing Neighbors

Another cluttering case is illustrated in Fig. 7c. When two road lines intersecting with less than a 90 degree angle have bristle lines, some of the bristle lines overlap as illustrated in Fig. 7c. For this case, we forbid bristle lines to cross neighbor road lines by placing the end point of a bristle line on the neighbor road line as shown in Fig. 7d. We first check the intersection of bristle blocks (colored boxes in Fig. 7) for the current road line on which we are generating bristle lines and its neighboring road lines by using the topology graph. If the blocks are intersected, we then check if a bristle line crosses the neighbor road lines by utilizing the intersection algorithm of 2D line segments [38]. This idea is based on the theory of amodal completion (or amodal perception) [39] in psychology that describes how the human visual system completes parts of an object even when it is only partially visible. Although the length of a bristle line represents data magnitude, benefits from cutting the length to avoid clutter dominate the side effects from data misunderstanding that could be caused by clutter. Moreover, when using redundant encoding utilizing bristle length and density as data magnitude, bristle density could help viewers complete parts of the bristle lines. Fig. 8 shows four image pairs before and after applying our clutter reduction strategies. Some improvements could also be considered in the future. For instance, our strategies still generate cluttered bristle lines in cases where road lines are very dense or close to others. We perform experiments in Section 6 to see how people understand the differences before and after clutter reduction. Here, we note that the experiments performed were for comparison and identification tasks. In these task types, line direction (as will be shown in the experiments) had little impact on the user results. However, in a cluster/delineate task in which users are asked to segment the data, the splitting of direction may influence the user's perception of cluster boundaries. As such, we recommend that map designers take caution in employing this scheme and use it only in appropriate map contexts. Future work will explore other schemes and design issues to handle neighbor crossings and influence on map design.

6 EVALUATION

To evaluate the effectiveness of our bristle maps, we conducted two quantitative controlled experiments. These studies are both comprised of an introductory session, and a training session. In the first study, five tasks were conducted to evaluate the efficiency of bristle maps compared to existing visualization methods (point, color (kernel density estimated—KDE), and line maps as shown in Figs. 1a, 1c, and 1d) and post-task questionnaires for qualitative feedback. In the second study, two tasks were conducted to evaluate the accuracy of users in estimating values from each of the map types (point, KDE, bristle, and line) as well as evaluating the perceived aesthetics of each image. Prior to each study, a pilot study was also conducted to ensure that each task contains a fair comparison among the techniques.

Participants. In the first study, thirty graduate students (23 males, seven females) in engineering, science, and statistics from our university participated in the study. All participants reported that they had experience in visualizing data on geographical maps using colors or icons (e.g., paper maps, online map services). The experience varied from almost daily (11 participants), 1-2 times a week (17 participants) to 1-3 times a month (two participants). For the identification/accuracy tasks and aesthetic comparisons (Tasks 6 and 7), a secondary study was run on 26 undergraduate students in engineering from our university.

Apparatus. The experiment was performed on a $30''$ monitor using our experimental application running on Windows XP, as shown in Fig. 9, where all visualizations were generated with $2,228 \times 1,478$ resolution. Each visualization was overlayed with numbered circles as shown in Fig. 9. Participants selected one of the numbers to answer the question in each trial using buttons in the interface panel on the top of the screen. Criminal incident reports collected in West Lafayette and Lafayette, Indiana from 1999 to 2010 were used in each trial, but different types of crimes were selected to generate visualizations in the training phase and in the actual study.

Design. We employed a repeated measure design of tasks incorporating variations of the images shown in Figs. 1a, 1c, and 1d and line maps similar to those of
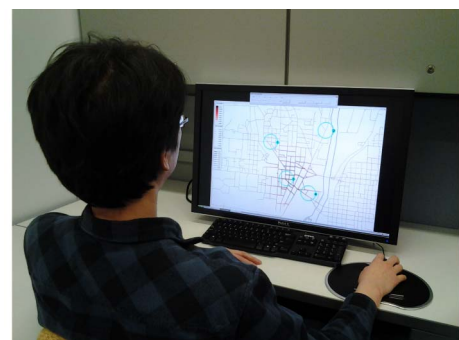

Fig. 9. Example setup for our experiment.

TABLE 2
The Number of Data Sets, Techniques
(Cases in Fig. 8 for Task 5), and Trials

	Data sets	Techniques (or cases)	Trials
Task 1	18	5	90
Task 2	18	5	90
Task 3	15	6	90
Task 4	12	4	48
Task 5	7	8	56
Task 6	2	4	24
Task 7	2	4	2

Fig. 3(right). Table 2 shows the number of data sets, techniques (cases as shown in Fig. 8 for Task 5), and trials in each task. For example, in Task 1 we utilized 18 different data sets to compare five different techniques (i.e., point map, color map, line map, and bristle maps using two different encoding schemes). Hence, each participant performed 18×5=90 trials in Task 1. In Task 3, we compared six different techniques (i.e., point map, bivariate color map, line maps in two different encoding schemes, and bristle maps in two different encoding schemes) with 15 data sets, resulting in 90 trials. Due to the difficulty of creating good examples to be used from our real crime data, we used fewer data sets in Tasks 4 and 5. In summary, each participant performed a total of 374 trials in Tasks 1 to 5, and it generally took 90 minutes.

Since the design of Tasks 1-5 focused on questions of comparing regions, a secondary study was also conducted. This study was again a repeated measure design of tasks incorporating variations of the images shown in Figs. 1a, 1c, and 1d and line maps similar to those of Fig. 3(right). However, here the subjects were asked to identify the values of regions in the image. Areas of homogeneous visual variables were circled in each image and the subjects were asked to approximate the amount of crime per region. As a final task, the subjects were simultaneously presented with a point map, color map, bristle map, and line map and asked to rank order the images based on their aesthetic values.

For all Tasks, trial order was varied using a magic square method [40] in each task. Completion time and participants' answers were recorded for a quantitative metric. The collected data from each task was subjected to an analysis of variance (ANOVA) test to determine if the average time and accuracy of task completion were significantly different among techniques. A Post-Hoc Tukey HSD test was then performed to determine significance between the techniques. P-values reported in this study come from the resultant Tukey HSD test. Before the study, participants were introduced to our experiment application and the techniques through an introductory session and a training session. During the training session, participants could ask questions and receive guidance in the use of the experiment application and analysis of each visualization. Once the training was completed, participants moved to the actual study. After completing each task (Tasks 1 to 4) participants were asked to answer the questionnaire to rate the efficiency of the techniques using a five-point Likert scale [41]. After completing Task 5, participants were also asked to describe their impression with regards to visual complexity for before and after image pairs applying our clutter reduction. In the questionnaire, we stated that the visual complexity is high if a participant felt any kind of difficulty or confusion in understanding the density, length, and color of bristle lines that encode the underlying data. Finally, after finishing all tasks, participants were asked to rate the overall efficiency among techniques.

Hypotheses. In this experiment, we hypothesized that our bristle maps would be better than or equally as good when compared to the other techniques in terms of task completion time and accuracy. Specifically, we hypothesized that our bristle maps would be better than other techniques as the complexity level of tasks increased from univariate to multivariate. The rationale of this assumption is that the line map and bivariate color map use at most two variables, whereas the several encoding parameters in our bristle map have the potential to create effective encoding combinations. We also hypothesized that our clutter reduction strategies would be useful to minimize cluttering on areas where a large number of bristle lines are created. In our follow-on experiment exploring identification of values, we hypothesized that bristle maps would be as accurate as all other representations in determining values. We also hypothesized that bristle maps would be ranked higher in terms of their aesthetics.

Tasks. We tested seven tasks: three for univariate, bivariate, and multivariate data encoding, respectively, one for temporal variance encoding, one for the clutter reduction, one for accuracy comparisons among the rendering styles, and one for aesthetic comparisons.

In Task 1, when given four regions highlighted in circles on the map, participants were asked to "find the region with the highest crime rate" in different visualizations representing spatiotemporal crime data using point, color, line-T (data encoded in the line (T)hickness), bristle-CLD (a redundant data encoding using (C)olor, (L)ength, and (D)ensity), and bristle-LD (a redundant data encoding using (L)ength and (D)ensity).

In Task 2, four regions were highlighted in circles on the map. Participants were asked to "find the region with the highest crime rates at both (or either) day and night time," using point (encoding day/night time crime rates in different colors), color, line-TO (data encoded as line (T)hickness and using (O)rientation for day/night crime rates), bristle-CLDO (redundant data encoding using (C)olor, (L)ength ,and (D)ensity, and using (O)rientation to indicate day/night crime rates), and bristle-LDO (data encoded using (L)ength and (D)ensity, but in a constant color, using (O)rientation to indicate day/night crime rates). The point map had differently colored points for day and night time crime rates, and two maps (day and night time color maps) were given in different colors for the color map.

In Task 3, four regions were highlighted in circles on the map. Participants were asked to "find the region with the highest crime rates for both (or either) two crimes (crime 1 and 2)," using point map (encoding two crimes in different colors), bivariate color map (Color-B), line-TO (a data encoding using (T)hickness in different colors, and using (O)rientation to indicate crime types), line-CT (encoding crime 1 using (C)olor and crime 2 using (T)hickness),

bristle-LDO (a redundant data encoding using (L)ength and (D)ensity, and using (O)rientation to indicate crime types), and bristle-CD (an encoding using (C)olor to indicate crime 1 and (D)ensity to indicate crime 2, with constant length).

In Task 4, participants were given two regions highlighted in circles on the map. Then, they were asked to "find the region with the highest temporal variance" in different visualizations using point maps, color maps, line maps, and bristle-LDV (a redundant data encoding using (L)ength and (D)ensity, and representing (V)ariance in the variance part of a bristle line). For the point, color, and line maps, multiple images were displayed on the screen to provide visualizations during several years. Our bristle map embedded the variance in the variance part of the bristle length as shown in Fig. 2 (third stage) and Fig. 5 (right column).

In Task 5, given two regions predefined in circles on bristle maps, participants were asked to "answer if crime rates on this given two regions look either different or the same as each other." Fig. 8 shows representative image pairs before and after applying our clutter reduction method. In trials, participants compared each case in Fig. 8 to a base case (i.e., bristle lines on a single straight road).

In Task 6, subjects were presented with a series of images with a single predefined circle, which covered an area consisting of homogeneous visual variables (i.e., identical color, bristle length, thickness, etc.). A univariate encoding was explored, and the Bristle-CLD settings were utilized for the bristle map. Participants were asked to estimate the amount of crime in the area using the provided scale (or scales in the case of bristle and line maps). Time and accuracy of the results were measured.

In Task 7, subjects were presented simultaneously with four images representing the same data set. These images consisted of a point map, a color map, a bristle map, and a line map. Subjects were asked to rank order the images in order of most to least aesthetically pleasing.

7   RESULTS AND DISCUSSION

After all tasks were completed, times and answers collected during the study were analyzed using a single-factor ANOVA. A Post-Hoc Tukey HSD test was then performed to determine significance between the techniques. P-values reported in this study come from the resultant Tukey HSD test. For accuracy, the percentage of correct answers was computed.

Task 1. A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with highest crime rates within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps, and bristle maps. There was a significant effect of visualization type on time at the $p < 0.05$ level for the conditions [$F(4, 145) = 35.366, p = 0.0000001$] and a significant effect of visualization type on accuracy at the $p < 0.05$ level for the conditions [$F(4, 145) = 3266.782, p = 0.0000000006$]. Because statistically significant results were found, we computed a Tukey posthoc test with results reported in Table 3. In Table 3, p-values < 0.05 indicate that groups were statistically different from one another.

TABLE 3
Tukey HSD Results for Task 1

Time	p-value $<$	Point map	KDE map	Line-T
	Bristle-CLD	.00001	.00001	.00042
	Bristle-LD	.00001	.00001	.01811
Accuracy	p-value $<$	Point map	KDE map	Line-T
	Bristle-CLD	.00001	.1554	.01851
	Bristle-LD	.00001	.3214	.00602

The result showed that the bristle maps groups were both significantly different than the point, color, and line maps in terms of speed (at the $p < 0.05$ level). Specifically, the bristle map groups average times were 50.7 and 56.6 seconds for the CLD and LD conditions, respectively, which was slightly faster than the Line-T condition at 69 seconds and much faster than the point map condition at 102.6 seconds. However, the color map group was the fastest at 34.6 seconds.

For accuracy, the bristle maps groups were both significantly different than the point map group in terms of accuracy (at the $p < 0.05$ level). Specifically, the bristle map groups accuracy ratings were 99.6 and 99.8 percent for the CLD and LD conditions, respectively, which was much higher than the point map condition with accuracy of 41.4 percent. No accuracy differences were found when compared to the other groups. See Table 8 for more specific results.

The comparison between color maps and bristle maps showed that color maps were better than the bristle map in terms of average time, and were not significantly different in terms of accuracy. This shows that bristle maps as a redundant encoding scheme has the same potential to convey data as single parameter encoding schemes; however, traditional schemes such as color maps may allow for a quicker comparison in the univariate case.

Comparing Bristle-LD and Line-T, we saw that the length of the bristle map matches the thickness of the line map. Hence, the bristle density was useful to find answers in Task 1 in terms of completion time and accuracy. Some participants also mentioned bristle density in their qualitative feedback as *"Bristle map is especially good when density of the bristles is also used"* and *"In bristle map, length, and density were more noticeable than color difference."* In this univariate encoding test, the point map showed the worst results and the color map was the best results in terms of time and accuracy as shown in Table 8.

Task 2. A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with highest crime rates at both day and nighttime within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps and bristle maps. There was a significant effect of visualization type on time at the $p < 0.05$ level for the conditions [$F(4, 145) = 2.717, p = 0.032$] and a significant effect of visualization type on accuracy at the $p < 0.05$ level for the conditions [$F(4, 145) = 89.89, p = 0.0000002$]. Because statistically significant results were found, we

TABLE 4
Tukey HSD Results for Task 2

Time	p-value <	Point map	KDE map	Line-TO
	Bristle-CLDO	**.01713**	.70091	.05943
	Bristle-LDO	**.02024**	.81621	.07166

Accuracy	p-value <	Point map	KDE map	Line-TO
	Bristle-CLDO	**.00001**	.07062	.36692
	Bristle-LDO	**.00001**	.99999	**.01283**

TABLE 5
Tukey HSD Results for Task 3

Time	p-value <	Point map	KDE-B	Line-TO	Line-CT
	Bristle-LDO	**.00009**	**.00515**	.58128	.73239
	Bristle-CD	**.01131**	**.03469**	.15506	.20693

Accuracy	p-value <	Point map	KDE-B	Line-TO	Line-CT
	Bristle-LDO	**.00001**	**.00001**	.27189	**.02771**
	Bristle-CD	**.00001**	**.00001**	.07002	.41194

computed a Tukey posthoc test with results reported in Table 4. In Table 4, p-values < 0.05 indicate that groups were statistically different from one another.

As we hypothesized, the result showed that the bristle maps groups were both significantly different than the point maps in terms of speed (at the $p < 0.05$ level). Specifically, the bristle map groups average times were 86.3 and 87.2 seconds for the CLDO and LDO conditions, respectively, which was slightly faster than the point map condition at 106.2 seconds.

For accuracy, the bristle maps groups were both significantly different than the point map group in terms of accuracy (at the $p < 0.05$ level). Specifically, the bristle map groups accuracy ratings were 90.5 and 93.3 percent for the CLDO and LDO conditions, respectively, which was much higher than the point map condition with accuracy of 63.1 percent. See Table 8 for more specific results.

The comparison between color maps and bristle maps showed that color maps were better than the bristle map in terms of average time, and were not significantly different in terms of accuracy. This shows that bristle maps as a reduandant encoding scheme has the same potential to convey data as single parameter encoding schemes; however, traditional schemes such as color maps may allow for a quicker comparison in the univariate case.

Findings also indicated that Bristle-LDO was better than Line-TO in terms of accuracy, whereas Bristle-CLDO was not significantly different from Line-TO in terms of accuracy. This indicated that the bristle density seems to be useful in finding correct answers in Bristle-LDO, but it was not in Bristle-CLDO. Further testing in combinations of visual variables and the ability to determine levels of sparseness will be done in the future.

Task 3. A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with highest crime rates in two types of crimes within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps, and bristle maps. There was a significant effect of visualization type on time at the $p < 0.05$ level for the conditions $[F(5, 174) = 6.655, p = 0.00001]$ and a significant effect of visualization type on accuracy at the $p < 0.05$ level for the conditions $[F(5, 175) = 144.24, p = 0.00000001]$. Because statistically significant results were found, we computed a Tukey posthoc test with results reported in Table 5. In Table 5, p-values < 0.05 indicate that groups were statistically different from one another.

The result showed that the bristle maps groups were both significantly different than the point maps and color maps in terms of speed (at the $p < 0.05$ level). Specifically, the bristle map groups average times were 88.2 and 94.5 seconds for the LDO and CD conditions, respectively, which was faster than the point map condition at 118.3 seconds and the color map condition at 115.3 seconds.

For accuracy, the bristle maps groups were both significantly different than the point map group and the color map group in terms of accuracy (at the $p < 0.05$ level). Specifically, the bristle map groups accuracy ratings were 94.4 and 90.4 percent for the LDO and CD conditions, respectively, which was much higher than the point map condition with accuracy of 26.6 percent and the color map condition with accuracy of 72.6 percent. See Table 8 for more specific results.

Note that we separated parameters for different crime types in Bristle-CD: (C)olor encodes crime 1 and (D)ensity encodes crime 2. Bristle-CD showed a significant effect compared to the bivariate color map as shown in Table 5. However, generation on this type of bristle maps should be selected carefully because one parameter could dominate the other. For instance, when we use color and length to separate two crime data, short bristle length for low crime rates in crime 2 removes bristle lines in dark color for high crime rates in crime 1. In our experiment, we selected color and density for two crimes, with constant length of bristles.

Task 4. A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with high temporal variance within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps, and bristle maps. There was a significant effect of visualization type on time at the $p < 0.05$ level for the conditions $[F(3, 116) = 42.051, p = 0.00001]$ and a significant effect of visualization type on accuracy at the $p < 0.05$ level for the conditions $[F(3, 116) = 42.33, p = 0.00001]$. Because statistically significant results were found, we computed a Tukey posthoc test with results reported in Table 6. In Table 6, p-values < 0.05 indicate that groups were statistically different from one another.

The result showed that the bristle maps groups were both significantly different than the point maps, line maps and color maps in terms of speed (at the $p < 0.05$ level). Specifically, the bristle map groups average time was 48.4 seconds for the LDV condition, which was faster than the point map condition at 194 seconds, the color map

TABLE 6
Tukey HSD Results for Task 4

Time	p-value $<$	Point maps	KDE maps	Line maps
	Bristle-LDV	**.00001**	**.00001**	**.00001**
Accuracy	p-value $<$	Point maps	KDE maps	Line maps
	Bristle-LDV	**.00001**	**.00001**	**.00001**

TABLE 7
Average Rank Ordering by Aesthetics

	Point	KDE Map	Bristle	Line
Average	2.26	2.6	2.79	2.34
Std Dev	1.18	.97	1.19	1.10

TABLE 8
Average Time and Accuracy

	Technique	Average time (seconds)	Accuracy (%)
Task 1	Point map	102.6	41.4
	KDE map	**34.6**	**100**
	Line-T	69	98
	Bristle-CLD	50.7	99.6
	Bristle-LD	56.6	99.8
Task 2	Point map	106.2	63.1
	KDE map	90	93.1
	Line-TO	100.5	87.9
	Bristle-CLDO	**86.3**	90.5
	Bristle-LDO	87.2	**93.3**
Task 3	Point map	118.3	26.6
	KDE-B	115.3	72.6
	Line-TO	**84**	**96.4**
	Line-CT	86.1	86.8
	Bristle-LDO	88.2	94.4
	Bristle-CD	94.5	90.4
Task 4	Point maps	194	53.6
	KDE maps	171.8	61.9
	Line maps	178.9	75.5
	Bristle-LDV	**48.4**	**94.7**

condition at 171.8 seconds, and the line map condition at 178.9 seconds.

For accuracy, the bristle maps groups were both significantly different than the point maps, line maps, and color maps in terms of speed (at the $p < 0.05$ level). Specifically, the bristle map groups accuracy rating was 94.7 percent for the LDV condition, which was much higher than the point map condition with accuracy of 53.6 percent, the color map condition with accuracy of 72.6 percent, and the line map condition with accuracy of 75.5 percent. See Table 8 for more specific results.

As we hypothesized, we found that the representation of temporal variance in bristle maps was significantly faster and accurate in terms of both average time and accuracy compared to providing several images of the point, color, and line maps. Moreover, we found that techniques showed the increasing pattern from the point maps to Bristle-LDV as shown in Table 8. This indicates that changes among several images would be better perceived in line patterns than in points or colors.

Task 5. A one-way between-subjects ANOVA was conducted to compare the effect of different map visualizations on a subject's time and accuracy in determining areas with high temporal variance within a given visualization. Conditions varied based on the given visualization, point maps, kernel density estimated color maps, line maps, and bristle maps. There was no significant effect of visualization type on time at the $p < 0.05$ level for the conditions $[F(1, 56) = 0.328, p = 0.569]$ and no significant effect of visualization type on accuracy at the $p < 0.05$ level for the conditions $[F(1, 56) = 0.315, p = 0.315]$. In Task 5, we found that bristle lines with and without clutter reduction did not differ significantly w.r.t. both average time and accuracy for all cases (Fig. 8). This means that the base bristle lines and bristle lines before applying clutter reduction and the base and bristle lines after applying our clutter reduction are perceived similarly by participants. Moreover, when told that the bristle line orientation does not encode data, the opposite orientations of bristle lines on a single straight road caused by virtual nodes (Fig. 7b) did not affect accuracy (87.7 percent). Other cases showed 42-58 percent of accuracy.

Task 6. For Task 6, we hypothesized that subjects would be as accurate as all other representations in determining

values. In Task 6, we found that bristle maps did not differ significantly w.r.t. accuracy when compared with point map, color map, and line map identification (ANOVA results of p-value = 0.18093, $F = 1.63$). However, we found that bristle maps did differ significantly w.r.t. time when compared with point map, color map and line map (ANOVA results of p-value = 0.0314, $F = 2.622$). Particularly, we found line maps and heat maps to both be significantly faster than point maps and bristle maps in identifying values (Tukey HSD test value of $p < 0.05$). Overall, these results indicate that in terms of accuracy, all geographical representations were equally useful; however, participants were (on average) over 1 second quicker in value judgments on both line maps and colors maps. This is most likely due to the fact that participants were quicker at making color judgments as compared to counting points and mentally linking multiple variables for the bristle maps.

Task 7. In Task 7, we found that users had a highly variable rating of which image appeared to be more aesthetically pleasing. The average positions and standard deviations are summarized in Table 7. Here, we find that while bristle maps have a slightly higher average ranking, there is no significant difference between the aesthetic ordering. A one-way between-subjects ANOVA was conducted to compare the rankings of map visualizations by subject in determining which visualization was ranked highest in aesthetics. There was no significant effect of visualization type on aesthetics at the $p < 0.05$ level for the conditions $[F(3, 183) = 1.79, p = 0.149]$.

Qualitative evaluation. Fig. 10 shows the results from qualitative feedback. Among the 30 participants, 27 participants (90 percent) agreed or strongly agreed that the bristle map was efficient for day and night time comparison in Task 2, 26 for two color maps and 23 for line map. Twenty-four participants (80 percent) agreed or strongly agreed that the bristle map was efficient for the comparison of two crimes in Task 3, 26 for the line map and 19 for the
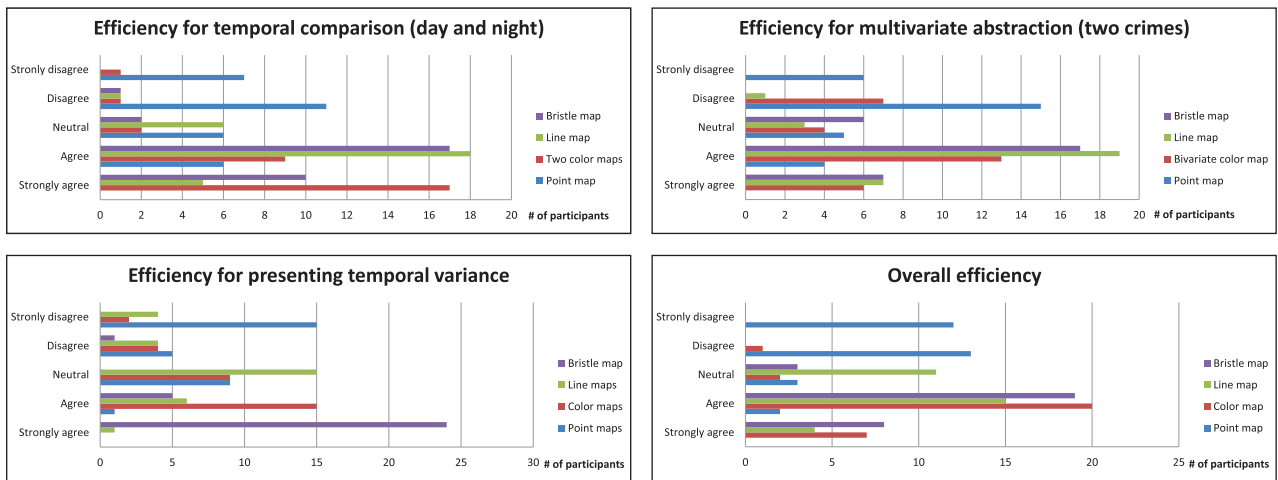
Fig. 10. Results from qualitative feedback for Tasks 2, 3, 4, and overall efficiency.

bivariate map. Twenty-nine participants (96.6 percent) agreed or strongly agreed that the bristle map was efficient for temporal variance representation. In the question for overall efficiency, 27 participants (90 percent) agreed or strongly agreed that bristle maps and color maps were overall efficient, and 19 (63.3 percent) for line maps. For point maps, 25 participants (83.3 percent) disagreed or strongly disagreed.

Participants were also asked to answer visual complexity and preference questions regarding the before (NCR) and after (CR) image pairs applying our clutter reduction. For the circular case (Fig. 8c), 96.5 percent of participants felt that NCR has higher visual complexity and 78.5 percent preferred CR. For the curved case (Fig. 8d), 65.5 percent of participants answered that CR has a higher visual complexity and 64 percent preferred NCR. While both cases use a technically identical clutter reduction algorithm, participants reported different visual complexity and preference for them. This indicates that our clutter reduction could be improved by considering the complexity of the underlying network structure.

Summary and Limitations. As a univariate encoding, the bristle maps were significantly different (in terms of speed and accuracy) than the point, color, and line maps. In the case of the point and line maps, bristle maps use resulted in a higher average correctness and speed; however, the color map for the univariate case had the fastest response and accuracy totals. This seems to indicate that the redundant encoding scheme is actually not beneficial in these cases. As such, use of bristle maps for single variable encoding is not recommended.

With regards to bivariate and multivariate encoding, bristle maps and line maps outperformed color and point maps. This is not surprising as bristle and line maps are able to combine variables into a single image, whereas in the case of point and color maps, the user must mentally combine the two images together. Bristle-(C)LD also showed a significant effect of the bristle density compared to Line-T. As a bivariate encoding, using orientation in bristle maps was not significant compared to two color maps. However, in the comparison with the bivariate color map, Bristle-LDO showed a significant effect in terms of average time and accuracy. As such, we have that Bristle-(C)LDO as a

bivariate encoding scheme created a middle level of cognitive load in-between two color maps and a bivariate color map. Bristle maps also showed potential as a multivariate encoding technique in a single view. Based on the results in Task 3, a point map using various colors and a multivariate color map would considerably increase users' cognitive load. In Tasks 1-3, we also observed that there is no significant effect between the bristle maps using the different encodings. The representation of temporal variance in the bristle map was significantly different from other methods. Our results also showed the differences among point, color, and line maps. Participants could better find the region with higher temporal variance when using line maps than using point and color maps. In the qualitative evaluation, 90 percent of the participants agreed or strongly agreed the overall efficiency of bristle maps to find answers. However, users also strongly preferred the color map in these cases as well.

Finally, we found that with regards to accuracy in identifying values, no technique outperformed any others. However, users were significantly faster in identifying values in both the color and line map scenarios. We hypothesize that in both cases the user focused only on the color, whereas in the point map case they needed to count the points and in the bristle map case they needed to reconfirm the univariate value by double checking several of the encoding legends.

Overall, this technique would be recommended when encoding large amounts of multivariate spatiotemporal point data. As the number of point samples increase, aggregation techniques are need to allow for quick summaries of the data, and, as is evidenced by our studies, pure spatial location representation by glyphs results in too much overlap for accurate measurement and evaluation. As the number of variables increase, color map representations allow for the encoding of variables only along a single visual variable (resulting in bivariate color maps or small multiple plots).

In using multivariate encodings, it is extremely important to understand the interaction effects that the visual variables will introduce in one another. Research into the perceptual interactions among different visual variables was performed by Acevedo and Laidlaw [42]. They

measured the perceptual interference of icon size, spacing, and brightness, noting that brightness outperforms spacing and size while being subject to interferences from both spacing and size. Acevedo and Laidlaw also noted that spacing also outperformed size, which contradicted some previous results; however, this result seems to align with our participants noting that the bristle spacing was a useful cue. Their results were reportedly due to the spacing sampling along a sinusoidal curve. The sampling of our bristles follows a uniform pattern within classification bins. Thus, there seems to be sufficient scientific evidence to justify using sparsity as a discriminating variable in the case of the bristle maps; however, further studies on this are warranted. Stone [43] has also studied the effect of size in color perception, noting that color appearance changed dramatically with the size being viewed. As such, it may be better to utilize fewer map classifications (color bins) when using bristle maps to increase the perceptual distance between each color being visualized.

The main limitations of the bristle map technique is that the combinations of data encoding can potentially prove overwhelming for the designer, and a poor choice on variable encoding can result in a suboptimal visualization. In particular, previous studies have provided results that can be used to predict that certain combinations of visual variables will either enhance or impede map reading. For example, the combination of length and density form an emergent property akin to Bertin's definition of grain. Such effects cannot be ignored; however, bristle maps can be encoded to take advantage of such combinations, as shown in Tasks 3 and 4.

Finally, with regards to scalability of the bristle map technique, in areas of dense roadways, different aggregation methods would need to be considered. As the roads become dense, the ability to plot lines of perceptually different length would become untenable. However, a solution to this would be to draw only the most important roads, thereby removing smaller roads from the analysis, or utilizing bristle maps in a focus+context manner.

8 CONCLUSIONS

In this work, we have described our novel multivariate data encoding scheme, the Bristle Map. This scheme provides a novel approach for encoding color, length, density, and orientation as data variables and allowing the user to explore correlations within and between variables on a single view. Given the number of parameters available within this encoding, this article has presented only a subset of potential encodings and examples. Here, we have shown the use of encoding bristle lines with redundant information, multivariate attributes for variable comparison, and temporal variance. We also showed a means of potentially encoding data uncertainty. To minimize overlap of bristle lines, we generated a topology graph from underlying geographical line features and employed strategies for clutter reduction. Then, to evaluate the effectiveness of bristle maps, we performed an evaluation study, where we explored different visual encoding combinations within the bristle maps and compared with existing techniques in several tasks. Based on our experiment results, we believe that our bristle map technique has much potential to increase the amount

of information that can be visualized on a single map for geovisualization.

REFERENCES

[1] A. MacEachren, D. Xiping, F. Hardisty, D. Guo, and G. Lengerich, "Exploring High-D Spaces with Multiform Matrices and Small Multiples," *Proc. IEEE Symp. Information Visualization (InfoVis)*, pp. 31-38, 2003, IEEE CS, doi: 10.1109/INFVIS.2003.1249006.

[2] C. North and B. Shneiderman, "Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata," *Proc. Working Conf. Advanced Visual Interfaces (AVI)*, pp. 128-135, 2000, ACM, doi: 10.1145/345513.345282.

[3] C. Weaver, "Cross-Filtered Views for Multidimensional Visual Analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 16, no. 2, pp. 192-204, Mar./Apr. 2010, doi: 10.1109/TVCG.2009.94.

[4] D.S. Ebert, R.M. Rohrer, C.D. Shaw, P. Panda, J.M. Kukla, and D.A. Roberts, "Procedural Shape Generation for Multi-Dimensional Data Visualization," *Computers & Graphics*, vol. 24, no. 3, pp. 375-384, June 2000, doi: 10.1016/S0097-8493(00)00033-9.

[5] S. Bachthaler and D. Weiskopf, "Continuous Scatterplots," *IEEE Trans. Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1428-1435, Nov./Dec. 2008, doi: 10.1109/TVCG.2008.119.

[6] R. Maciejewski, S. Rudolph, R. Hafen, A.M. Abusalah, M. Yakout, M. Ouzzani, W.S. Cleveland, S.J. Grannis, and D.S. Ebert, "A Visual Analytics Approach to Understanding Spatiotemporal Hotspots," *IEEE Trans. Visualization and Computer Graphics*, vol. 16, no. 2, pp. 205-220, Mar./Apr. 2010, doi: 10.1109/TVCG.2009.100.

[7] J.J. van Wijk and A. Telea, "Enridged Contour Maps," *Proc. Conf. Visualization (VIS)*, pp. 69-74, 2001, IEEE CS, doi: 10.1109/VISUAL.2001.964495.

[8] R.J. Phillips and L. Noyes, "An Investigation of Visual Clutter in the Topographic Base of a Geological Map," *Cartographic J.*, vol. 19, no. 2, pp. 122-132, Dec. 1982, doi: 10.1179/000870482787073225.

[9] S. Openshaw, "The Modifiable Areal Unit Problem," *Concepts and Techniques in Modern Geography*, vol. 38, Geo Books, 1984.

[10] M. Swink and C. Speier, "Presenting Geographic Information: Effects of Data Aggregation, Dispersion, and Users' Spatial Orientation," *Decision Sciences*, vol. 30, no. 1, pp. 169-195, Jan. 1999, doi: 10.1111/j.1540-5915.1999.tb01605.x.

[11] C.L. Eicher and C.A. Brewer, "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation," *Cartography and Geographic Information Science*, vol. 28, no. 2, pp. 125-138, Apr. 2001, doi: 10.1559/152304001782173727.

[12] R. Spence, *Information Visualization*. Addison-Wesley, 2001.

[13] L. Wilkinson, *The Grammar of Graphics*, second ed. Springer-Verlag, 2005.

[14] A.M. MacEachren, *How Maps Work: Representation, Visualization, and Design*. Guilford Press, 1995.

[15] S. Chainey, L. Tompson, and S. Uhlig, "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime," *Security J.*, vol. 21, no. 1/2, pp. 4-28, Feb.-Apr. 2008, doi: 10.1057/palgrave.sj.8350066.

[16] B.W. Silverman, *Density Estimation for Statistics and Data Analysis (Monographs on Statistics and Applied Probability)*, vol. 26, Chapman and Hall, 1986.

[17] C. Ahlberg and B. Shneiderman, "Visual Information Seeking Using the FilmFinder," *Proc. Conf. Companion Human Factors in Computing Systems (CHI)*, pp. 433-434, 1994, ACM, doi: 10.1145/259963.260431.

[18] Y.-H. Fua, M.O. Ward, and E.A. Rundensteiner, "Structure-Based Brushes: A Mechanism for Navigating Hierarchically Organized Data and Information Spaces," *IEEE Trans. Visualization and Computer Graphics,* vol. 6, no. 2, pp. 150-159, Apr.-June 2000, doi: 10.1109/2945.856996.

[19] A. Dix and G. Ellis, "By Chance: Enhancing Interaction with Large Data Sets through Statistical Sampling," *Proc. Working Conf. Advanced Visual Interfaces (AVI),* pp. 167-176, 2002, ACM, doi: 10.1145/1556262.1556289.

[20] A. MacEachren, "Visualizing Uncertain Information," *Cartographic Perspectives,* vol. 13, pp. 10-19, 1992.

[21] R. Dunn, "A Dynamic Approach to Two-Variable Color Mapping," *The Am. Statistician,* vol. 43, no. 4, pp. 245-252, Nov. 1989, doi: 10.1080/00031305.1989.10475669.

[22] J. Olson, "Spectrally Encoded Two-Variable Maps," *Annals Assoc. Am. Geographers,* vol. 71, no. 2, pp. 259-276, June 1981, doi: 10.1111/j.1467-8306.1981.tb01352.x.

[23] A. MacEachren and D. DiBiase, "Animated Maps of Aggregate Data: Conceptual and Practical Problems," *Cartography and Geographic Information Systems,* vol. 18, no. 4, pp. 221-229, Oct. 1991, doi: 10.1559/152304091783786790.

[24] H. Hagh-Shenas, S. Kim, V. Interrante, and C. Healey, "Weaving Versus Blending: A Quantitative Assessment of the Information Carrying Capacities of Two Alternative Methods for Conveying Multivariate Data with Color," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1270-1277, Nov./Dec. 2007, doi: 10.1109/TVCG.2007.70623.

[25] T. Saito, H.N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda, "Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data," *Proc. IEEE Symp. Information Visualization (InfoVis),* pp. 173-180, 2005, IEEE CS, doi: 10.1109/INFOVIS.2005.35.

[26] M. Sips, J. Schneidewind, D.A. Keim, and H. Schumann, "Scalable Pixel-Based Visual Interfaces: Challenges and Solutions," *Proc. 10th Int'l Conf. Information Visualization (IV),* pp. 32-38, 2006, IEEE CS, doi: 10.1109/IV.2006.95.

[27] C. Panse, M. Sips, D. Keim, and S. North, "Visualization of Geo-Spatial Point Sets via Global Shape Transformation and Local Pixel Placement," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 749-756, Sep./Oct. 2006, doi: 10.1109/TVCG.2006.198.

[28] D. Dorling, A. Barford, and M. Newman, "Worldmapper: The World as You've Never Seen It Before," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 757-764, Sep./Oct. 2006, doi: 10.1109/TVCG.2006.202.

[29] P.C. Wong, K. Schneider, P. Mackey, H. Foote, G. Chin, R. Guttromson, and J. Thomas, "A Novel Visualization Technique for Electric Power Grid Analytics," *IEEE Trans. Visualization and Computer Graphics,* vol. 15, no. 3, pp. 410-423, May/June 2009, doi: 10.1109/TVCG.2008.197.

[30] D. Fisher, "Hotmap: Looking at Geographic Attention," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1184-1191, Nov./Dec. 2007, doi: 10.1109/TVCG.2007.70561.

[31] C. Tominski, P. Schulze-Wollgast, and H. Schumann, "3D Information Visualization for Time Dependent Data on Maps," *Proc. Ninth Int'l Conf. Information Visualization (InfoVis),* pp. 175-181, 2005, IEEE CS, doi: 10.1109/IV.2005.3.

[32] J. Bertin, *Semiology of Graphics.* ESRI Press, 2011.

[33] J. Tarbell, "Substrate," *Web Site & Simulation,* http://www.complexification.net/gallery/machines/substrate/, 2003, Feb. 2012.

[34] T. Isenberg, "Visual Abstraction and Stylisation of Maps," *Cartographic J.,* vol. 50, no. 1, pp. 8-18, Feb. 2013, doi: 10.1179/1743277412Y.0000000007.

[35] P. Rheingans, "Task-Based Color Scale Design," *Proc. SPIE,* vol. 3905, pp. 35-43, 2000, SPIE, doi: 10.1117/12.384882.

[36] C. Ware, "Color Sequences for Univariate Maps: Theory, Experiments and Principles," *IEEE Computer Graphics and Applications,* vol. 19, no. 5, pp. 41-49, Sep./Oct. 1988, doi: 10.1109/38.7760.

[37] C.A. Brewer, *Designing Better Maps: A Guide for GIS Users.* ESRI Press, 2005.

[38] M. Prasad, "Intersection of Line Segments," *Graphics Gems II,* J. Arvo, ed., pp. 7-9, Academic Press, 1991.

[39] A. Michotte, G. Thinès, and G. Crabbé, *Les Complements Amodeux des Structures Perceptives (Amodal Completions of Perceptual Structures),* Louvain: Institut de Psychologie del'Université de Louvain, France: Studia Psychologica, 1964.

[40] M.S. Farrar, *Magic Squares.* BookSurge Publishing, 1996.

[41] R.A. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology,* vol. 22, no. 140, pp. 5-55, 1932.

[42] D. Acevedo and D. Laidlaw, "Subjective Quantification of Perceptual Interactions among Some 2D Scientific Visualization Methods," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 1133-1140, Sept. 2006, doi: 10.1109/TVCG.2006.180.

[43] M. Stone, "In Color Perception, Size Matters," *IEEE Computer Graphics & Applications,* vol. 32, no. 2, pp. 8-13, Mar./Apr. 2012, doi: 10.1109/MCG.2012.37.
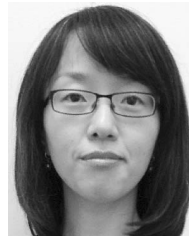
SungYe Kim received the master's degree in computer science and engineering from Chung-Ang University, South Korea in 2000, and the PhD degree in electrical and computer engineering from Purdue University in May, 2012. She is currently a graphics software engineer at Intel Corporation. Prior to this, she was employed as a research engineer at the Electronics and Telecommunications Research Institute from 2000 to 2006. Her research interests include computer graphics, illustrative visualization, visual analytics, and information visualization.

Ross Maciejewski received the PhD degree in electrical and computer engineering from Purdue University in December, 2009. He is currently an assistant professor at Arizona State University in the School of Computing, Informatics & Decision Systems Engineering. Prior to this, he served as a visiting assistant professor at Purdue University and was at the Department of Homeland Security Center of Excellence for Command Control, and Interoperability in the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) group. His research interests include geovisualization, visual analytics, and nonphotorealistic rendering. He is a member of the IEEE.

Abish Malik received the BS degree in electrical engineering from Purdue University in 2009, and is currently working toward the PhD degree in the School of Electrical and Computer Engineering at Purdue University. He is a research assistant at the Purdue University Rendering and Perception Lab. His research interests include visual analytics, correlation, and predictive data analytics.
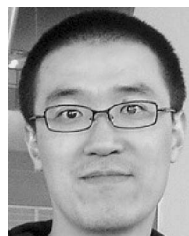
Yun Jang received the bachelor's degree in electrical engineering from Seoul National University, South Korea in 2000, and the master's and doctoral degree in electrical and computer engineering from Purdue University in 2002 and 2007, respectively. He is an assistant professor of computer engineering at Sejong University, Seoul, South Korea. He was a postdoctoral researcher at CSCS and ETH Zürich, Switzerland from 2007 to 2011. His research interests include interactive visualization, volume rendering, visual analytics, and data representations with functions. He is a member of the IEEE.

David S. Ebert received the PhD degree in computer science from Ohio State University. He is a professor in the School of Electrical and Computer Engineering at Purdue University, the University Faculty scholar, the director of the Purdue University Rendering and Perceptualization Lab, and the director of the Purdue University Regional Visualization and Analytics Center. His research interests include novel visualization techniques, visual analytics, volume rendering, information visualization, perceptually based visualization, illustrative visualization, and procedural abstraction of complex, massive data. He is a fellow of the IEEE and a member of the IEEE Computer Society's Publications Board.

Tobias Isenberg received the doctoral degree from the University of Magdeburg, Germany. He is a senior research scientist with INRIA in France. Previously, he held positions as assistant professor for computer graphics and interactive systems at the University of Groningen, the Netherlands, and as a postdoctoral fellow at the University of Calgary, Canada. He works on topics in interactive nonphotorealistic and illustrative rendering as well as computational aesthetics and explores applications in scientific visualization. He is a member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

Comparison of Known Food Weights with Image-Based Portion-Size Automated Estimation and Adolescents' Self-Reported Portion Size

Christina D. Lee,[1] Junghoon Chae, B.S.,[2] TusaRebecca E. Schap, M.Sc., R.D.,[3]
Deborah A. Kerr, Ph.D.,[4] Edward J. Delp, Ph.D.,[2] David S. Ebert, Ph.D.,[2]
and Carol J. Boushey, Ph.D., M.P.H., R.D.[3,5]

Abstract

Background:

Diet is a critical element of diabetes self-management. An emerging area of research is the use of images for dietary records using mobile telephones with embedded cameras. These tools are being designed to reduce user burden and to improve accuracy of portion-size estimation through automation. The objectives of this study were to (1) assess the error of automatically determined portion weights compared to known portion weights of foods and (2) to compare the error between automation and human.

Methods:

Adolescents ($n = 15$) captured images of their eating occasions over a 24 h period. All foods and beverages served were weighed. Adolescents self-reported portion sizes for one meal. Image analysis was used to estimate portion weights. Data analysis compared known weights, automated weights, and self-reported portions.

Results:

For the 19 foods, the mean ratio of automated weight estimate to known weight ranged from 0.89 to 4.61, and 9 foods were within 0.80 to 1.20. The largest error was for lettuce and the most accurate was strawberry jam. The children were fairly accurate with portion estimates for two foods (sausage links, toast) using one type of estimation aid and two foods (sausage links, scrambled eggs) using another aid. The automated method was fairly accurate for two foods (sausage links, jam); however, the 95% confidence intervals for the automated estimates were consistently narrower than human estimates.

Conclusions:

The ability of humans to estimate portion sizes of foods remains a problem and a perceived burden. Errors in automated portion-size estimation can be systematically addressed while minimizing the burden on people. Future applications that take over the burden of these processes may translate to better diabetes self-management.

J Diabetes Sci Technol 2012;6(2):428-434

Author Affiliations: [1]College of Liberal Arts, Purdue University, West Lafayette, Indiana; [2]College of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana; [3]Department of Nutrition Science, Purdue University, West Lafayette, Indiana; [4]School of Public Health, Curtin University, Perth, Western Australia; [5]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii

Corresponding Author: Carol J. Boushey, Ph.D., M.P.H., R.D., Epidemiology Program, University of Hawaii Cancer Center, 1236 Lauhala Street, Suite 407G, Honolulu, HI 96813; email address *cjboushey@cc.hawaii.edu*

Comparison of Known Food Weights with Image-Based Portion-Size
Automated Estimation and Adolescents' Self-Reported Portion Size

Lee

Introduction

In 2010, there were approximately 25.8 million people or 8.3% of the U.S. population affected by diabetes.[1] Noninsulin-dependent diabetes, adult-onset diabetes, or type II diabetes that is now referred to as type 2 diabetes composes about 90–95% of those with diabetes.[2] Currently, type 2 diabetes is an epidemic in the United States and worldwide,[3,4] affecting both children and adults largely due to the rising rates of obesity.[5,6] There are many long-term complications that come with diabetes, such as retinopathy, nephropathy, and peripheral neuropathy, which are associated with high medical costs.[2,7] More importantly, diabetes was the seventh leading cause of death by diseases in 2010.[1]

Diabetes self-management is an integral component of diabetes care and requires considerable attention to diet, exercise, and medication use.[1,8–10] For most persons with diabetes, managing diet and exercise routines can be challenging. Therefore, to enhance self-management of diabetes, industry and researchers have turned to technology. A growing body of literature suggests that using information technology such as computers, the Internet, multimedia, and mobile devices could improve and positively affect the process of care for people with diabetes.[11,12]

Diet is a critical element of self-management and probably the most challenging to assess and monitor.[13] To date, few mobile technologies that address dietary adherence have been developed.[8–11,14] Individuals with diabetes described recording food intake using a smartphone, personal computer, and a blog as motivating; however, several usability problems were identified as reducing the likelihood of sustained use of these tools.[14] Specifically, the steps needed to identify the foods and estimate the portion and serving sizes were described as limited and lacking flexibility.[14] An emerging area of research is in the use of images for dietary records[15,16] using small mobile devices with embedded cameras (e.g., mobile telephone). These tools are being designed to identify foods and beverages and their portion sizes accurately through automation.[17] The objectives of this article are to (1) assess the error surrounding the mean estimate of automatically determined portion weights based on images taken by adolescents compared with known portion weights of selected foods and beverages, (2) compare the error between automation and human estimation of selected foods, and (3) envision future application of these automated, image-based methods for diabetes management and prevention.

Methods

Recruitment and Study Design

A convenience sample of healthy adolescents between 11 and 18 years was recruited from the local community.[16,17] On the day of the study, participants were transported to a university campus early in the morning prior to consuming any food or beverages. The participants were served all meals at set times and snacks were provided *ad libitum* over a 24 h period while being closely monitored.[18] Between eating occasions, camp-like activities were provided. At the end of the day, all individuals estimated the portion sizes of their breakfast foods. The study methods were approved by the Purdue University Institutional Review Board, and informed assent and consent were obtained from the volunteers and their parents, respectively.

Eating Occasions and Use of the Mobile Telephone Application

Foods served represented common foods reported by adolescents.[19,20] For each eating occasion, all foods and beverages were preweighed separately to one-tenth of a gram prior to plating.[18] Participants received instruction for using a mobile telephone to capture images of each eating occasion. In order to obtain an image useful for image analysis, participants were instructed to include in each image (1) all food and beverages and (2) the fiducial marker, a small credit card-sized item used for color correction and volume (**Figure 1A**). Participants were instructed to eat to satiation and to request seconds, if desired. The procedures for capturing images were then repeated for any additional portions.

HTC p4351 mobile telephones (HTCAmerica, Bellevue, WA) running Windows Mobile 6.0 (Microsoft, Redmond, WA) were used.[16,21,22] The user was given a choice to retake the image or save the image. Once the user was satisfied with the image, the mobile telephone prompted the user to eat before proceeding to the next screen as shown in **Figure 1A**. After eating, the user was prompted to take an image of the place setting regardless of whether food and beverages remained. **Figure 1B** is an example of the final screen with before-and-after images.

Comparison of Known Food Weights with Image-Based Portion-Size
Automated Estimation and Adolescents' Self-Reported Portion Size

Lee

Self-Reported Portion-Size Estimation of Breakfast Foods

While viewing an image of their breakfast meal, participants were asked to estimate the amount of each food item consumed 14 h after the breakfast meal.[17] Numerous methods for estimating portion size are available so at least two of the most common methods of portion-size estimation were used: (1) multiple measurement descriptors (MDes) pertinent to each specific food from the What's In The Foods You Eat *Search Tool*, 3.0. (*http://www.ars.usda.gov/Services/docs.htm?docid=17032*) and (2) two-dimensional (2D) food portion visual with 2D images of standard-sized plates and bowls with cubes depicting ¼ cup, ½ cup, 1 cup, and 2 cups (Block Dietary Data Systems, Berkeley, CA; *http://www.nutritionquest.com*). Self-reporting of beverage portions was not done. Participants were randomly divided into two groups. One group (*n* = 8) used the 2D portion estimation aid and the other group (*n* = 7) used the MDes portion estimation aid. For portion evaluation, the self-reported intake of each food was converted into grams.

Automated Portion-Size Estimation of Foods and Beverages Served

Images from the mobile telephone were sent to a server for image analysis. Methods for automatic identification of food using image analysis have been described previously.[22] Methods used to automatically estimate volume of foods served using computer algorithms have been previously described.[23,24] Briefly, once a food was identified automatically, volume was estimated using camera location, orientation, and other parameters for use in 3D reconstruction of the food volume from an image. A fiducial marker (e.g., checkerboard square in **Figure 1B**) was used for size and spatial location of food on the plate.

The system partitioned the space of the food objects into two geometric shapes, cylinders and squares, each with their own set of parameters. The spherical approximation models drew upon spheres and prismatic approximation models. For the foods and beverages served at meals, the automated volume was estimated as cubic centimeters and converted into weight (g) using density values derived from rapeseed volumeter measures of duplicate plates of each meal.[18,25]

Data Analysis

Means of the gram weights of each food and beverage actually served during meals and consumed were computed. Means of the estimated portions, i.e., automated and
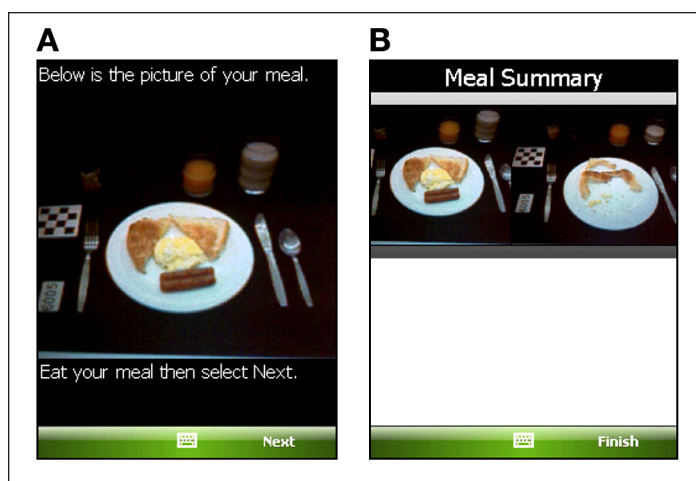
Figure 1. Screen shots from mobile telephone. **(A)** Example of a before image. **(B)** Example of end-of-meal screen showing before and after meal.
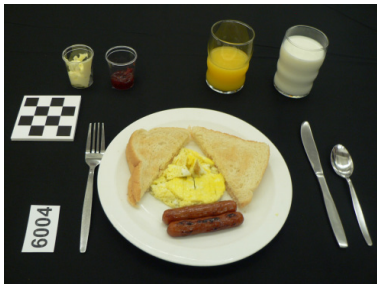
self-report, were computed. Accuracy of the portion-size estimates was assessed by plotting the mean of the ratio of the estimates, i.e., automated and self-reports, against the actual weights of the foods and beverages, as well as, the 95% confidence intervals (CIs) of the ratios. Thus, a ratio >1 would indicate overestimation and a ratio <1 would indicate underestimation. Descriptive analysis included frequencies and percents. When needed, differences between quantitative variables were assessed using a paired *t*-test. All analyses were conducted using SPSS 17.0.

Results

A total of 15 adolescents (3 girls, 12 boys) participated in this study and took images of their meals. Each before image contained the foods and beverages served. A total of 19 unique foods were served over the three meals. Milk was served at each of the meals, thus 45 images that included milk were available for image analysis. Soda was served at lunch and dinner, thus 30 images were used for image analysis. For the 17 remaining foods served, only 15 images (i.e., 1 for each participant) contained those foods. The number of children self-reporting food portions using two different methods for the breakfast foods only is shown in **Table 1**. An example of a before image from the breakfast meal as taken by a participant is also shown in **Table 1**. The comparison of known gram weights to automatic gram weights by all foods is shown in **Table 2**.

Ratios of the automatic estimation to actual weights and the 95% CIs for breakfast foods are shown in **Figure 2**. Accuracy of the self-reported portions, as depicted by

Comparison of Known Food Weights with Image-Based Portion-Size
Automated Estimation and Adolescents' Self-Reported Portion Size

Lee

Table 1.
Foods Served to Adolescents (*n* = 15, 11–18 years) at Breakfast and the Number of Self-Reported Food Portions

Breakfast[a]	Food descriptions	Self-report per food	
		2D[b]	MDes[c]
	Scrambled eggs	8	7
	Sausage links	8	7
	White toast	7	7
	Margarine	7	7
	Strawberry jam	8	7
	Orange juice	—	—
	2% milk	—	—

[a] Example of an image of foods and beverages taken by one of the adolescents.
[b] 2D portion-size estimation aid used for self-reported food portions (2% milk and orange juice not estimated).
[c] MDes is multiple descriptors of common household and weight measures used, e.g., cups, teaspoon, etc. (2% milk and orange juice not estimated).

Table 2.
Automated Volume Analysis Converted to Weight (g) and Energy (kcal) Compared with Known Weight and Energy Based on Images Taken by 15 Adolescents (11–18 years) of Foods During a 24 h Period under Controlled Conditions

FNDDS food code	Brief name	Mean weight (g ± SD)[a]			Ratio of estimate to known[b]	Energy (kcal ± SD)[a]		
		n	Known	Estimate measures		*n*	Known	Estimate measures
11112110	2% Milk	45	220.0 ± 0.0	208.7 ± 9.8[c]	0.95	45	110.0 ± 0.0	104.3 ± 4.9[c]
25221660	Sausage links	15	46.5 ± 1.0	41.5 ± 2.8[c]	0.89	15	148.7 ± 3.1	132.8 ± 8.9[c]
32105000	Scrambled eggs	15	61.5 ± 0.7	108.5 ± 27.4[c]	1.77	15	91.0 ± 1.1	160.6 ± 40.6[c]
51101010	Toast	15	47.7 ± 3.4	80.0 ± 17.9[c]	1.67	15	139.9 ± 10.1	234.5 ± 52.3[c]
51121040	Garlic bread	15	41.1 ± 3.0	119.8 ± 15.3[c]	2.92	15	155.7 ± 11.3	454.0 ± 57.9[c]
53106050	Chocolate cake w/ icing	15	81.5 ± 12.5	105.7 ± 17.5[c]	1.31	15	298.3 ± 45.7	387.0 ± 63.9[c]
53241500	Sugar cookie	15	27.8 ± 1.9	31.6 ± 3.2[c]	1.14	15	132.1 ± 9.1	150.0 ± 15.2[c]
58132310	Spaghetti w/ sauce, cheese	15	240.3 ± 2.6	214.5 ± 60.9	0.89	15	377.3 ± 4.2	336.8 ± 95.6
61210220	Orange juice	15	124.0 ± 0.0	128.6 ± 10.3	1.04	15	52.1 ± 0.0	54.0 ± 4.3
63135140	Peaches	15	69.3 ± 9.9	116.1 ± 18.4[c]	1.69	15	37.4 ± 5.3	62.7 ± 9.9[c]
63137170	Pear halves	15	75.6 ± 4.9	138.9 ± 20.7[c]	1.84	15	37.8 ± 2.5	69.5 ± 10.4[c]
71401020	French fries	15	70.5 ± 4.3	204.7 ± 31.1[c]	2.90	15	94.5 ± 5.7	274.3 ± 41.7[c]
74401010	Ketchup	15	15.5 ± 0.4	17.0 ± 7.1[c]	1.10	15	15.0 ± 0.4	16.5 ± 6.9[c]
75114000	Lettuce (salad)	15	48.3 ± 4.8	220.3 ± 35.5[c]	4.61	15	8.2 ± 0.8	37.4 ± 6.0[c]
81103040	Margarine	15	27.8 ± 0.6	40.4 ± 12.4[d]	1.45	15	149.3 ± 3.4	216.9 ± 66.8[d]
83202020	French dressing	15	35.7 ± 1.0	32.7 ± 1.5[c]	0.92	15	71.4 ± 2.0	65.4 ± 3.0[c]
91402000	Strawberry jam	15	21.1 ± 1.1	21.4 ± 5.3	1.01	15	54.9 ± 2.9	55.6 ± 13.7
92410310	Coke	30	227.2 ± 2.3	305.7 ± 27.6[c]	1.35	30	84.1 ± 0.9	113.1 ± 10.2[c]
99999999[e]	Cheeseburger sandwich	15	198.8 ± 11.5	187.2 ± 34.5	0.95	15	361.8 ± 20.9	340.7 ± 62.8

FNDDS, food and nutrient database for dietary studies; SD, standard deviation.
[a] Paired *t*-tests were used to evaluate differences between known and estimated values.
[b] Ratio of estimated weight to known weight. A value >1 indicates an overestimation. A value <1 indicates an underestimation.
[c] $p < .001$.
[d] $p < .01$.
[e] 99999999 represents a combination food composed of six separate foods with food codes within FNDDS.

Comparison of Known Food Weights with Image-Based Portion-Size
Automated Estimation and Adolescents' Self-Reported Portion Size

Lee
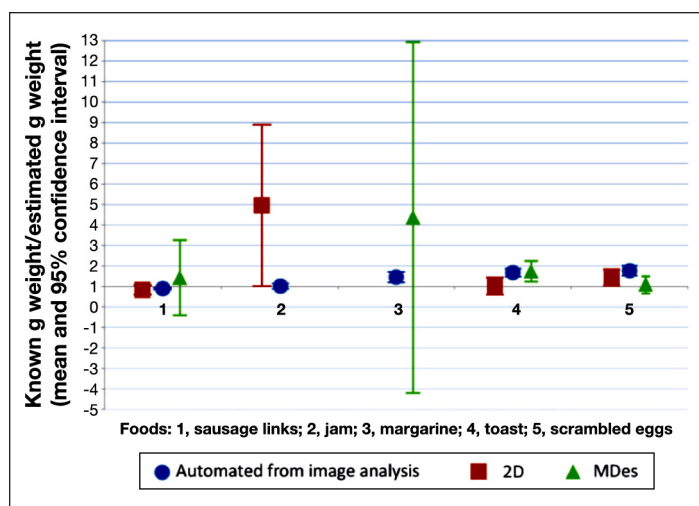
Figure 2. Weight error from images taken by 15 adolescents (11–18 years) at a breakfast meal. Ratio greater >1 is overestimated and ratio <1 is underestimated (mean and 95% CI). 2D, two dimensional portion estimation aid; MDes, multiple descriptors, e.g., cup, teaspoon. See text for further description.

ratios and 95% CIs, is compared with the automated method. The mean ratio of automatic weight estimation to known weight for the 19 foods is shown in **Table 2** along with energy comparisons. The mean energy for the food served to the 15 adolescents at all meals was 2723 ± 51 kcal. The mean energy estimated from the automatic volume computations was 3588 ± 180 kcal.

Discussion

To our knowledge, this is the first report of automatic volume estimation of food in meals over an entire day using images taken by adolescents. The weights of some foods and beverages were estimated fairly close and others were not. Of the 19 foods and beverages that represented commonly consumed foods, about 50% were estimated within an acceptable range, i.e., within 15% of truth (2% milk, sausage links, sugar cookie, spaghetti with sauce and cheese, orange juice, ketchup, French dressing, strawberry jam, cheeseburger sandwich). On the other hand, a food, such as lettuce, which has a large amount of void space, was overestimated by over 400%. Because of the low energy density of lettuce, the average energy difference between known weight and estimated weight was less than 30 kcal, which is small for an entire day. The gram weights were overestimated for at least 12 of the 20 foods, but depending on the food, the error in energy varied. For all foods combined, the mean energy was overestimated, suggesting that the energy dense foods seem to be most affected by any error. Thus, energy density of a food can help guide

the focus of which foods need the most attention with regard to improving the accuracy of automated estimation.

A common problem in using self-reported dietary intake data is the amount of error present,[26] which reduces the ability to find statistically significant associations between diet and health outcomes. In certain cases (i.e., sausage links, toast, scrambled eggs), the adolescents estimated portion size better than the automated estimate. However, in all cases, the degree of error, or spread around the means for each food and beverage, was substantially less for the automated estimates over the human estimated amounts. The potential for reduced error in portion-size estimation due to investment in automation is promising. In the future, data generated from automation may mimic the results of nutrition studies in controlled clinical studies.

Using simple shapes, the automated volume estimation differs substantially from human estimation. Human estimation is biased by social desirability or lack of knowledge about sizes,[27,28] whereas the automated volume estimation is not influenced by these factors. Challenges for the automated system include foods with void space or porosity, reflection from containers, and containers. Therefore, a food such as margarine may be overestimated, an unlikely occurrence among humans who are self-reporting portions. At this point, the results of the correlation between the true energy served and energy based on image analysis for volume still need improvement. The development of an expanded selection of shape templates may enhance results, as well as the benefits accrued from the rapid advances in technology.

Portion-size estimation is difficult for people and regarded as burdensome.[17,29] Several attempts have been made to provide creative methods to improve portion-size estimation.[30,31] For example, Matheson and colleagues[30] provided 8–12-year-old girls with clay to mold and shape a known quantity of a bread stick and crinkled paper strips for spaghetti with sauce and salad with dressing. This method worked well for some foods. Correlations between actual and estimated intakes with two portion-size measurement aids were high for three foods ($r = 0.56$–0.79, all $p < .001$) and low for the bread ($r = 0.16$, $p = .43$). Translating this method into practice, i.e., providing individuals with clay and paper for quickly estimating dietary intakes, is questionable. Despite attempts to improve human's abilities to estimate portion sizes, accuracy remains elusive. Conversely, advancements in technology that can equate to improvements in automation can

Comparison of Known Food Weights with Image-Based Portion-Size
Automated Estimation and Adolescents' Self-Reported Portion Size

Lee

offer substantially more benefits, e.g., low burden and more precision, than pursuing methods that involve an unreasonable amount of time on the part of individuals.

Besides the necessity to improve portion-size estimation prior to launching a system that can provide immediate feedback to a user, other technology challenges need to be addressed. The computational capacity of mobile telephones is still incapable of running the image analysis for automatic food identification and portion estimation. Thus, either the computational capacity of mobile telephones or speed of data transfer must improve substantially. However, given the proliferation of technological advances, the introduction of these tools for use in the future is assured.

Conclusions

Environmental factors such as diet and exercise play significant roles in the prevention of type 2 diabetes[32] and in self-management of diabetes. These lifestyle factors are potentially modifiable.[33] Progress in mobile technologies holds promise to reduce health care costs and move medical care more toward preventive care.[34] **Figure 3** shows an example of the type of message that could be displayed on the screen of a mobile telephone in the future that would immediately inform individuals

about their diets. The information regarding the amount of carbohydrate or other key nutrients can assist with modifications in food intake and medication levels. Although individuals with diabetes found recording food using smartphones to be motivating, they were less likely to continuously use these applications that required them to manually identify foods and estimate portion sizes.[14] Applications that take over the burden of these processes, as described in this article, may in the future encourage regular daily use, which may translate to better self-management.

Figure 3. A futuristic example of immediate dietary feedback using automated food identification and volume estimation running on a mobile application.

References:

1. Centers for Disease Control and Prevention. National Diabetes Fact Sheet, 2011. http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf.

2. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010;33 Suppl 1:S62–9.

3. Bonow RO, Gheorghiade M. The diabetes epidemic: a national and global crisis. Am J Med. 2004;116(5A):2S–10S.

4. Zimmet P, Alberti KGMM, Shaw J. Global and societal implications of the diabetes epidemic. Nature. 2001;414(6865):782–7.

5. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults 1999–2008. JAMA. 2010;303(3):235–41.

6. Ogden CL, Carroll MD, Curtin LR, Lamb MM, Flegal KM. Prevalence of high body mass index in US children and adolescents, 2007–2008. JAMA. 2010;303(3):242–9.

7. Engelgau MM, Gelss LS, Saadine JB, Boyle JP, Benjamin SM, Gregg EW, Tierney EF, Rios-Burrows N, Mokdad AH, Ford ES, Imperatore G, Narayan KM. The evolving diabetes burden in the United States. Ann Intern Med. 2004;140(11):945–50.

8. Bu D, Pan E, Walker J, Adler-Milstein J, Kendrick D, Hook JM, Cusack CM, Bates DW, Middleton B. Benefits of information technology-enabled diabetes management. Diabetes Care. 2007;30(5):1137–42.

9. Katz DL, Nordwall BA. Novel interactive cell-phone technology for health enhancement. J Diabetes Sci Technol. 2008;2(1):147–53.

10. Franklin VL, Greene A, Waller A, Greene SA, Pagliari C. Patients' engagement with "sweet talk"—a text messaging support system for young people with diabetes. J Med Internet Res. 2008;10(2):e20.

11. Faridi Z, Liberti L, Shuval K, Northrup V, Ali A, Katz DL. Evaluating the impact of mobile telephone technology on type 2 diabeteic patients' self-management: the NICHE pilot study. J Eval Clin Pract. 2008;14(3):465–9.

12. Jackson CL, Bolen S, Brancati FL, Batts-Turner ML, Gary TL. A systematic review of interactive computer-assisted technology in diabetes care. J Gen Intern Med. 2006;21(2):105–10.

Comparison of Known Food Weights with Image-Based Portion-Size
Automated Estimation and Adolescents' Self-Reported Portion Size

Lee

13. Franz MJ, Powers MA, Leontos C, Holzmeister LA, Kulkarni K, Monk A, Wedel N, Gradwell E. The evidence for medical nutrition therapy for type 1 and type 2 diabetes in adults. J Am Diet Assoc. 2010;110(12):1852–89.

14. Arsand E, Tufano JT, Ralston JD, Hjortdahl P. Designing mobile dietary management support technologies for people with diabetes. J Telemed Telecare. 2008;14(7):329–32.

15. Zhu F, Bosch M, Woo I, Kim SY, Boushey CJ, Ebert DS, Delp EJ. The use of mobile devices in aiding dietary assessment and evaluation. IEEE J Sel Top Signal Process. 2010;4(4):756–66.

16. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. Evidence-based development of a mobile telephone food record. J Am Diet Assoc. 2010;110(1):74–9.

17. Schap TE, Six BL, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. Adolescents in the United States can identify familiar foods at the time of consumption and when prompted with an image 14 h postprandial, but poorly estimate portions. Public Health Nutr. 2011;14(7):1184–91.

18. Six BL, Schap TE, Kerr DA, Boushey CJ. Evaluation of the Food and Nutrient Database for Dietary Studies for use with a mobile telelphone food record. J Food Compost Anal. 2011;24:1160–7.

19. Jensen JK, Gustafson D, Boushey CJ, Auld G, Bock MA, Bruhn CM, Gabel K, Misner S, Novotny R, Peck L, Read M. Development of a food frequency questionnaire to measure calcium intake of Asian, Hispanic, and White youth. J Am Diet Assoc. 2004;104(5):762–9.

20. Novotny R, Boushey C, Bock MA, Peck L, Auld G, Bruhn CM, Gustafson D, Gabel K, Jensen JK, Misner S, Read M. Calcium intake of Asian, Hispanic and White youth. J Am Coll Nutr. 2003;22(1):64–70.

21. Kim S, Schap T, Bosch M, Maciejewski R, Delp EJ, Ebert DS, Boushey CJ. Development of a mobile user interface for image-based dietary assessment. Proceedings of the 9th International Conference on Mobile and Unbiquitous Multimedia (MUM), Limassol, Cyprus, 2010 (NIHMSID: NIHMS327454).

22. Zhu F, Mariappan A, Boushey CJ, Kerr D, Lutes KD, Ebert DS, Delp EJ. Technology-assisted dietary assessment. Proc SPIE. 2008;6814(681411):1–10.

23. Woo I, Otsmo K, Kim S, Ebert DS, Delp EJ, Boushey CJ. Automatic portion estimation and visual refinement in mobile dietary assessment. Proc SPIE. 2010;7533:753300 (NIHMSID: NIHMS327456).

24. Chae J, Woo I, Kim S, Maciejewski R, Zhu F, Delp EJ, Boushey CJ, Ebert DS. Volume estimation using food specific shape templates in mobile image-based dietary assessment. Proc SPIE. 2011;7873:78730K. Doi: 10.1117/12.876669.

25. Vanhamel S, Van Den Ende L, Darius PL, Delcour JA. A volumeter for breads prepared from 10 grams flour. Cereal Chem. 1991;68(2):170–2.

26. Foster E, Matthews JNS, Nelson M, Harris JM, Mathers JC, Adamson AJ. Accuracy of estimates of food portion size using food photographs—the importance of using age-appropriate tools. Public Health Nutr. 2006;9(4):509–14.

27. Harnack L, Steffen L, Arnett DK, Gao S, Luepker RV. Accuracy of estimation of large food portions. J Am Diet Assoc. 2004;104(5):804–6.

28. Matheson DM, Wang Y, Klesges LM, Beech BM, Kraemer HC, Robinson TN. African-American girls' dietary intake while watching television. Obes Res. 2004;12:32S–7S.

29. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. Eur J Clin Nutr. 2009;63 Suppl 1:S50–7.

30. Matheson DM, Hanson KA, McDonald TE, Robinson TN. Validity of children's food portion estimates. Arch Pediatr Adolesc Med. 2002;156(9):867–71.

31. McGuire B, Chambers E, 4th, Godwin S, Brenner S. Size categories most effective for estimating portion size of muffins. J Am Diet Assoc. 2001;101(4):470–2.

32. Orchard RJ., Temprosa M, Goldberg R, Haffner S, Ratner R, Marcovina S, Fowler S, Diabetes Prevention Program Research Group. The effect of metformin and intensive lifestyle intervention on the metabolic syndrome: the Diabetes Prevention Program Randomized Trial. Ann Intern Med. 2005;142(8):611–19.

33. Horton ES. Effects of lifestyle changes to reduce risks of diabetes and associated cardiovascular risks: results from large scale efficacy trials. Obesity (Silver Spring). 2009;17 Suppl 3:S43–8.

34. Fletcher RR, Poh MZ, Eydgahi H. Wearable sensors: opportunities and challenges for low-cost health care. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:1763–6.

Enabling Syndromic Surveillance in Pakistan

Ross Maciejewski*[1], Shehzad Afzal[2], Adam J. Fairfield[1], Arif Ghafoor[2], David S. Ebert[2], Naeem Ayyaz[4] and Maaz Ahmed[3]

[1]Computer Science, Arizona State University, Tempe, AZ, USA; [2]Purdue University, West Lafayette, IN, USA; [3]King Edward Medical University, Lahore, Pakistan; [4]University of Engineering and Technology, Lahore, Pakistan

Objective

This work presents our first steps in developing a Global Real-time Infectious Disease Surveillance System (GRIDDS) employing robust and novel in-fectious disease epidemiology models with real-time inference and pre/exercise planning capabilities for Lahore, Pakistan. The objective of this work is to address the infectious disease surveillance challenges (specific to developing countries such as Pakistan) and develop a collaborative capability for monitoring and managing outbreaks of natural or manmade infectious diseases in Pakistan.

Methods

Utilizing our partner hospitals in the Lahore, Punjab area, we have begun developing a theoretical model of patient hospital visits with respect to diseases and syndromes within Pakistan. Our first thrust has focused on the collection, categorization and cleansing of data based on expert knowledge from our partnering institutions in Pakistan. Data consists of a patient's home address and chief complaint which is then categorized into syndromes. Home addresses are geocoded utilizing the Google API with a resultant 72% accuracy. Unknown geolocations are aggregated only at the hospital level. Using this cleaned data, we employ methods similar to our previous work [1] on syndromic surveillance for early disease detection. Currently, we have collected over 600,000 patient records over 1.5 years.

We employ the use of choropleth maps, isopleth maps utilizing kernel density estimation of patient addresses, traditional control chart methods such as exponentially weighted moving averages (EWMA), and a non-parametric time series analysis approach (seasonal trend decomposition using loess smoothing (STL) [2]) which requires only 90 days of historical data to be put into operation. The time series models are deployed as part of a real-time surveillance system in which temporal anomalies over regions can be analyzed and disease outbreaks reported.

Results

Figure 1 illustrates our visual analytics toolkit in operation. Here we see the location of our partner hospital in the Lahore region. The hospital coverage is in the most populous location of the city, providing data as a sentinel site for the overall health of the city. Currently, our system employs the use of interactive filters and linked isopleth or choropleth maps with time series analysis on mouse over.

Conclusions

Currently our research has focused on one partner location within the city of Lahore. Our ongoing work is focusing on the adoption of such a system to other regions of the country and the development of disease spread simulations (particularly Dengue Fever) utilizing baseline data collected by our partners. We plan to integrate these models into our visual analytics system for real-time planning and simulation.
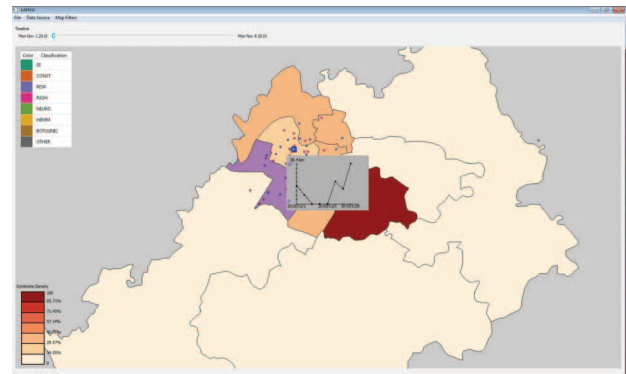
Keywords

syndromic surveillance; visual analytics; pakistan

Acknowledgments

References

[1] Maciejewski, R., et al.., "Forecasting Hotspot – A Predictive Analytics Approach," IEEE Transactions on Visualization and Computer Graphics, 17(4): 440-453, 2011.

[2] Hafen, R. P., et al., "Syndromic surveillance: STL modeling, visualizing and monitoring disease counts," BMC Medical Informatics and Decision Making, 2009.

*Ross Maciejewski
E-mail: rmacieje@asu.edu

Evaluating the Role of Time in Investigative Analysis of Document Collections

Bum chul Kwon, *Student Member*, *IEEE*, Waqas Javed, *Student Member*, *IEEE*,
Sohaib Ghani, *Student Member*, *IEEE*, Niklas Elmqvist, *Member*, *IEEE*,
Ji Soo Yi, *Member*, *IEEE*, and David S. Ebert, *Fellow*, *IEEE*

Abstract—Time is a universal and essential aspect of data in any investigative analysis. It helps analysts establish causality, build storylines from evidence, and reject infeasible hypotheses. For this reason, many investigative analysis tools provide visual representations designed for making sense of temporal data. However, the field of visual analytics still needs more evidence explaining how temporal visualization actually aids the analysis process, as well as design recommendations for how to build these visualizations. To fill this gap, we conducted an insight-based qualitative study to investigate the influence of temporal visualization on investigative analysis. We found that visualizing temporal information helped participants externalize chains of events. Another contribution of our work is the lightweight evaluation approach used to collect, visualize, and analyze insight.

Index Terms—Qualitative evaluation, investigative analysis, temporal visualization, insight-based evaluation

✦

1 INTRODUCTION

TIME has always received special treatment in the visualization literature [1]. It is used for a wide variety of tasks such as understanding causality, discovering trends, and predicting future events. A wide array of techniques has been developed over the years for visualizing temporal data, such as for time-varying quantitative data, event sequences, and storytelling.

Despite much of this prior work including results from empirical user studies, there exists very little knowledge on the actual role of temporal data and temporal visualization for **investigative analysis** [2]. In the field of visual analytics the concept of time is central to the analytical process [1], [3], [4] and widely utilized in many visual analytics tools. However, there has not been sufficient work on the role and impact of temporal information on the thinking process of an investigative analyst. In particular, recent progress in empirical evaluation of visual analytics systems [4], [5], [6], [7] have failed to clearly deal with this topic.

In this paper, we attempt to address this issue by presenting and discussing results from a qualitative evaluation comparing the performance of participants conducting a investigative analysis task using a visual analytics tool with and without access to temporal visualization. It should be noted that the purpose of this work is **not** to answer the

question whether temporal information and temporal visualization is useful or not—the answer to this question is a clear "yes"—but rather to study differences in how users utilize temporal information when it is explicitly presented in a temporal visualization, as opposed to when no such visualization is available. Our ambition is that these findings will in turn allow us to derive practical and workable results that have general application across a wide array of visual analytics tools.

Having said that, it is important to realize that evaluation of visualization and visual analytics is difficult and still in its infancy [8], [9], [10]. While the field of human-computer interaction has a long tradition of performing aptitude tests on low-level cognitive and perceptual tasks, it is not clear that extrapolating such tradition to higher level sensemaking and decision making tasks is possible [11]. The overarching investigative analysis task is generally too individual, volatile, and amorphous to afford quantitative evaluation and comparison. Our evaluation is, therefore, qualitative in nature, and we make no efforts to derive quantitative measures on time and error, which is often meaningless in the context of investigative analysis. Instead, our findings revolve around observation, semistructured interviews, and informal performance analysis. We, thus, follow in the footsteps of Kang et al. [4] but focus on a hitherto neglected aspect of visual analytics. We have also quite deliberately taken a lightweight approach to this qualitative comparison that we think may be of general use for evaluating visual analytics tools.

Thus, we see the main contributions of this paper as the following: 1) results and observations from a qualitative comparison of investigative analysis with and without access to temporal visualization; 2) design implications on how to best design and utilize temporal visualization in visual analytics tools; and 3) a novel evaluation approach for *lightweight qualitative comparison* that strikes a balance between time and cost versus depth and explanatory power.

• *B.c. Kwon and J.S. Yi are with the School of Industrial Engineering, Purdue University, 315 N. Grant Street, West Lafayette, IN 47907. E-mail: {kwonb, yij}@purdue.edu.*
• *W. Javed, S. Ghani, N. Elmqvist, and D.S. Ebert are with the School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Avenue, West Lafayette, IN 47907-2035. E-mail: {wjaved, sghani, elm, ebertd}@purdue.edu.*

2 BACKGROUND

This work explores the role of time and temporal visualization in *investigative analysis*. In this section, we first discuss how temporal information has been used in intelligence analysis and existing work that supports temporal data and visualization. We then motivate our evaluation by reviewing the state of the art in visual analytics evaluation.

2.1 Investigative Analysis and Time

Investigative analysis is defined as making discoveries and finding hidden truths in large collections of data [2]—in a way, detecting the expected and discovering the unexpected. It is a cognitively taxing task performed by a wide variety of user groups including business analysts, journalists, scientists, intelligence analysts, and law enforcement officers.

More specifically, investigative analysis involves understanding the connections, causality, and relationships between different entities, scattered in multiple documents, collected from multiple sources, and represented in multiple different formats [2]. Due to the limited working memory of the human mind, the analytical process becomes increasingly difficult as the number of entities involved in the analysis grows [12].

In addition to the complexity and large scale of the data, identifying potential explanations and hypotheses, as well as testing these hypotheses by finding evidence from collected data, is an onerous task. Because of this problem, the dominant approach of investigative analysts is similar to that of historians rather than that of scientists [13]. Instead of deriving all possible cases and scientifically evaluating them, which is often difficult if not impossible, they tend to find a coherent narrative to explain the interesting phenomena. Thus, collecting evidence to confirm or reject hypothetical stories is an important step for intelligence analysts [14]. Because of this tendency of creating stories out of evidence, time becomes very important in investigative analysis. Time is essential for suggesting and sometimes determining sequential orders, thereby clarifying cause and effect relationships. Temporal information can also be used to rule out unlikely hypotheses (e.g., if Bob's visit to a particular place happened much earlier than a suspicious event, his visit may not be related to the event) and identify impossible hypotheses (e.g., Bob was killed in 1998, so he cannot have bombed a building in 2000).

However, investigating the temporal aspects of evidence on top of already complex data is challenging. Visualization and visual analytics can aid the analysis by providing effective graphical representations of the available information—what cognitive scientists call *external cognition* [15], [16]—and by allowing for interactive exploration of these representations.

2.2 Temporal Visualization and Analysis

As discussed in Section 2.1, time is an inherent dimension in data analysis because of its unique semantic meaning; for this reason, it has always received special attention in visualization [1]. Therefore, it is not surprising that many visual analytics tools include some functionality for temporal analysis or temporal visualization. The most common mechanism is the timeline view, where events and time-varying data are visualized on a chart where one dimension (often the horizontal) is time. Another example is representing temporal dynamics in geo-spatial visualizations [17]. For investigative analysis, the data usually consist of discrete events in time, so we largely ignore the considerable body of work on visualization of time-varying or time-series data. Examples of discrete event timeline views can be found in Jigsaw [2], LifeLines [18], [19], and Similan [20]. Certain tools take temporal aspects a step further: GeoTime [3] has a story building mechanism for constructing narratives from event sequences, and CzSaw [21] maintains analysis provenance to facilitate reflection and replay.

Despite the prevalence of temporal visualization, we have not been able to find any studies that particularly investigate the role of temporal visualization in the analytical process. In fact, there exists very little work that empirically studies the analytical process in general (e.g., [4]), let alone its temporal aspects.

2.3 Evaluating Visual Analytics

Most empirical evaluations in visualization and visual analytics study low-level analytical tasks like search, navigation, and queries, and are therefore more of a physical aptitude than a cognitive nature [9]. Only a few studies investigate higher order analytical activities like sensemaking, decision-making, or even comparison, correlation, and organization.

In a sense, this dearth of empirical knowledge is an effect of the difficulty of evaluating visualization in general [9], [22], and investigative analysis in particular [4]. This is mostly due to the open-ended nature of many visual analytics tasks, which makes drawing clear conclusions from quantitative data difficult [23]. In fact, it is sometimes difficult to even *collect* quantitative data in the first place: What should really be measured? This is also the reason for the heavy emphasis on more qualitative and exploratory user studies of visual analytics tools in the literature.

Many such existing studies are relevant to our purposes. In separate work, Bier et al. [5] and Jeong et al. [6] studied quantitative performance for professional analysts solving sensemaking tasks in intelligence and financial analysis, respectively. Similarly, Isenberg et al. [24] and Robinson [25] independently performed exploratory studies of collaborative sensemaking in paper-based settings; both of these papers are particularly interesting due to their use of timeline visualizations to present results. Gotz and Wen [26] conducted an empirical study of user interaction behavior during visual analysis to propose general guidelines for user-driven visual analytics tools. Chin et al. [14] compared group and individual performance with collaborative information visualization environments through a quantitative experiment and derived a stage model that explains the users' collaboration process. Park et al. [27] qualitatively reviewed how collaborators in virtual environments work together to perform several tasks on visualized oceanographic data. Also of interest is the description of Plaisant et al. [10] of how the VAST contest judged the utility of the submitted visual analytic tools.

The recent qualitative study [4] on the Jigsaw [2] system is particularly instructive. The authors conducted a between-participants study, which divided participants into
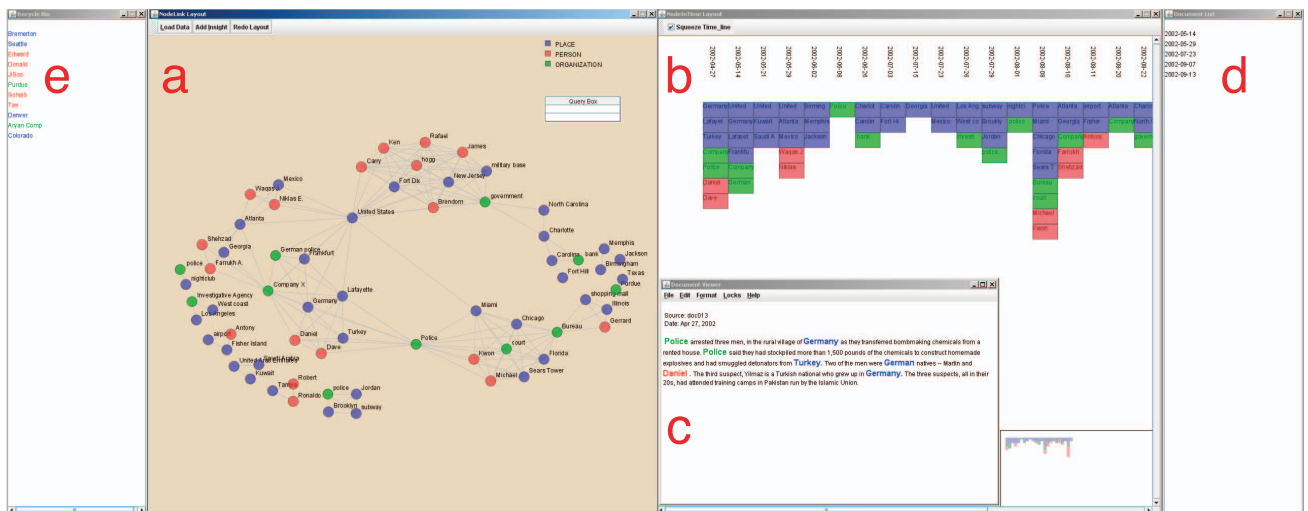
Fig. 1. Overview of the TIMEINVESTIGATOR tool. (a) Entity-relationship view for color-coded entities and their co-occurrence in reports. (b) Timeline view showing entity distribution in time. (c) Document view with color-coded entities highlighted. (d) Document list for currently matched documents. (e) Recycle bin showing entities that have been discarded as irrelevant. Blue entities are places, red are persons, and green are organizations.

groups of four, each group having only partial access to a subset of functions of Jigsaw, and made all of them work on the same intelligence reports to identify a fictional terrorist plot. Two external raters graded the score of findings in the experiment based on correctness of answers and also provided subjective grade on narrative debriefings. They also measured the elapsed time, amount of notes, documents viewed, etc. Rather than providing statistical measures, they tried to deduce particular strategies being used in each group to understand the role of the visual analytics system in the analysis.

Insight-based evaluation is another option. Saraiya et al. [28] used insight reports to collect findings for microarray data analysis. The reports were then evaluated and proved helpful for understanding what kinds of insights that participants generated while using different tools. This kind intrusive methodology rather than measuring time and accuracy of users' performance could be useful in order to capture the cognitive analysis process [29].

3   INVESTIGATIVE ANALYSIS SOFTWARE

As our literature review shows, there already exists a number of investigative analysis tools such as Jigsaw [2], CzSaw [21], and Analyst's Notebook [30] that we could use in our evaluation of temporal visualization. However, these tools require significant training and they often include many different views and methods for solving a particular task. In addition, it is difficult for outsiders to instrument these tools to collect user interaction data (e.g., click stream). Therefore, we felt that a better approach would be to identify the canonical tasks in investigative analysis and develop a minimalistic tool that supports them.

3.1   Canonical Tasks

Based on our focus on time and on the interaction categories proposed by Yi et al. [31], we derive the below canonical tasks for investigative analysis.

- **Reading documents.** Reading is a central activity in investigative analysis [2]. (*Elaborate* [31])

- **Viewing relationships.** Relationships between entities suggest association, information exchange, and causality. (*Connect* [31])
- **Selecting.** Marking entities allows for structuring work and correlating relationships. (*Select* [31])
- **Filtering.** Entities that are irrelevant to the analysis should be possible to discard. (*Filter* [31])
- **Viewing temporal relationships.** Causality is an important relationship [1]. (*Reconfigure* [31])

3.2   TimeInvestigator

Guided by these canonical tasks, we developed an investigative analysis tool, called TIMEINVESTIGATOR, consisting of five cross-linked views where an operation in one view (e.g., selecting and filtering) would affect all other views accordingly (Fig. 1).

- The **Entity-Relationship (ER) view** shows entities and their co-occurrences using a graph (Fig. 1a).
- The **Timeline view** shows entity occurrences on a timeline (Fig. 1b).
- The **Document view** shows reports with the entities highlighted and color-coded (Fig. 1c).
- The **Document list** shows names and dates of currently matched documents (Fig. 1d).
- The **Recycle bin** contains entities that have been removed from other views; e.g., filtering out irrelevant entities from the Timeline (Fig. 1e).

Using these views, users were able to dynamically add and remove entities from the application—this essentially meant moving entities to and from the recycle bin. On starting up the application, no entities were shown in the main views. The analyst could then add whole ranges of entities, or just select a few.

3.3   Entity-Relationship View

The Entity-Relationship view (Fig. 1a) is the main view of TIMEINVESTIGATOR and is designed to partly mimic the Graph view of Jigsaw. The view displays the entities in the document collection as nodes and their co-occurrence in

documents as link relations between the nodes. Nodes are labeled with their entity names and are color-coded depending on their type (i.e., Places in blue, Organizations in green, and Persons in red). Nodes can be moved so that the user can partition the space during the analysis.

Beyond browsing, the view also supports free text search using a query box (top right in Fig. 1a). Matched nodes are highlighted in yellow. Finally, the ER view also incorporates an entity legend (just above the query box in Fig. 1a) that supports toggling visibility of nodes by entity type simply by clicking on the label. Finally, entities can be filtered out using a double-click (sending them to the recycle bin).

3.4 Timeline View

The Timeline view is a temporal visualization that displays entity occurrences organized along a temporal axis (Fig. 1b). This is done by aggregating all reports in the document collection by their dates, and then showing all of the entities for each particular date. Each entity in the Timeline view is represented by a labeled box that is color-coded according to the entity type. Since a single entity may appear in more than one report at different dates, entity boxes may be duplicated for several dates along the timeline.

For each date, entity boxes are grouped according to their type to make the display consistent; for our example in Fig. 1b, we group entities in the order of "Place," "Organization," and "Person" from top to bottom. Furthermore, the order within each entity type group depends on the number of occurrences of a particular entity in the whole document collection, organized in descending order (i.e., the entity with most occurrences is placed at the top). The view can be scrolled horizontally to support long event sequences contained in large document collections. A small viewport in the right bottom corner shows an overview of the whole timeline to aid overview and navigation. To further ease temporal navigation, the user can toggle timeline compression, where all dates containing no currently selected entity are removed. Finally, a user can remove an entity using a double-click (sending it to the recycle bin).

3.5 Reading Documents

The Document list enumerates documents where matched or selected entities occur (Fig. 1d), and updates as the user selects, queries, and filters entities. The Document view allows for reading the actual text of reports in the document collection (Fig. 1c), a vital part of investigative analysis [2]. Just like in Jigsaw, the view highlights entities using color-coding based on type, and also draws them using a bigger font. Any number of reports can be open at a time.

3.6 Recycle Bin

The Recycle bin is a list of entities that have been removed (i.e., filtered out) in order of removal (Fig. 1e). Entities can be returned into the data set by double-clicking on its entry. The ER and Timeline views also support undo and redo for the delete operation, moving entities to and from the recycle bin.

While the recycle bin is not typically a canonical component found in investigative analysis tools, its existence in TIMEINVESTIGATOR is a side effect of the decision to allow participants to discard entities from views and recover those if necessary. Discarding entities corresponds to filtering in many existing tools. Furthermore, no participant reported difficulties in understanding this function, probably because it resembles the trash can in major operating systems.

3.7 Recording Insights

For the purposes of performing insight-based evaluation [28], we created a view called the Insight Report view. A new insight report can be generated at any time at the click of the "Add Insight" button; doing so will take a screen capture of the whole TIMEINVESTIGATOR desktop, and will open a text field where the user can type in free text related to the insight. The insight report also asks which view (Document, Entity-Relationship, or Timeline view) helped inspire the insight. The text, screenshot, and time stamp are saved when submitting the report.

The insight report view was a pure byproduct of the evaluation methodology, and we tried to minimize its impact on the analysis process. In particular, existing reports created at an earlier stage were *not* available for consultation at later stages. In the study (see below), we stressed the need to report findings using this mechanism, but we did not explicitly remind participants to do this during sessions. Despite these steps to minimize its impact, it is entirely possible that the inclusion of insight report generation changed the structure of the analysis process; this was also noted by Saraiya et al. [28] in their original work.

4 EVALUATION

Our ambition with this work, as noted above, was to study the influence of temporal visualization on investigative analysis of document collections. Below we discuss the general method we employed, as well as specifics on participants, equipment, and task.

4.1 Method

Out of the evaluation methods reviewed in the literature, we found the controlled study approach by Kang et al. [4] that involved single nonexpert users in contrasting conditions working on an extensive constructed scenario with ground truth to be the most appropriate for our work. We, therefore, decided to adopt this methodology, but to reduce time investments by not video recording sessions, and instead to use a combination of observations, screen captures, click streams, and insight-based evaluation [28] to collect deeper insights about the analytical process.

We call this *lightweight qualitative comparison*, and submit that it may be a useful evaluation method that strikes a balance between in-depth qualitative (or even ethnographic) evaluation performed using domain experts, and low-overhead quantitative evaluation involving nonexpert participants.

4.2 Participants and Apparatus

We recruited 12 paid participants ($10 per hour)—seven males and five females recruited from the engineering student population at our university—randomly divided into two groups: six participants with *no* access to the

Timeline view in the TIMEINVESTIGATOR tool (Group N), and six participants with full access to the Timeline (Group T). The reason for choosing students as participants as opposed to professional analysts is that we were unable to get access to such analysts in our traditional university setting. We discuss the implications of this limitation further in Section 6.6.

The study was performed on a desktop computer equipped with two $19''$ monitors ($1,280 \times 1,024$ pixels) to accommodate the multiple views of TIMEINVESTIGATOR. Participants were not told the name of the tool, nor the special emphasis on temporal analysis to minimize any unexpected biases. Prior to starting the experiment, participants underwent a training session of approximately 20 minutes using a dummy data set.

4.3 Task

The task consisted of identifying a hidden terrorist plot in a collection of 50 fictional intelligence reports. This data set was the same that was used in the recent evaluation by Kang et al. [4]. Participants were allowed to take up to 1 hour to complete the task and were encouraged to make use of the full time.

Participants were instructed to create insight reports whenever they learned something significant about the document collection. They were told that these reports would be the main evaluation instrument in the study, and thus that creating reports was important.

Upon finishing the experiments, participants were told to write a short narrative on the suspected terrorist plot. They were then issued a questionnaire on their experiences of the method, strategy, and view primarily used to perform the task. Finally, we also conducted exit interviews with all participants.

4.4 Measures

We collected several measures to understand the experiences of the two groups of participants (Groups N and T), including interviews and insight reports (which view a participants got the insight from, text, and screen capture). We also instrumented the TIMEINVESTIGATOR tool to collect participant usage patterns (i.e., the uses of the different views and clicks of entities with timestamps). The purpose was to use these quantitative measures to aid our understanding.

We coded insight reports systematically using two coders, who are also authors of this paper, working independently and using a shared coding rubric (Cohen's kappa coefficient $= 0.49$, which is considered "good" or "moderate"). The independent code streams were then merged, discussed, and unified. Score was based on five main plot points (and a number of subplots per plot point) that we had extracted from the ground truth of our data set; taken together, these five plot points explained the full story. Every plot point that was discussed in an insight report was scored from 0.0 to 1.0 depending on the accuracy of the insights and the coverage of subplots, so each insight report was scored between 0.0 and 5.0. The final score per participant is based on the accuracy and comprehensiveness of cumulative insight reports.
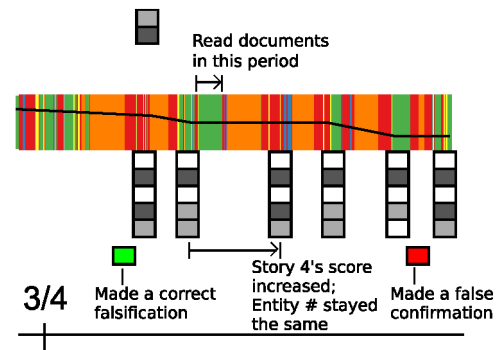
Fig. 2. Key for the activity timeline in Fig. 3.

Some participants falsified—i.e., disproved certain elements of the data set as not being relevant or correct—story components that were not included in the main plot and/or reported incorrect speculation and confirmation. Though these are notable aspects of insight reports, including them into scores is problematic (e.g., PT3 successfully falsified story components in 10 different insight reports, but how much is each successful falsification worth?). Thus, such elements are separately codified and not included in the scores.

5 RESULTS

We collected results using a combination of interaction logs, observations, interviews, and insight reports.

5.1 Visualizing Evaluation Results

To aid our understanding of participant analysis processes, we decided to visualize our study results. Inspired by timeline visualizations created by Isenberg et al. [24] and Robinson [25], we created the visualization in Fig. 3 to show the temporal event sequences we collected during the study: which view a participants interacted with, when an insight report was submitted, the individual scores for insight reports, and the number of entities visible in the tool. Fig. 2 gives a legend to aid in understanding Fig. 3. Some notable observations follow.

First, the lack of patterns in the visualization suggests the great variation in analysis method between individuals. Participants demonstrated wide variation in their final scores, how frequently insight reports were submitted (e.g., PN4, PT2, PT3, and PT6 submitted insight reports more frequently than the average across all participants, but PN1 and PN5 did less frequently), and which views they frequently used (e.g., PT4 used the Timeline view heavily, but PN4, who did not have access to the Timeline, mainly used the ER and the Document list). These large differences made us doubt that simply recruiting more participants would yield statistically significant results.

Second, as discussed previously, Group N viewed fewer entities in the TIMEINVESTIGATOR tool than Group T when their first insight reports were submitted. The black solid lines in the colored band in Fig. 3 indicate that the numbers of entities are generally increasing in Group N as time progresses, but decreasing in Group T. In other words, Group N seemed to base their analysis on progressively adding supporting evidence, whereas Group T instead
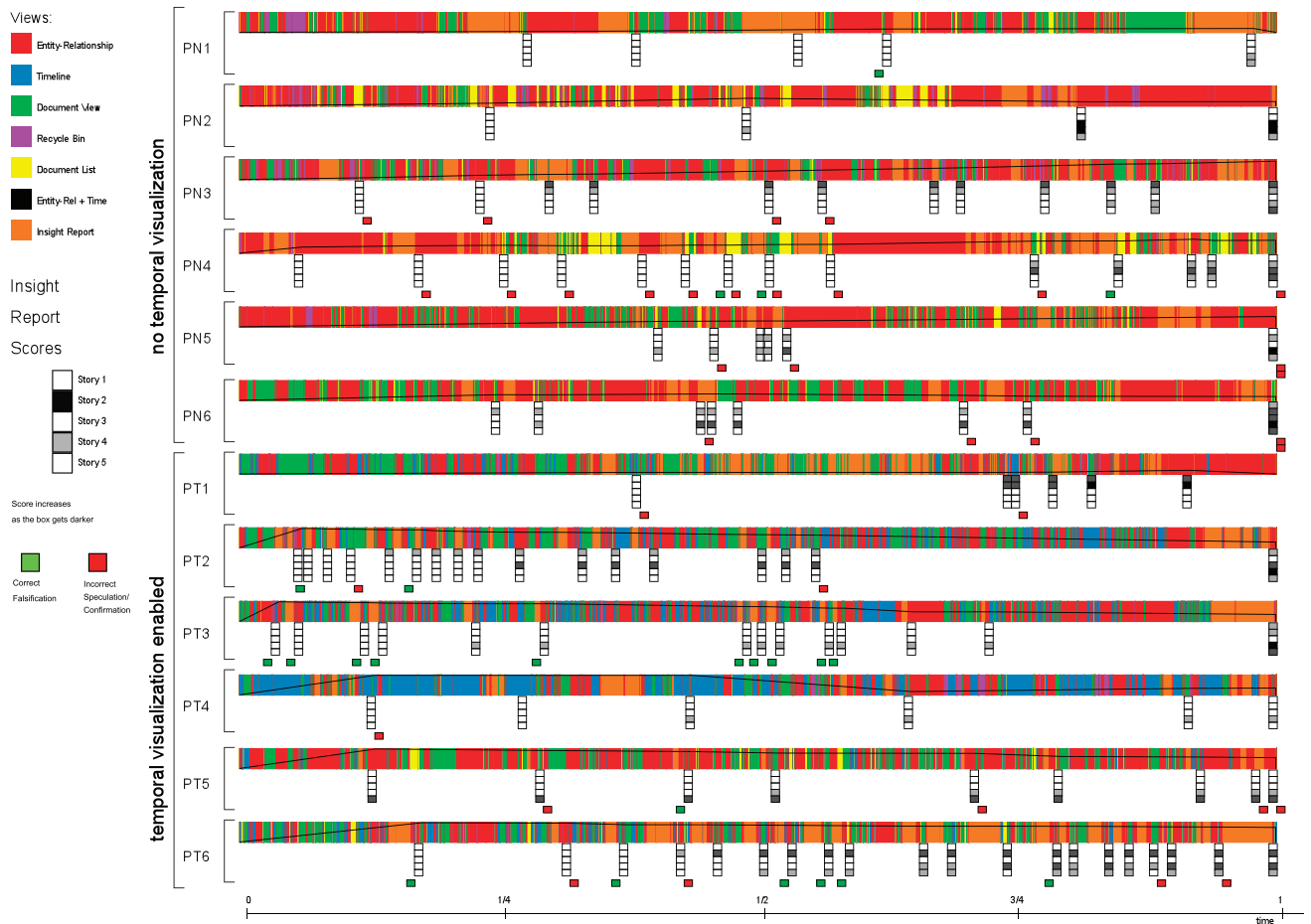
Fig. 3. Activity timeline for study sessions for all 12 participants (inspired by Isenberg et al. [24] and Robinson [25]). The color-coded band indicates which view the participants were interacting with. Black lines across the color-coded band show the number of entities shown in the ER view. A stack of boxes under the color-coded band shows the score of an insight report. Each box in a stack corresponds to one of the five major storylines; higher scores results in darker color of each box for that particular storyline.

iteratively removed circumstantial or unrelated evidence. This result is in line with the interview results reported in Section 5.2. Participants in Group N seemed to have difficulties in dealing with many entities, so they tended to carefully add entities. In contrast, ones in Group T, except for PT1, started with almost all entities on the screen and progressively filtered out ones that were irrelevant.

Third, some participants (PT3 and PT6) in Group T successfully falsified irrelevant plot points, indicated in green boxes in Fig. 3, but Group N participants collectively demonstrated only four successful falsifications. Actually, insight reports submitted by PN3, PN4, and PN6 included quite a few incorrect speculations and confirmations, indicated by red boxes.

These findings suggest that Groups T and N employed significantly different analysis methods, and the likely cause for this is the absence or presence of the Timeline view. The benefits of the Timeline seems to be in organizing and externalizing the events and then helping in discarding irrelevant entities. Below we study these aspects in more depth.

5.2 Interview

The results of interviews revealed three notable benefits of the Timeline view: for 1) making sense of the order of event

chains; 2) identifying important nodes; and 3) discovering relevant documents easier.

The first and rather obvious benefit of the Timeline view is that it helps making sense of the sequential and/or logical order of event chains, pointing to the prevalence of stories in temporal reasoning. When asked about how the Timeline view helped, PT1 said, "[It helps to] figure out how relationships change over time." PT2 also added, "The timeline helped me understand the order [of events]."

Second, the Timeline view also seemed helpful in identifying important nodes. In a question asking "Please describe how you knew you had found the main plot," PT4 remarked, "The frequency of a node appears to be significant. If a node appears multiple times in the [Timeline] view, it is more important." PT5 also said, "[In the Timeline view,] strong connections can be shown if entities show up multiple times." PT6 described his or her strategy as "See a person first. Follow the timeline. See if they are linked to the plot. Pay attention to areas with lots of blocks [i.e., events] in the Timeline view." In contrast, Group N participants seemed to have difficulties in discerning which is important or not. For example, PN3 reported, "[I] keep them all till the end because there is no way to decide if any of them are important."

TABLE 1
Summary Statistics of All Participants with No Access to Temporal Visualization (Group N)

| Measures | Group N | | | | | | | Statistics |
	PN1	PN2	PN3	PN4	PN5	PN6	Avg	Mann-Whitney U
Final score	0.37	2.17	1.45	0.98	1.23	2.42	1.44	$U = 22, p = 0.5887$
Earliest insight report (min:sec)	16:04	20:23	7:08	3:40	29:15	15:14	15:00	$U = 25, p = 0.3095$
# of insight reports	5	4	12	14	6	8	8.18	$U = 9.5, p = 0.1962$
# of entities in initial insight report	6	19	11	40	36	40	25.33	$*U = 4, p = 0.0292$
# of entities in final insight report	29	30	120	85	69	27	60	$U = 17, p = 0.9361$

TABLE 2
Summary Statistics of All Participants with Access to Temporal Visualization (Group T)

| Measures | Group T | | | | | | | Statistics |
	PT1	PT2	PT3	PT4	PT5	PT6	Avg	Mann-Whitney U
Final score	1.17	1.73	1.67	0.20	0.63	1.27	1.11	$U = 22, p = 0.5887$
Earliest insight report (min:sec)	25:35	4:02	2:11	6:53	7:24	10:35	9:00	$U = 25, p = 0.3095$
# of insight reports	6	17	14	6	9	20	12	$U = 9.5, p = 0.1962$
# of entities in initial insight report	12	123	127	127	127	126	107	$*U = 4, p = 0.0292$
# of entities in final insight report	28	36	44	44	71	95	53	$U = 17, p = 0.9361$

Third, Group T seemed to find it easier to discover relevant documents than Group N. The Timeline view not only shows the temporal information but also serves as an easy access to documents ordered chronologically, which helped users to easily follow suspicious entities over time. (PT2: "Read through time. [Find] what they did in the past and the future." and PT3: "With the remaining nodes, [I] used the Timeline view to make sense of stories.") In contrast, interviews with Group N hinted at the impact of having no access to temporal visualization: they were forced to read a lot of reports. When asked "How did you go about confirming that the plot is actually threatening?," Group N participants stated that they had to search multiple documents without any order. (PN1: "I could confirm the story only after reading the whole document" and PN5: "[I] verified it by reading reports associated with it. I would find the suspected node and verify it by using the [Document] view.")

5.3   Summary Statistics

First, we reviewed summary statistics to see if there were any significant differences between Groups N and T. However, because the number of participants is small, it was difficult to calculate statistical analyses with reasonable confidence. As expected, we did not see many notable differences between the two groups, as evidenced by Tables 1 and 2.

The only statistically significant difference was the number of entities that they placed on the ER view[1] (not in the Recycle bin) when they submitted the first insight report. This is in line with our observations from Section 5.1. Except for PT1, all Group T participants had over 100 entities on the views when the first insight reports were submitted. In contrast, Group T participants had on average less than 40 entities. This difference is statistically significant (Mann-Whitney $U = 4, p = 0.0292$). However, the difference in the number of entities diminished when they submitted the final insight reports (no statistical difference).

We also investigated how much time was spent for each view and from which views participants gained their insights (see Table 3). For both groups, participants tended to spend most of their time on the ER view, which is somewhat surprising to us because we had anticipated that the Document view would consume significant amount of time because they should read reports anyway to know the details. It is also interesting to see that the majority of reports were based on insights gained from the Document view (see Table 3), a result consistent with Stasko et al. [2]. However, there is no statistically significant difference between the two groups except for those due to the presence and absence of the Timeline view.

5.4   Insight Reports

The insight reports proved to be a rich source of qualitative information on the investigative strategies employed by our participants. One main finding is that the timeline helped Group T participants in finding the correct results more quickly. Below we pull out the main such trends and discuss them in depth.

5.4.1   Falsification

We first studied insight reports submitted by PN4, whose reports had 10 instances of incorrect speculations/confirmation. What we found is that PN4 often started with a suspicious activity based on the layout of entities in the ER view (PN4-0[2]: "There is a group of six people that communicate with each other a lot. They could be plotting something.") The suspicion continued in PN4-1 and ended at PN4-7, when PN4 opportunistically found evidence showing that PN4's initial suspicion was incorrect. While proceeding with the investigation, PN4 basically found initial cues, uncovered additional entities related to the clue, and expanded the network around these initial suspects. The procedure was then repeated. This seems to be a fairly natural investigative analysis process that mixes intuition

---

1. The number of entities on the Timeline view is identical to those on the ER view because the views are synchronized.

2. PN4-0 stands for the first insight report from participant PN4. Note that the insight report number starts from 0.

TABLE 3
Time Usage and Insights Generated from Each View

TimeInvestigator view	Group N		Group T	
	Avg.	S.D.	Avg.	S.D.
Usage times (min:sec):				
Entity-Relationship	36:01	10:01	22:41	6:34
Timeline view	–	–	11:30	7:40
Document list	8:43	4:38	1:18	1:52
Recycle bin	3:58	3:28	0:50	0:43
Document view	11:29	3:30	14:26	5:03
Insight report	10:38	4:42	9:48	5:31
Number of insights:				
Entity-Relationship	2.83	2.79	4.17	3.87
Timeline view	–	–	2.5	2.59
Document view	5.33	3.83	5.67	6.09

Fig. 4. A portion of the screenshot submitted with PT3-12. Note that Arnold and Abu H. appeared in two adjacent documents (2002-10-21 and 2002-10-22), which reveal fingerprint matches between two identities.

and guesswork with evidence and reasoning, and is consistent with earlier results [4], [24], [25].

We found that the Timeline view for Group T had a significant impact on the investigative process. Some participants made falsifications based on the duration of entities co-occurring throughout the timeline. For example, PT3 simply removed three entities based on their occurrence (PT3-0: "Since **Robert D'Onfrio**, **Tampa**, and **Jesus Vazquez** were only mentioned once and mentioned in the same report, I am removing them as suspects"). If PT3 would not have had access to the Timeline view, PT3 may have followed more wrong leads similar to PN4 because the three entities look closely connected in the ER view. Furthermore, in one instance, PT6 stopped tracing a person because the person had no appearance after a specific time (PT6-2: "**Julio** and **David** were removed since their act does not connect with the terrorist attack at this moment.").

In summary, Group N participants lacked the additional cues that Group T had from the Timeline view. The lack of these additional cues made Group N participants (particularly, PN3, PN4, and PN6) consider irrelevant information as a part of main plots.

5.4.2 Alias Detection

The 50 intelligence reports contain three aliases that are crucial for understanding the terrorist network because seemingly disconnected networks suddenly become connected when two separate names turn out to denote the same person. Although we cannot show statistical significance, we found that Group N noticed fewer instances of such aliases and made more mistakes in dealing with aliases than Group T. More specifically, Group N correctly identified a total of five aliases in PN1-4, PN2-2, PN5-0, PN5-2, and PN6-0 while Group T identified nine in PT1-1, PT1-3, PT2-16, PT3-12, PT3-13, PT5-3, PT6-9, PT6-11, and PT6-17. In addition, PN4 and PN6 treated one person with two aliases as two separate people (PN4-9 and PN6-2), and PN5 found a wrong alias for an entity (PN5-1), while no participants in Group T made such mistakes.

Note that identifying such aliases does not require a global understanding of the terrorist network. Instead, it requires simply reading a specific document containing evidence like "**Abu H.**, who was released from custody after the September 11 incidents and whose fingerprints
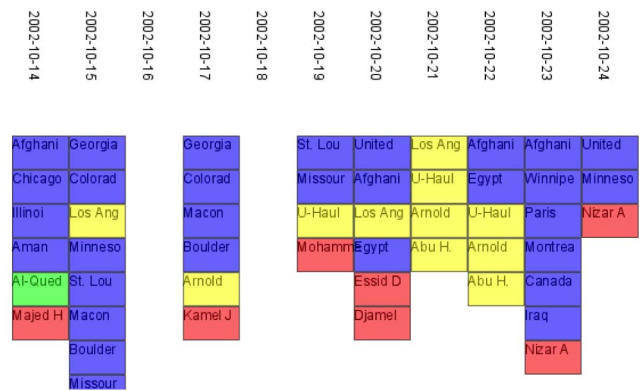
were found in the U-Haul truck rented by **Arnold C.** (report on 2002-10-22)" Thus, the performance of identifying aliases largely depends on the ability to identify such a document.

Even after investigating all of the insight reports and associated screenshots relevant with these aliases, we failed to find a single and universal explanation why there is a difference in identifying aliases between the two groups. One speculation is that it is a mere positive side effect of the Timeline view as an additional overview that helped participants find more relevant information efficiently as discussed in Section 5.2. Another speculation is that the Timeline view may make a document containing evidence for an alias more salient than the ER view. In the ER view, such a document appears as a single link between the two identities, which could be easily overlooked in a complex network. However, in the Timeline view, the two names would separately appear in different documents except for the one or two documents containing the evidence showing the connection between the two. This visualization would be more visible than a single link in the ER view as shown in Fig. 4.

5.4.3 Screen Captures

We also analyzed the screen captures that were taken at the moment that participants submitted reports.

Interestingly, the layouts of entities in the ER view generated by Groups N and T are drastically different, as exemplified by Figs. 5 and 6. It was clear that several participants (PN2, PN5 and PN6) in Group N tried to place
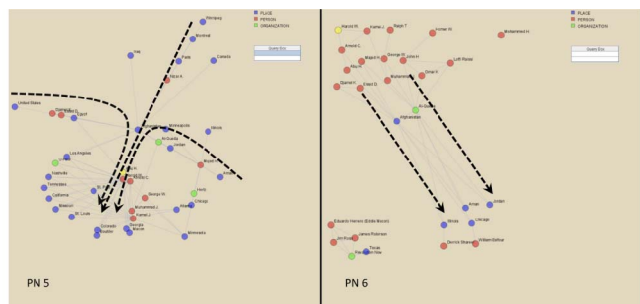
Fig. 5. The layouts in the entity-relationship view for PN5-4 (left) and PN6-7 (right).
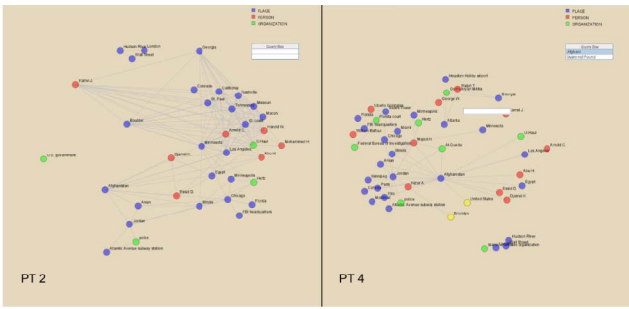
Fig. 6. The layouts of the entity-relationship view for PT2-16 (left) and PT4-5 (right).

entities in temporal orders on the ER view to reflect the identified storyline (PN1 and PN4 reorganized the ER view, but there was no temporal order visible). Fig. 5 shows the ER view with dotted lines indicating these storylines. There is also an exception to this pattern: PN3 did not change the layout of the ER view at all for any spatial organization.

That Group N used the ER view for recovering temporal order is not unexpected. Group N did not have any external media (not even paper and pencil) to record or build storylines. Thus, the ER view is a natural medium for them to externalize their storylines while conducting investigative analysis. This might also explain why the initial numbers of entities on the screen were lower in Group N. We speculate that the participants probably did not want their storylines to be polluted with irrelevant entities.

In contrast with Group N, we did not find a particular layout pattern on the ER view for Group T. Instead, we found that entities on the ER view were more or less randomly spread or held their initial positions without much changes. Fig. 6 exemplifies how PT2 and PT4 organized entities on the ER view.

The random entity layouts generated by Group T were unexpected. Because Group T participants also did not have any other external media for recording storylines, they likely used the Timeline instead of the ER view for externalizing analysis. This is interesting because the Timeline does not allow users to change the layout of entities on the screen, which we initially thought would make it unsuitable as a story mechanism. This is discussed in more detail next.

6 DISCUSSION

Our emphasis with this work was on understanding the role of temporal visualization using the Timeline view in TIMEINVESTIGATOR as an instantiation. As acknowledged earlier, our intention is not to provide statistical comparisons between the two groups. We also acknowledge that the addition of the Timeline view may improve the analysis by simply providing another external visual representation. However, the results shown above clearly indicate that participants were indeed influenced by the presence and absence of this view. Therefore, we focused our attention on comparing behavior as opposed to comparing low-level performance between the two groups.

6.1 Benefits of the Timeline View

In summary, we noted three benefits of the Timeline view from our study: the Timeline view 1) is vital for uncovering

important entity relations; 2) allows for filtering out unimportant entities; and 3) helps identify patterns that are invisible in the ER view.

First, the most obvious benefit of having access to the Timeline view is that it aids in uncovering important relationships. The ER view only shows binary relations, i.e., whether there was a relationship between certain entities. On the other hand, the Timeline view shows the development of relationships over time. This, in turn, seemed to provide Group T participants with 1) chronological and logical order of events; 2) the importance of events/entities; and 3) changes in relationships over time (based on feedback from PT1, PT4, and PT6). The absence of temporal visualization may be a roadblock for Group N because the chronological order of events were not visually available, so they needed to be remembered or recorded by some other method. Group N's heavier use of the Document List (average 8:43 in Group N versus average 1:18 in Group T in Table 3), which listed the dates of reports in chronological order, indirectly shows that participants needed external cues to organize events in time. The interesting layouts of nodes and links, built by PN3 and PN4, in Fig. 5 also showcase a tendency to want to make sense of stories and time.

Second, the Timeline view also provides additional cues to identify unimportant and irrelevant information. Two groups had unique patterns in the use of falsifications. Group N tended to be hesitant to falsify and thus discard entities, presumably due to the difficulty of overseeing long-term implications of such an action. Interview quotes from Group N clearly showed this aspect: for example, PN3 reported that he or she felt that he/she could not remove any entities because they may somehow be important. Because of the absence of temporal visualization, participants in Group N seemed to struggle in following entities of interest. On the other hand, Group T could follow entities through time and discard if they turned out not to be important. For example, when some entities appeared only in a certain time period, which was clearly visible in the Timeline view, these entities were easily disregarded (PT1 and PT6). This seemed to help Group T make falsifications with confidence.

Third, the Timeline view appears to be effective in highlighting a specific pattern, such as aliases, which could be easily obscured in the ER view. Identifying aliases is a particularly interesting activity in the context of investigative analysis because evidence showing aliases is often subtle but may drastically influence the analysis outcome. More specifically, it makes the aliased entity very suspicious and often helps to better understand its local neighborhood of entities. Although we failed to collect evidence showing that the Timeline view directly helped identify such aliases, we speculate that the Timeline view generally makes such subtle evidence more visible.

These benefits of the Timeline view were slightly different from our initial speculations derived from the literature (Section 2.1). We thought that time would be important because it can definitely rule out some hypotheses because they are impossible or unlikely due to causality (i.e., effects cannot come before causes). However, we found that causal relations cannot be easily visualized

because these subtle causalities are often buried in documents. Entity co-occurrence in a single document is far from enough for causal relations distributed in time (and thus generally over several documents). Therefore, an analyst must often read several documents carefully to extract these casual relationships. We found that while temporal visualizations cannot directly provide causality information, they can certainly support the task of deriving causality. Entity frequencies, which are relatively simple to visualize, served as important cues for our participants to discern which nodes might be more important than other nodes.

6.2 The ER View versus the Timeline View

Another interesting aspect of the Timeline view is that it not only provided additional benefits on its own but also influenced how the ER view was used. As shown repeatedly, Group N tended to start with a small number of entities and added more entities as the analysis progressed, while Group T did the opposite. In other words, Group N seemed to use the ER view in the additive fashion of "drawing on a canvas" to record the progress of investigation. Thus, they did not want to overload the ER view with irrelevant entities, which may distract them. We also noticed that, when too many entities were loaded into the ER view, participants tried to divide and allocate particular regional space in the ER view for specific use (e.g., Fig. 5). This intelligent use of space for simplifying choice, perception, and computation in the real world has also been observed in general cognitive science studies, e.g., by Kirsh [32].

Group T, on the other hand, seemed to use the ER view in a subtractive fashion of "carving a sculpture." This is a radically different approach. Rather than adding important entities to the ER view, participants in Group T chose to eliminate unrelated entities from the view. The ER view did not seem to be used for story building as discussed in Section 5.4.3.

This difference is most likely caused by the fact that Group N participants used the ER view as a story building mechanism, whereas Group T participants did not. The unorganized layouts on the ER view generated by Group T participants are particularly intriguing. Assuming that Group T also built storylines for their investigative analysis, Group T participants appeared to use the Timeline view as a story building mechanism. However, it should be noted that the Timeline view does not provide much degree of freedom in organizing visual elements because the horizontal and vertical locations of entities are predetermined. The only additional interaction that participants could do is either highlighting or eliminating entities. Thus, the Timeline view appears to be more limited compared to other story building mechanisms, such as the Shoebox feature in Jigsaw [2], which provide various evidence marshaling capabilities.

But if this is true, how were Group T participants able to build stories using the Timeline view? One speculation is that people may not need sophisticated interface support to build a story (as evidenced by oral storytelling tradition). Perhaps the single key feature of the Timeline is that it presents events in chronological order, the precise order necessary to conduct investigative analysis. The temporal entity order and highlight features may be sufficient to help people construct stories out of various reports. Perhaps, Group T participants built stories in their minds using only the external media to help organize the story more efficiently. This speculation is in line with the notion of distributed cognition [33] and its application to visualization [34].

However, we acknowledge that the complexity of the main storyline in our study may not be complex enough, allowing the Timeline view to serve as a story building mechanism in only this particular case. For a more complex storyline (e.g., multiple branching and merging of storylines along the main plots), its lack of spatial interaction may cause it to be insufficient as a story building mechanism because it does not support the expressive power needed by the analyst. In addition, the role of the ER view seem to be still important even though it was not used as a story building mechanism for Group T participants. The time spent on the ER view for Group T is still substantial, and highlighting (i.e., brushing) entities appeared to help Group T participants connect the two views and see relational and temporal information simultaneously. The speculations made on top of our findings in this study should be tested in future research.

6.3 Externalizing Temporal Data

In general, the above benefits of the Timeline view all seem to stem from the fact that this view externalizes temporal relationships in a form more amenable to human comprehension than many other representations. This is an instantiation of the concept of *external cognition* [16], which has been quoted as one of the primary mechanisms of general visualization. Adapting Scaife and Rogers' terminology, we think that timeline representations aid the user in the following ways:

- **Computational offloading.** Timeline representations make the story of the data explicit in the world (i.e., on the computer screen), reducing the need for users to mentally formulate and store this information in their minds.
- **Re-representation.** Temporal order is key to uncovering causal information, which in turn is central to identifying an overarching plot in investigative analysis. Unlike other visual representations, timeline representations make the temporal order between events clearly visible.
- **Graphical constraining.** Mapping time onto a screen dimension provides an explicit graphical constraint on the temporal order of events that is not present in the ER view, where the nodes are placed according to different criteria. This constraining allows users to quickly rule out impossible hypotheses (e.g., effects before causes).

Naturally, the above mechanisms are true for all visual representations in general. However, this treatment makes some progress toward understanding the actual mechanics of why timeline visualizations are useful. Time is a fundamental aspect in our world and for visual analytics [1], so these findings will be useful in acquiring a better understanding of how visualization helps the user understand the data.

6.4 Design Guidelines

Based on our findings and discussions above, we provide below a set of design guidelines for temporal visualization when used in investigative analysis:

Supporting temporal analytic tasks. We found that temporal visualization could help discern which entities and relationships are important or not by presenting the following patterns:

1. entities appearing multiple times over the timeline;
2. several entities co-occurring multiple times over the timelines;
3. entities appearing before and after a certain time;
4. two entities appearing only once together but separately appearing multiple times over the timeline.

When designing future temporal visualizations, the designer should confirm that these patterns are indeed visible.

Combining temporal and relational information. We also found that the ER and Timeline views were often used together by Group T. Thus, our recommendation is to provide both relational and temporal information, or, better yet, to create visualizations that combine relational and temporal information in the same view. In addition, interaction techniques such as brushing and linking, which combine data from temporal and relational views, should be used.

Supporting story building. Although the importance of supporting story building is well understood and accepted in the visual analytics community, our findings on the use of the Timeline view are intriguing and suggest a need to study this topic in much more detail. In our particular study, the single key feature of the temporal visualization seemed to be that it showed selected events in chronological order: this was enough to off-load the participants sufficiently so that they felt no need to externalize the storyline in the ER view. While this may be an effect of the relatively small data set we used in the study, this in turn suggests that visual representations do have a significant impact on the cognitive effort of the user. A story building mechanism need not be overly complex and full-featured if the temporal visualization provides sufficient information for the user to be able to reconstruct the storyline in their head.

6.5 Evaluating Visual Analytics: Lessons Learned

We did not record video due to its high cost in codifying and analysis, which went against our lightweight evaluation approach. We also found in pilot testing that video was of limited use since our study was based on single-user analysis restricted only to the TIMEINVESTIGATOR tool and with no external aids. In other words, the optimal use of video in our study would be to record the contents of the screen. In general, we feel that this gives rise to a recommendation on how to capture user behavior during investigative analysis: *select user behavior capture mechanisms by carefully considering analysis costs versus potential gains.*

Based on this reasoning, our TIMEINVESTIGATOR system was heavily instrumented to capture large amounts of interaction data and screenshots during each experimental session. However, this left us with tens of thousands of lines of log events. Our solution was to turn the analysis of visual analytics evaluation into a visual analytics problem of its own. In this paper, we have explored ways of applying visualization techniques to both analyze our data as well as to expose it to our audience. We are surprised by the scarcity of such approaches in the literature (notable exceptions include Isenberg et al. [24], Robinson [25], and Tang et al. [35]), and thus we feel our recommendation on this is both novel and useful: *use visualization to analyze complex evaluation results.*

Furthermore, inspired by the insight-based evaluation employed by Saraiya et al. [28], we introduced a replacement of the think-aloud protocol in the form of our impromptu insight reports and screenshots that we use for collecting the participants' thought process throughout analytical sessions. We think that this is a useful technique for visual analytics evaluation because of its smooth integration into the analytical process (noting down intermediate thoughts and ideas is not uncommon when studying complex problems), so we recommend to *allow participants to record insights and results throughout a session, and not just at the end.*

6.6 Limitations

Below we discuss the most important limitations of our study and their potential impact on our findings.

6.6.1 Custom Tool

We developed a custom tool—TIMEINVESTIGATOR—for this evaluation rather than using an existing, established tool like Jigsaw [2], CzSaw [21], or the Analyst's Notebook [30]. This means that our results may be more difficult to generalize to other tools for investigative analysis. Furthermore, it could also be argued that our implementations for different views are suboptimal compared to those of established tools.

However, we made this decision because of the need to be able to fully instrument the tool with our testing environment, and to constrain the tool to have a minimal subset of operations. Existing tools often have several ways of accomplishing the same task, whereas our approach allowed us to make the different views of the tool as orthogonal as possible.

6.6.2 Participant Expertise

Just like Kang et al. [4], our evaluation included only novice analysts from the student population at our university. As a result, our results may have been different if the study participants had been professional intelligence analysts. However, as noted by Kang et al., intelligence analysts are a small and highly inaccessible population, so including them in exploratory studies of this nature is difficult, at least in a university setting. Given the lightweight evaluation methodology used in this paper, we wanted to investigate the depth and breadth of findings possible even with nonprofessional analysts as study participants.

What impact this choice had on our results is difficult to establish. The fact that we constrained participants to using an unfamiliar tool would eliminate effects of practice and presumably uncover emerging strategies that would be same across both populations. All participants also received 20 minutes of training using a small data set before each session.

Furthermore, our participants were all engineering students, whereas many analysts may come from a broader social or political science background. Again, we are unable to predict what impact this difference would have on the results: one hypothesis may be that social and political science majors are more accustomed to reading and summarizing large amounts of text, whereas the visual representations used in our tool would benefit engineering students better. Additional evaluation is needed to answer such questions.

6.6.3 Solution Grading

Kang et al. [4] used external reviewers (graduate students in the larger research group) to grade the solutions derived by each participant to avoid bias. We used two of the authors of the paper to grade and code the insight reports independently.

To maintain objectivity in spite of this fact, we established a strict coding rubric, performed the two coding sessions independently, and then merged and discussed the results into a single coding metric. This approach is common practice in much qualitative evaluation in social science. Therefore, we do not think it affected the results significantly.

7 CONCLUSION AND FUTURE WORK

We have presented a qualitative evaluation of temporal visualization for investigative analysis. Our evaluation clearly showed that having a temporal visualization (the Timeline view) provides participants with additional aids to find important clues and falsify irrelevant information, so that they more easily can find the correct solution. These positive outcomes are a result of the temporal view not only serving as a passive view showing temporal information, but also serving as an external memory aid for viewing complex event sequences and for building storylines.

It is clear that visual analytics evaluation is still a wide-open research topic. Our future work will focus on studying the analytical process in more detail. In particular, we think that the low-cost evaluation approach used in this paper will be helpful in extending our studies of investigative analysis to other settings, scenarios, and tasks beyond the intelligence domain.

ACKNOWLEDGMENTS

REFERENCES

[1] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visual Methods for Analyzing Time-Oriented Data," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 1, pp. 47-60, Jan./Feb. 2008.

[2] J.T. Stasko, C. Görg, and Z. Liu, "Jigsaw: Supporting Investigative Analysis through Interactive Visualization," *Information Visualization,* vol. 7, no. 2, pp. 118-132, 2008.

[3] R. Eccles, T. Kapler, R. Harper, and W. Wright, "Stories in GeoTime," *Information Visualization,* vol. 7, no. 1, pp. 3-17, 2008.

[4] Y. Kang, C. Görg, and J. Stasko, "Evaluating Visual Analytics Systems for Investigative Analysis: Deriving Design Principles from a Case Study," *Proc. IEEE Symp. Visual Analytics Science and Technology,* pp. 139-146, 2009.

[5] E.A. Bier, S.K. Card, and J.W. Bodnar, "Entity-Based Collaboration Tools for Intelligence Analysis," *Proc. IEEE Symp. Visual Analytics Science and Technology,* pp. 99-106, 2008.

[6] D.H. Jeong, W. Dou, F. Stukes, W. Ribarsky, H.R. Lipford, and R. Chang, "Evaluating the Relationship between User Interaction and Financial Visual Analysis," *Proc. IEEE Symp. Visual Analytics Science and Technology,* pp. 83-90, 2008.

[7] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort, "The Sandbox for Analysis: Concepts and Methods," *Proc. ACM Conf. Human Factors in Computing Systems,* pp. 801-810, 2006.

[8] S. Carpendale, "Evaluating Information Visualizations," *Information Visualization: Human-Centered Issues and Perspectives,* vol. 4950, pp. 19-45, 2008.

[9] C. Plaisant, "The Challenge of Information Visualization Evaluation," *Proc. ACM Conf. Advanced Visual Interfaces,* pp. 109-116, 2004.

[10] C. Plaisant, G.G. Grinstein, J. Scholtz, M. Whiting, T. O'Connell, S.J. Laskowski, L. Chien, A. Tat, W. Wright, C. Görg, Z. Liu, N. Parekh, K. Singhal, and J.T. Stasko, "Evaluating Visual Analytics at the 2007 VAST Symposium Contest," *IEEE Computer Graphics and Applications,* vol. 28, no. 2, pp. 12-21, Mar./Apr. 2008.

[11] C. North, "Toward Measuring Visualization Insight," *IEEE Computer Graphics and Applications,* vol. 26, no. 3, pp. 6-9, May/June 2006.

[12] G. Klein, B.M. Moon, and R.R. Hoffman, "Making Sense of Sensemaking 1: Alternative Perspectives," *IEEE Intelligent Systems,* vol. 21, no. 4, pp. 70-73, July/Aug. 2006.

[13] R.J. Heuer, *Psychology of Intelligence Analysis.* United States Govt. Printing Office, 1999.

[14] G. Chin Jr., O.A. Kuchar, and K.E. Wolf, "Exploring the Analytical Processes of Intelligence Analysts," *Proc. ACM Conf. Human Factors in Computing Systems,* pp. 11-20, 2009.

[15] D.A. Norman, "Cognition in the Head and in the World," *Cognitive Science,* vol. 17, no. 1, pp. 1-6, Jan.-Mar. 1993.

[16] M. Scaife and Y. Rogers, "External Cognition: How Do Graphical Representations Work?," *Int'l J. Human-Computer Studies,* vol. 45, no. 2, pp. 185-213, 1996.

[17] D. Peuquet, "It's about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems," *Annals of the Assoc. of Am. Geographers,* vol. 84, no. 3, pp. 441-461, Sept. 1994.

[18] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman, "LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records," *Proc. ACM Conf. Human Factors in Computing Systems,* pp. 221-227, 1996.

[19] T.D. Wang, C. Plaisant, A.J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman, "Aligning Temporal Data by Sentinel Events: Discovering Patterns in Electronic Health Records," *Proc. ACM Conf. Human Factors in Computing Systems,* pp. 457-466, 2008.

[20] K. Wongsuphasawat and B. Shneiderman, "Finding Comparable Temporal Categorical Records: A Similarity Measure with an Interactive Visualization," *Proc. IEEE Symp. Visual Analytics Science and Technology,* pp. 27-34, 2009.

[21] N. Kadivar, V. Chen, D. Dunsmuir, E. Lee, C. Qian, J. Dill, C. Shaw, and R. Woodbury, "Capturing and Supporting the Analysis Process," *Proc. IEEE Symp. Visual Analytics Science and Technology,* pp. 131-138, 2009.

[22] J. Scholtz, "Beyond Usability: Evaluation Aspects of Visual Analytic Environments," *Proc. IEEE Symp. Visual Analytics Science and Technology,* pp. 145-150, 2006.

[23] M. Tory and T. Möller, "Evaluating Visualizations: Do Expert Reviews Work?" *IEEE Computer Graphics and Applications,* vol. 25, no. 5, pp. 8-11, Sept./Oct. 2005.

[24] P. Isenberg, A. Tang, and M.S.T. Carpendale, "An Exploratory Study of Visual Information Analysis," *Proc. ACM Conf. Human Factors in Computing Systems,* pp. 1217-1226, 2008.

[25] A.C. Robinson, "Collaborative Synthesis of Visual Analytic Results," *Proc. IEEE Symp. Visual Analytics Science & Technology,* pp. 67-74, 2008.

[26] D. Gotz and Z. Wen, "Behavior-Driven Visualization Recommendation," *Proc. Int'l Conf. Intelligent User Interfaces,* pp. 315-324, 2009.

[27] K.S. Park, A. Kapoor, and J. Leigh, "Lessons Learned from Employing Multiple Perspectives in a Collaborative Virtual Environment for Visualizing Scientific Data," *Proc. ACM Conf. Collaborative Virtual Environments,* pp. 73-82, 2000.

[28] P. Saraiya, C. North, V. Lam, and K.A. Duca, "An Insight-Based Longitudinal Study of Visual Analytics," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 6, pp. 1511-1522, Nov./Dec. 2006.

[29] J. Scholtz, E. Morse, and T. Hewett, "In Depth Observational Studies of Professional Intelligence Analysts," *Proc. Human Performance, Situation Awareness and Automation Technology Conf.,* 2004.

[30] i2 Incorporated "Analyst's Notebook," http://www.i2inc.com/, Mar. 2011.

[31] J.S. Yi, Y. ah Kang, J.T. Stasko, and J.A. Jacko, "Toward a Deeper Understanding of the Role of Interaction in Information Visualization," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1224-1231, Nov./Dec. 2007.

[32] D. Kirsh, "The Intelligent Use of Space," *Artificial Intelligence,* vol. 73, pp. 31-68, 1995.

[33] E. Hutchins, *Cognition in the Wild.* MIT Press, 1995.

[34] Z. Liu, N.J. Nersessian, and J.T. Stasko, "Distributed Cognition as a Theoretical Framework for Information Visualization," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1173-1180, Nov./Dec. 2008.

[35] A. Tang, M. Pahud, S. Carpendale, and B. Buxton, "VisTACO: Visualizing Tabletop Collaboration," *Proc. ACM Conf. Interactive Tables and Surfaces,* pp. 29-38, 2011.

Bum chul Kwon received the master degree in industrial engineering from Purdue University in West Lafayette, Indiana, in 2010. Currently, he is working toward the PhD degree in the School of Industrial Engineering at Purdue University in West Lafayette, Indiana. His research interests include information visualization, visual analytics and their applications in consumer decision making and the healthcare industry. He is a student member of the IEEE.

Waqas Javed received the bachelor of science degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2007. Currently, he is working toward the PhD degree in the School of Electrical and Computer Engineering at Purdue University in West Lafayette, Indiana. His research interests include information visualization, visual analytics, and human-computer interaction. He is a student member of the IEEE.

Sohaib Ghani received the bachelor of science degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2007. Currently, he is working toward the PhD degree in the School of Electrical and Computer Engineering at Purdue University in West Lafayette, Indiana. His research interests include information visualization, visual analytics, and human-computer interaction. He is a student member of the IEEE.

Niklas Elmqvist received the PhD degree from Chalmers University of Technology in Göteborg, Sweden, in 2006. Currently, he is working as an assistant professor in the School of Electrical and Computer Engineering at Purdue University in West Lafayette, Indiana. He was previously a postdoctoral researcher at INRIA in Paris, France. He is a member of the IEEE and the IEEE Computer Society.

Ji Soo Yi received the PhD degree from Georgia Institute of Technology, Atlanta, in 2008. Currently, he is working as an assistant professor in the School of Industrial Engineering at Purdue University in West Lafayette, Indiana. He is a member of the IEEE.

David S. Ebert received the PhD degree from the Ohio State University in Columbus, in 1999. Currently, he is the Silicon Valley professor in the School of Electrical and Computer Engineering at Purdue University in West Lafayette, Indiana. He is a fellow of the IEEE and the IEEE Computer Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

Feature-Driven Data Exploration
for Volumetric Rendering

Insoo Woo, *Member, IEEE*, Ross Maciejewski, *Member, IEEE*,
Kelly P. Gaither, *Member, IEEE*, and David S. Ebert, *Fellow, IEEE*

Abstract—We have developed an intuitive method to semiautomatically explore volumetric data in a focus-region-guided or value-driven way using a user-defined ray through the 3D volume and contour lines in the region of interest. After selecting a point of interest from a 2D perspective, which defines a ray through the 3D volume, our method provides analytical tools to assist in narrowing the region of interest to a desired set of features. Feature layers are identified in a 1D scalar value profile with the ray and are used to define default rendering parameters, such as color and opacity mappings, and locate the center of the region of interest. Contour lines are generated based on the feature layer level sets within interactively selected slices of the focus region. Finally, we utilize feature-preserving filters and demonstrate the applicability of our scheme to noisy data.

Index Terms—Direct volume rendering, transfer function, focus+context visualization.

✦

1 INTRODUCTION

THE emergence of web-based scientific simulation portals [1], [2] has created an abundance of volumetric data, and an effective means of exploring volumetric data is through direct volume rendering. One of the most common methods for direct volume rendering is to employ the use of interactive transfer function widgets which users interactively use to define a mapping of the volumetric data values to optical properties (color and opacity). Unfortunately, creating an appropriate transfer function often involves tedious adjustment and fine tuning of parameters, resulting in a trial-and-error type approach by the user. This problem is further compounded in scientific portals (e.g., the NanoHub [3]), as the scientists who need to analyze the data are often novices in volumetric rendering. Furthermore, relevant sparse features within the volumetric data can easily be occluded by semitransparent surrounding values, making it difficult to find the specific value range within the occluded regions that should be enhanced for better visualization and analysis. This volumetric exploration problem is illustrated in Fig. 1.

In Fig. 1, a scientist has created an Indium-Arsenic (InAs) quantum dot simulation [4], and the scientist wants to explore the electron wave function data encapsulated in the resultant volumetric output. Here, the important structure to the end user is the vertical stack consisting of three InAs quantum dots. However, this structure is not easily found using 1D and 2D histogram widgets. The transfer function widget in Fig. 1 (middle) uses a 2D histogram widget [5] plotting the value versus value gradient magnitude of the volumetric data. The user has drawn a series of rectangular boxes within the widget. The structures found in region A of the histogram are of little to no importance to the scientist, and, while the most important structures of the data are located in the histogram region of Fig. 1 marked B, the data in region B are so sparse that there are no visual cues to guide the user to select those regions in the histogram space. Note that both a linear and logarithmic mapping of the opacity was applied in the 2D histogram with similar results.

Thus, novice end users of direct volume rendering tools need methods where rendering parameters can be suggested in a way that incorporates the end-user's expert knowledge of interest within the data set and the properties of the volumetric data. In this work, we propose a novel interaction approach to semiautomatically generate appropriate rendering parameters to help users efficiently explore and analyze their volumetric data sets. Key contributions include the following:

- A hardware-accelerated graphics pipeline (Fig. 2) that enables interactive visualization through a user-specified region of interest. Features within the region of interest are analyzed and automatically mapped to a set of rendering parameters based on the features. Moreover, our pipeline includes feature-preserving denoising methods (e.g., median filters) to support noisy volumetric data analysis.
- A novel selection scheme for color mapping and data enhancement based on data types (interval and ratio), utilizing measurement theory. Features of interest can be highlighted based on regions and focus values.

- *I. Woo is with the Purdue Visual Analytics Center, Purdue University, PO Box 519, 465 Northwestern Ave., West Lafayette, IN 47907. E-mail: iwoo@purdue.edu.*
- *R. Maciejewski is with the Arizona State University, PO Box 878809, Tempe, AZ 85287. E-mail: rmacieje@asu.edu.*
- *K.P. Gaither is with the Texas Advanced Computing Center, University of Texas, Research Office Complex 1.101, J.J. Pickle Research Campus, Building 196, 10100 Burnet Road, Austin R8700, TX 78758-4497. E-mail: kelly@tacc.utexas.edu.*
- *D.S. Ebert is with Purdue Visual Analytics Center, Purdue University, 465 Northwestern Ave., West Lafayette, IN 47907. E-mail: ebertd@purdue.edu.*

Fig. 1. Comparison of transfer function design between a 1D histogram widget (left), a 2D histogram widget (center), and our semiautomated transfer function widget (right). In the 2D histogram, the region A includes the majority of the volumetric data; however, the most important information is found in the sparse areas (B) in the 2D histogram. In our method, the user is able to highlight meaningful boundaries using the value profile for the region of interest. The "top view" (rightmost figure) shows the ray (in red) passing through the volume and the line graph displays the scalar values and gradient magnitudes along the ray (in red).
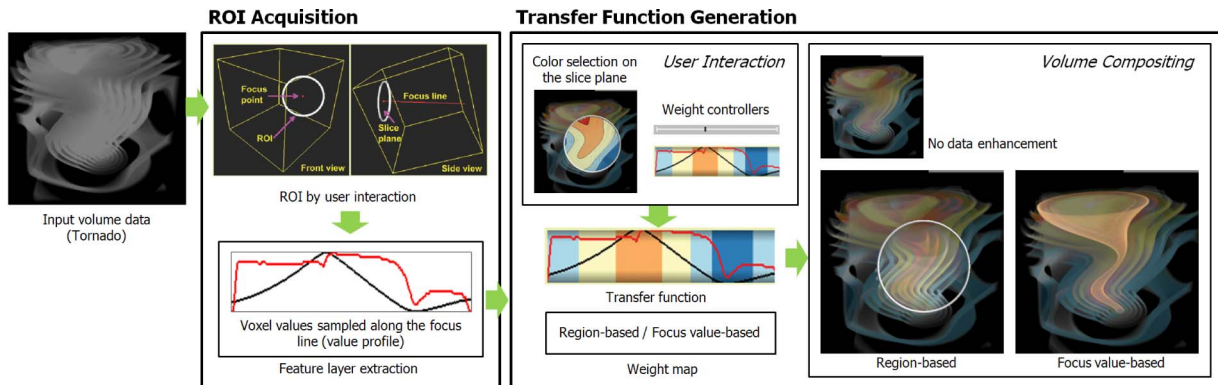
Fig. 2. Schematic diagram of our semiautomated data exploration method. The sequence consists of user's region of interest acquisition, feature extraction, color selection, and data enhancement.

- An intuitive user interface for transfer function design (Fig. 3), modification, and interaction utilizing line charts and contour lines within a slice view for enhancing local data features. Level sets are extracted
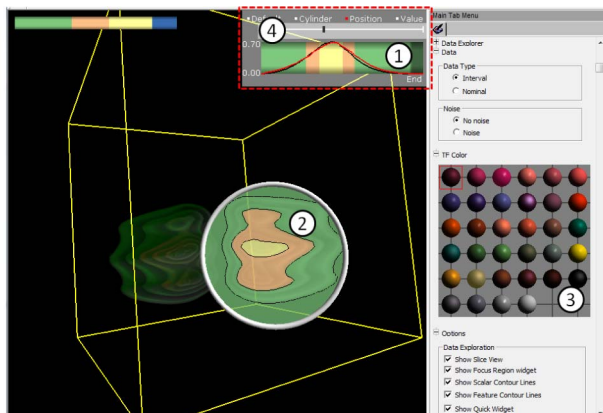
Fig. 3. The user-interface controls of our interactive feature-driven exploration tool. 1: The value profile widget. 2: The contour line widget. 3: The color palette selection widget. 4: The focus region widget.

from a slice view using a CUDA [6] kernel, contour lines are displayed on the view, and then colors can be changed on the slice view based on the level set.

The use of the feature selection tool enables users to generate appropriate transfer functions more effectively. Furthermore, the addition of the value profile widget provides simulation scientists with another view in which to analyze their data. Finally, our semiautomated tools make use of line graphs and heightmaps that are generated by sampling scalar values along a specific line segment, or on a specific slice surface within the volumetric data. By applying both line graph and contour line metaphors within a volume visualization application, users can further investigate and analyze data features by specifying the *spatial region* of interest within the volumetric data or by varying the *scalar value range* of interest to highlight features.

In this paper, we will discuss the advantages of our pipeline and interactive widgets and their ability to enhance the overall visualization and analysis process for the end user. In the next section, we summarize the related work in transfer function design, color selection, and image denoising. In Section 3, we describe our pipeline and algorithms

for interactive transfer function design. In Section 4, we explain the implementation details of the algorithm. Finally, we discuss our results and present feedback from end users in Section 5 and propose future work in Section 6.

2 RELATED WORK

Interactive transfer function design has been addressed through many different approaches. Direct manipulation widgets help users to explore volumetric data sets by using data probes, clipping planes, classification, shading, and color picking widgets [5], [7]. Some approaches have utilized design galleries [8] or parallel coordinate style interfaces [9] for transfer function design, while other recent approaches have focused on user-oriented design schemes. For example, Rezk-Salama et al. [10] provided a high-level transfer function model using widgets that map directly to volume features (such as controls for bone color and opacity). Wu and Qu [11] designed a framework to manipulate rendered images for multiple volumetric objects and merge the images into a single rendering. Bruckner and Gröller [12] employed a stylized transfer function widget that enables users to intuitively select a nonphotorealistic shading style from the style transfer function. Bajaj et al. [13] introduced the contour spectrum interface that presents users with a collection of data characteristics to select significant isovalues for isosurfacing. Lundström et al. [14] introduced the α-histogram which incorporates spatial coherence as a means of automated transfer function design. This work divides the data into subblocks (spatial sub regions), computes local histograms, amplifies peaks by raising histogram values to the power of α, and calculates an α histogram by summing up all the amplified histograms. The work by Lundström et al. provides transfer functions on a global scale as opposed to our work, which focuses on enhancing local data features. Correa and Ma [15] proposed a new volume exploration technique by introducing the concept of a scale field that allows users to assign colors and opacities based on size. However, the concept of a scale field works best when the data sets have a well-defined underlying shape, which is often not the case in scientific simulations.

Besides ill-defined volumetric structures, another issue is that the size and complexity of scientific simulations has also been drastically increasing. As such, recent work has focused on the use of focus+context visualization [16]. In order to enable users to select a region of interest with ease, research has focused on two primary tasks: defining a focus and context area, and effectively emphasizing the focused area. Lu et al. [17] inferred a user's regions of interest by capturing eye gaze positions as the user watched a rotating volume. Li et al. [18] proposed a system for automatically generating a variety of cut-away illustrations. Viola et al. [19] introduced cut-away views for volume rendering based on importance by suppressing less important information. However, their method requires data segmentation and user-assigned importance values to each data segment. Ropinski et al. [20] introduced a stroke-based transfer function design scheme computing a difference histogram for all pairs of the control points along inner and outer strokes [20]. For better peak detection, this work used

weights based on the voxels' visibility. It works properly when the data have distinctive volume object boundaries since the user can easily draw stroke lines along the intensity boundary. However, many simulation data sets lack well-defined edges and boundaries.

Along with focus+context methods, raycasting-based data analysis methods have also been employed to aid users when exploring their volumetric data. These methods create automated or semiautomated rendering parameter settings or provide an interface supporting volumetric layer exploration. Opacity peeling [21] generates an image of an opacity layer (image buffer) whenever an accumulated alpha value exceeds a threshold and allows the user to explore the opacity layers. Malik et al. [22] extended the concept by analyzing values sampled along rays through the volumetric data and extracting feature layers. Instead of accumulating opacity, Malik's work searches for transition points from the sampled values to extract feature layers with prominent peaks. Correa et al. [23] introduced the visibility histogram where, for a given viewpoint, visibility is weighted using opacity, which is considered as importance, and used to maximize the visibility of the value range of interest. Kohlmann et al. [24] used ray profiles to facilitate the process of finding and highlighting interest points in 2D slice views starting from the users' mouse click location in the 3D view. However, their main focus is on the construction of a ray profile library that stores ray profile samples for different CT data sets and on matching profiles to find close similarities.

Another factor to consider in scientific simulation data is that the data are often temporal in nature. As such, previous work has also focused on defining a transfer function for time-varying volumetric data. Recent work by Akiba and Ma [25], Akiba et al. [26] utilized parallel coordinate plots to create a volume rendering interface for exploring multi-variate time-varying data sets. By means of a prediction-correction process, Muelder and Ma [27] proposed a method to predict the feature regions in the previous frame, making the feature tracking coherent and easy to extract the actual features of interest. Maciejewski et al. [28] used density estimation for creating a temporally coherent set of transfer functions of a group of feature space histograms.

While these previous methods have attempted to provide user friendly interfaces for transfer function manipulation, there are still barriers to the adoption of the methods by scientists and general users. In this paper, we focus on a new data exploration and transfer function design scheme. We provide the end user with familiar ray profile plots and an interactive means of defining regions of interests, as shown in Fig. 3. From these tools, we are able to semiautomatically generate rendering parameters and reduce some of the transfer function design burden for the end user.

3 INTERACTIVE FEATURE-DRIVEN DATA EXPLORATION

Since analysis methods are affected by the structure and nature of data [29], we define our visual mapping parameter choices based on the underlying data type. In measurement theory, the measurement scale is categorized into four

TABLE 1
Color and Opacity Selections Based on Their Data Type

Data Type	Noise	Filtering	Color Sel.	Opacity Enhancement
Interval	with	O	Sequential	Local peak value per layer
	without	X		
Ratio	with	O	Diverging	Local maximum gradient magnitude per layer
	without	X		

categories (nominal, ordinal, interval, and ratio) [30]. Nominal data have no order, ordinal data have order, interval data have meaningful intervals, and ratio data have the height level of measurement. These different types of data have features that need to be highlighted in different ways [29]. In the case of volumetric data, we consider two categories of data (Table 1), interval and ratio data, and generalize a set of guidelines and rules for automated or semiautomated methods for our visualization techniques. Nominal and ordinal data types are left for future consideration.

The color selection schemes, emphasized areas, and filtering parameters are semiautomatically determined by the user-defined data type. These schemes are applied automatically in our transfer function generation pipeline depending on the data type. Fig. 2 illustrates the sequence of operations in our approach. First, we provide the user with an initial rendering of their data. During our first pass, transfer function parameters are generated using a sine function that maps the scalar values to an opacity, thus creating high and low contours [31] as seen in Fig. 2 (input volume data). Next, using the initial volume rendering, the user interacts with the volume. Mouse clicks indicate the user's region of interest (Fig. 2, ROI acquisition), and a focus point and a focus region are created resulting in a focus line that is a viewing ray passing through the focus point. Then, our system extracts feature layers to compute the number of colors and the data enhancement points and displays the results in a value profile widget. A color transfer function and a weight map are automatically defined based on the extracted feature layers (Fig. 2, transfer function and weight map). The weight map is utilized to modulate the opacity. In this way, users can explore their data in focus-region-based or focus-value-driven ways (Fig. 2, volume composition). In addition, we apply nonlinear filters to the value profiles and slice view images to support our transfer function generation approach for noisy volumetric data sets.

3.1 Region of Interest (ROI) Acquisition

The region of interest acquisition is the first step in our data exploration method. Earlier methods for ROI selection included the use of the mouse cursor along with several prespecified context layers from the volume data set [32]. However, this method entails the tedious process of manually creating several context layers that capture important features within the data set. Our work chooses to utilize a simpler interface design in which mouse movements are used to obtain the ROI within the volume data from the 2D screen. First, the user mouse clicks on the region of interest and drags the mouse in order to define the size of their region of interest. The mouse click defines the initial center of the ROI.

Once the ROI is determined, a ray profile is obtained and analyzed. The initial center of the ROI is then repositioned from the original location to a new locally optimal location based on analysis of the ray profile in the area selected and the user selected datatype. Based on the datatype, the ROI is centered at either the maximum gradient magnitude or at the maximum scalar value in the 3D neighborhood. Then, the user can shift the default position of the ROI along the viewing ray, allowing him or her to acquire a different ROI along this focus line.

The repositioning and analysis of the ray profile is the main step in our semiautomated data exploration. Since features can be identified with local peaks and maximum gradient magnitudes obtained by sampling along the ray at regular intervals, we employ the concept of a feature layer to generate the visual mapping based on these local peaks and maximum gradient magnitudes. Each layer corresponds to the minimal pairwise spatial distance calculated between peaks along the ray profile. A feature layer stores the minimum and maximum of the local scalar values and the minimum and maximum of the local gradient magnitudes to determine a value range for the color and data enhancement within each layer. Depending on the type of volumetric data (interval or ratio as defined by the user), the maxima and minima of these values are used to determine the transition points (as shown in the value profile widget of Fig. 3 and detailed in Fig. 4) that divide the profile into multiple layers.

We first extract feature layers based on the local extrema along the ray profile and then use these feature layers for data exploration. Malik et al. [22] employed such a ray profile for feature peeling. Multiple rays are cast to extract scalar values along the rays. This approach allows the users to explore volumetric data sets layer-by-layer. We also use the variation of scalar values along the rays to extract feature layers. In addition, we use the local peaks, both from scalar values and gradient magnitude, of each layer to provide color assignment and data enhancement for features of interest in layers according to the data type of the input volume data set. Thus, each layer is found
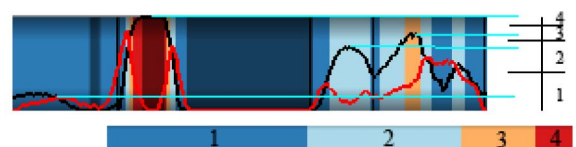
Fig. 4. The larger color bar on the top displays the scalar (black) and gradient magnitude (red) value profiles. Three vertical black lines indicate the positions of transition points that create four feature layers. The vertical bar on the right is divided into four regions bounded by the scalar value ranges of the corresponding colors in the color map shown in the color bar below.

Fig. 5. Value profile of a noisy MRI data. (a) Original value profile of the scalar values and (b) same profile filtered with a median low pass filter.

between two extrema of either the scalar or gradient values along the ray profile. The ordering of the layers by scalar value is done so that the correct color scheme will be mapped to each layer. As shown in Fig. 4, the layers are perceptually ordered by color; however, the ordering of the layers on the ray profile would not match the perceptual color ordering without this step.

In the color selection stage (Section 3.2.1), we assign a unique color to each identified layer. In our work, we utilize the Color Brewer [33] color schemes and match each scheme to the appropriate data type (nominal maps to qualitative, interval to sequential, ratio to divergent, and ordinal to sequential). However, in some cases, we may not be able to extract the necessary number of layers either due to the difficulty in determining feature layers (for example, in the case of gradually increasing scalar values) or due to the presence of excessive layers. To overcome this, we add a user-defined parameter to specify the number of feature layers, i.e., the number of colors, to use in these cases. When the number of extracted layers does not exceed this user-defined parameter, we iteratively sort the layers based on their value range. Then, we select the largest layer, and split it into two until the number of layers is equal to the user-specified parameter value. When the number of extracted layers exceeds this user-defined parameter, we automatically merge layers with similar local peaks until the number of layers is equal to the specified parameter value. The layers are sorted in increasing order of the maximum peak values, the peaks of two consecutive layers are compared and the two layers are merged into one if the distance between the two peaks was the smallest distance between all other peaks.

In the presence of noisy volumetric data, the problem of extracting layers from volume profiles is further compounded. In our data exploration and transfer function design scheme, we use contour lines to represent significant value ranges in the region of interest. Unfortunately, contour lines for noisy volumetric data sets contain many artifacts which can confuse the end user. Our work follows the contour conventions proposed by Viola et al. [34] while providing a more general interface for generating contours while removing noise. Our system uses nonlinear filters, parallelized using CUDA [6], and we compare the quality of the resulting data sets in Section 4. Fig. 5a shows a noisy value profile, here one can imagine that the number of transitions found would result in a large number of nonessential layers. As such, we utilize nonlinear filters such as the median filter that are used to smooth such value variations [22] as shown in Fig. 5b. This smoothes the ray profile and allows for easier feature layer extraction.

The denoising is only done on the ray profile and the slice view. The original data are never denoised, so the output of the pipeline is only slightly different in the noisy versus nonnoisy data. This is due to the fact that a user specifies a maximum number of layers, and layers found
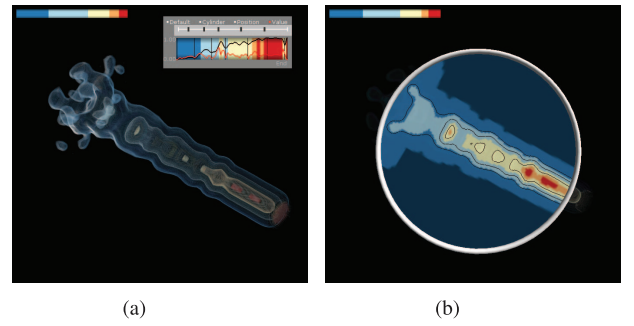
Fig. 6. Initial rendering of a *fuel simulation* data set with the default parameter settings. (a) Shows the changes of scalar values (in black on the top right) and gradient magnitudes (in red) and the regions enhanced based on this profile. (b) Shows the contour lines generated automatically to convey the shapes within the region of interest.

during the computational analysis that are only at a small distance from each other will be merged. In the case of noisy local peaks, the local peaks typically become merged. The denoising is only needed for the contour views.

3.2 Transfer Function Generation

In our approach, the value profile and contour lines are used to provide an interactive environment for data exploration. Fig. 6 illustrates the initial data visualization based on our value profile and contour lines using default color and opacity selection. The line graph on the top right in Fig. 6a shows the scalar and gradient value variations from the value profile. Fig. 6b shows the shapes of the volumetric objects with contour lines in the region of interest. Based on the feature layers, the number of colors and emphasized areas are computed. The color bar on the top left of the figure shows the color transfer function. A cool color is selected as an outer color and a warm color is selected as an inner object using the Color Brewer's diverging color scheme (Red-Yellow-Blue). The user can change the colors directly on the slice view. The ticks above the line graph indicate important features (maximum intensities in each local area) to be enhanced. The user can add or remove ticks for data enhancement.

3.2.1 Color Selection

We assign a unique color to every layer identified during the ROI acquisition phase. Depending on the data type, we use Color Brewer's color schemes (sequential, diverging, and qualitative) by default as listed in Table 1. Fig. 4 shows the assignment of a color map with four colors to a value profile with four layers. The scalar range of each color in the color map is obtained by computing the mean value of each set of adjacent peak values as marked on the vertical line on the right in this figure.

Ideally, the number of colors to be used can be determined from the number of layers. The process of merging similar layers and specifying the number of desired layers in the feature extraction step prevents the assignment of different colors to similar layers and generates an appropriate color transfer function.

3.2.2 Data Enhancement

Our method utilizes color to define a feature; however, feature enhancement is also an important setting for transfer function design. There are several approaches to

(a)   (b)
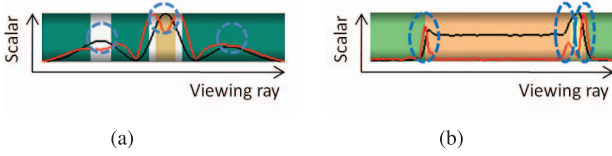
Fig. 7. Enhancement regions identified from the value profile of a region of interest based on the data type. (a) Local peaks of each layer are enhanced for interval data. (b) Local maximum gradients of each layer are enhanced for ratio data.

data feature enhancement. We modulate the opacity values to enhance the data features based on the extracted critical values. As a default opacity setting, we apply a sine function to map the scalar data value to opacity, which creates alternating low and high opacity contours [31], shown on the left side in Fig. 2.

However, different data types have different characteristics that require the emphasis of appropriate regions in the value profile as shown in Fig. 7 and Table 1. For example, ratio data are characterized by the difference of values rather than the sequence of values. Therefore, the boundaries of each layer are more important and need to be emphasized instead of the peaks of each layer. Fig. 8 illustrates the difference between enhancing local peaks and local maximum gradients for a ratio data set (the *tooth* data set).

For the data enhancement, the ranges of local peaks and maximum gradient magnitudes obtained from the value profile are maintained in a weight map and the weights are multiplied by initial opacities to obtain final opacities. We employ a 1D weight map to enhance the features of interest (e.g., local maximum value and local maximum gradient magnitude). Local maximum gradient magnitudes are usually used for the boundary enhancement [35]. However, simulation data sets (e.g., the probability distribution of atoms) tend to have smooth data variations. Therefore, we use the local maximum value for simulation data sets as the default setting, as shown in Fig. 7a. In Kindlmanns approach, the boundaries of objects are defined by sharp curves in the histogram. By utilizing the ray profile, the changes in density along the profile allow us to determine transitions between features that are not necessarily visible in the histogram (as shown in Fig. 1).

The weight map assigns a scalar value scaled from zero to one for each voxel as the voxel's weight. By default, the initial values for all the density bins are set to 0.2. The empirical value 0.2 deemphasizes the high opacities from the default opacities. Typically, the maximum scalar value of a layer and the maximum gradient magnitude should be emphasized in the case of sequential data sets (e.g., simulation data) and ratio data sets (e.g., medical data), respectively. We compute a value range $[\alpha_i, \beta_i]$ to be highlighted for each peak by obtaining the minimum and maximum of scalar values of the N closest neighboring sampling points of the ith peak along the ray. We set the value of N to 7 in this paper.

The weight map is updated using

$$w_j = \begin{cases} 1, & \alpha_i \leq v_j \leq \beta_i, j \in [0, \text{num of bins}), \\ 0.2, & \text{otherwise}, \end{cases} \quad (1)$$

where w_j is the jth weight in the ray profile histogram and v_j is the voxel scalar value. However, even weighted

(a)   (b)
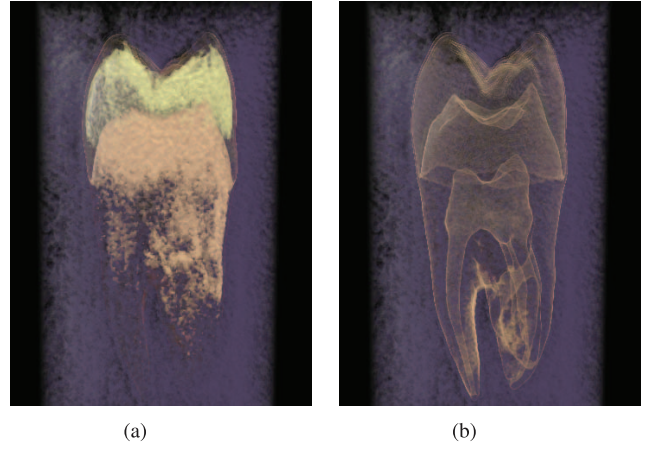
Fig. 8. Images showing the differences arising from enhancing different regions of the *tooth* data set. (a) Image obtained by enhancing local peaks of each layer. (b) Image obtained by enhancing local maximum gradients of each layer.

opacities cannot be used to display the values with zero opacity, as shown by Zhou et al. [36]. Thus, we use (2) to compute the opacity for the range of focus values

$$\alpha_{vd} = \max\{\alpha_o, \alpha_w\} \times w_j, \quad (2)$$

where, α_{vd} is the enhanced opacity, α_o is original opacity, and α_w is the opacity of the windowing function as defined in [36]. Values obtained for the weight mapping and nearest neighbor sampling were obtained through visual analysis of repeated trials working with end users. Future work will focus on determining automatic optimal parameter settings.

Additionally, we apply focus-region-based data enhancement along a ray cast through the volume data. For this enhancement, we assume that the region is cylindrical or spherical. For spherical regions, the center point (C) of the focus region and the radius (r) of the ROI are used to compute the weighted opacities, (5). Data enhancement is performed based on the distance between the position of a voxel and the center point of the focus region.

For cylindrical regions that emphasize the information around the focus line, the two end points (x_1, x_2) of a focus line, a radius (r) of a focus region, and the distance (d) between a sample position (v_p) and the line (x_1, x_2) are used for the computation of weighted opacities. The opacity values are computed according to their shape. For a cylinder, the opacity is computed as

$$\alpha_{rd} = \alpha_{vd} \times \left(k + \left(2 - \frac{d}{r} \right) \times I_c(v_p) \right), \quad (3)$$

$$I_c(v_p) = \begin{cases} 1, & d \leq r, \\ 0, & \text{otherwise}. \end{cases} \quad (4)$$

For a sphere

$$\alpha_{rd} = \alpha_{vd} \times \left(k + \left(2 - \frac{|v_p - C|}{r} \right) \times I_s(v_p) \right), \quad (5)$$

$$I_s(v_p) = \begin{cases} 1, & |v_p - C| \leq r, \\ 0, & \text{otherwise}, \end{cases} \quad (6)$$
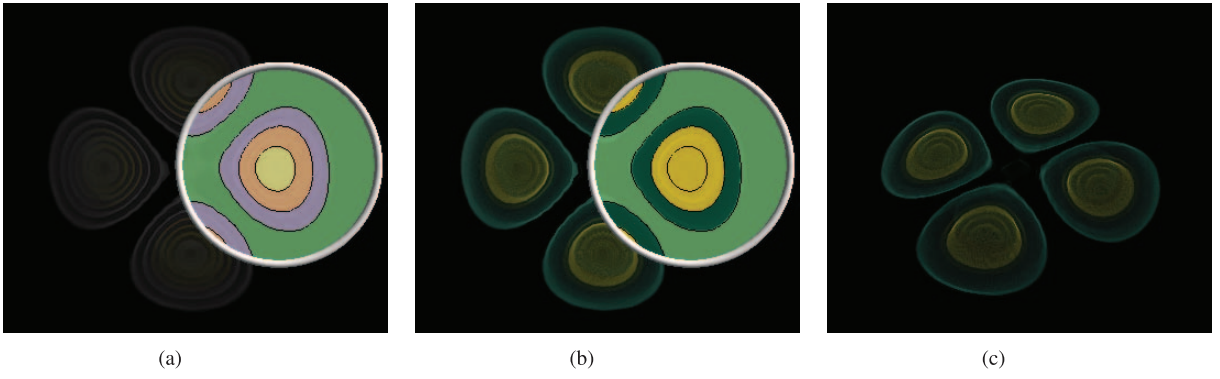
Fig. 9. User-defined color selection on a slice view displaying contour lines. (a) Shows a slice view. (b) Shows the resulting visualization after manually selecting colors and enhancing features of interest. (c) Shows the final resulting image.

where α_{vd} is the enhanced opacity from (2), α_{rd} is the resultant opacity, k is a constant (in this paper, we use $k = 0.5$ for all images for focus-region-based data enhancement), v_p is a sample position, and $I_c(v_p)$ and $I_s(v_p)$ are indicator functions that identify whether a voxel is within the region of interest or not.

3.2.3 Contour Line Extraction

Using a transfer function design interface to highlight a region on the slice view or change the color of the region is often unintuitive as there is no direct link between the slice view and the transfer function interface. In order to better guide users, our system employs the use of contour lines in the slice view to provide intuitive cues for ROI selection. Contour lines have been shown to be effective in non-photorealistic rendering to convey shape information [37] and can be used to illustrate value regions or shapes of simulation data. Moreover, contour lines on a slice view in the region of interest provide intuitive clues to select and enhance the region of interest. As such, we extract and visualize contour lines on the slice view to allow users to change visual properties of the region of interest as well as to convey shape in the selected regions of interest. Users can control the number of level sets by adjusting the sampling space of the contour lines as well as the number of the extracted feature layers. Whenever a user changes the position of the ROI, the view on the slice plane is updated and contour lines are regenerated. We extract contour lines from the image in the slice view to illustrate the shape of the volumetric objects based on the critical boundaries of the data in the regions of interest. Contour lines are extracted using the Marching Square algorithm [38] instead of image-based edge detection algorithms since 3D line primitives obtained from the former can be regenerated and displayed clearly regardless of the magnification and camera position unlike the image-based methods. A 2D slice of the user-specified size is placed and moved along the ray and the slice captures the values and stores them to a render target. The color buffer is then used to extract contour lines.

The number of the level sets of the contour lines is computed and isovalues ranging from 0 to 1 (where the value is internally scaled) are determined for each level to generate contour lines based on the extracted layers from the value profile. For volumetric data with no noise, the contour lines can be directly generated to show critical

boundaries as shown in Fig. 6b. Moreover, we also allow users to select and assign colors to various regions formed by the contour lines on the slice view plane. Fig. 9 shows the user-defined color selection using a slice view.

Unfortunately, noisy volumetric data produce many contour lines, as shown in Fig. 10, that clutter the display, reducing their utility in conveying shape information. Therefore, we apply median and bilateral filters to the image of a slice view to remove the noise before generating contours. We describe the implementation details of contour line generation in the next Section.

4 IMPLEMENTATION

We implemented the data exploration pipeline on a Windows XP PC with an Intel Pentium 4 3.40 GHz CPU, 2 GB RAM, and a GeForce 8800 GTX graphics card. The median filtering with a 3×3 window and contour line
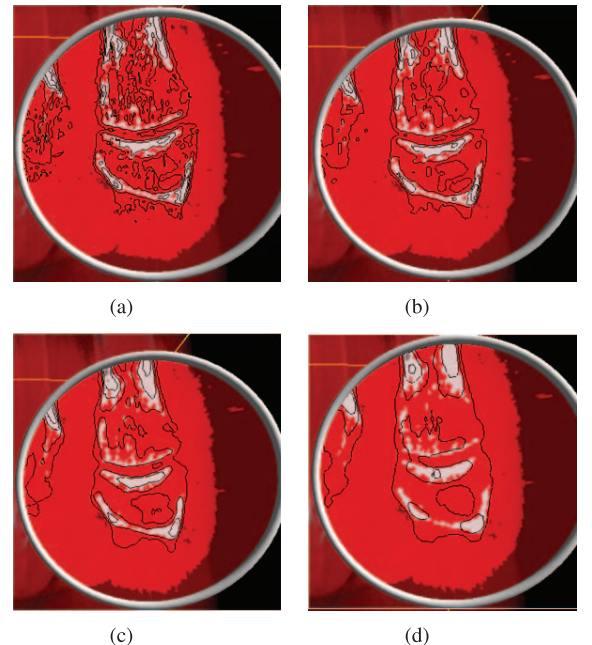
Fig. 10. Images of the *foot* data set rendered by a median filter with various sizes. (a) Shows contour lines when no filters are applied. (b), (c), and (d) Show the contour lines extracted from the image filtered by a 3×3, 5×5, and 9×9 size median filter, respectively.
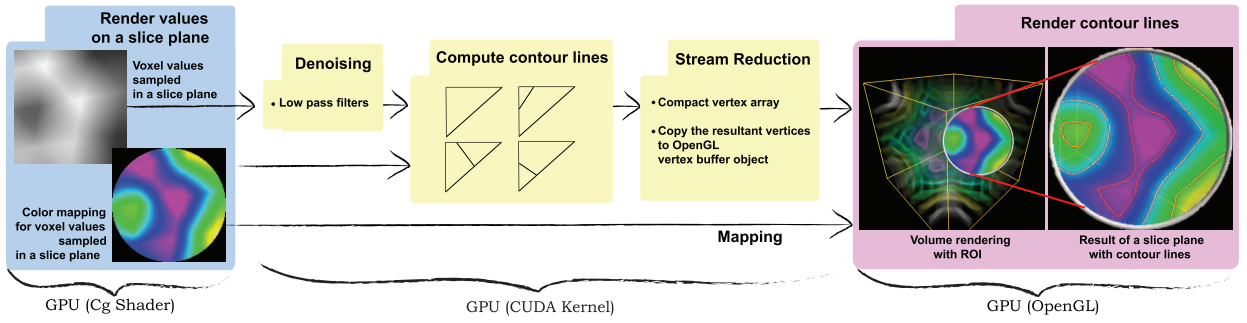
Fig. 11. Pipeline for contour line generation using the CUDA kernel. The first step, implemented in the Cg shader, generates a texture containing scalar values sampled from the volumetric data, which is then mapped to colors on the slice plane. The second step filters noisy data sets to remove noise. The third step computes contour lines and performs stream reduction on the CUDA kernel and the final step renders the computed contour lines.

generation takes 5.7-7.8 and 9.9-10.8 milliseconds for 128×128 and 256×256 grid textures on the GPU, respectively.

This section describes the implementation details of the key aspects in our data exploration pipeline namely value profiling, slice view generation, contour line generation, and image denoising.

Value profiling. To obtain a value profile from the user's mouse click in the ROI, we compute two points, one each on the near and far planes of the view frustum using gluUnproject. These two points, determined using the ray and bounding box intersection test, define the line segment for the value profile. To sample scalar and gradient magnitude values along this ray, sampling points are generated from the starting point to the ending point of the line segment and are stored in a 1D RGB component texture. A rectangle is rendered with this texture to pass the sampling points to the graphics pipeline. In a pixel shader, each texel is used as a sampling position to obtain the scalar value and the gradient magnitude at the corresponding position in the 3D volume texture. The sampled scalar value and the 3D gradient magnitude are stored in the render target. For data sets with noise, scalar, and gradient magnitude values are filtered with a median filter using the host CPU due to the low computational cost. The sampled values are copied to the host memory in order to compute the initial center position of the ROI as well as the feature layers.

Slice view generation. The pipeline for the slice view produces two color buffers in the render target. One stores the scalar and gradient magnitude values for the contour line generation and the other stores the color (RGB) of each scalar value and the transparency (Alpha) to display in the slice view. Images are generated from the slice view by treating it as a circle, and an orthographic camera facing this plane is used to sample values within the circle of the focus region. Regions outside of the circle are rendered as translucent by adjusting the alpha component as shown in Fig. 11.

Contour line generation. Level sets of the contour lines are determined based on the feature layers obtained from the value profile. Fig. 11 illustrates the contour line generation pipeline. The image (2D texture) is treated as a grid and contour lines are computed using the Marching Squares algorithm [38]. However, for better memory efficiency, triangle primitives are used instead of squares because each test using triangles produces zero or one contour line, while the Marching Square test produces zero

to three contour lines. It reduces the memory consumption during the stream reduction operation [39], [40] while the number of iterations over the CUDA kernel becomes twice as many as the level sets. This doubling is due to the fact that the contour lines are extracted from triangles (two triangles per a grid cell). This tradeoff in using triangle primitives over squares is justified because efficient memory consumption is of greater concern than computation time in most web-based portals, such as NanoHub [3], that have to support multiple simultaneous users.

Each triangle primitive is assigned to a thread, and each thread generates zero or one line, as shown in the third module (Contour lines generation) in Fig. 11, and the number of lines is stored in the *flag* in Fig. 12. The fourth step produces a scattered array of line vertices, *lines (scattered)* in Fig. 12. Unwanted elements (no line primitives) from the scattered array are removed by applying stream reduction (using the CUDA Data Parallel Primitive (CUDPP) library [40]) and the resultant array becomes compact like *lines (compact)* in Fig. 12. An OpenGL buffer object is created to store all the line primitives since the CUDA kernel and the OpenGL pipeline can both access the OpenGL buffer. The number of contour lines computed in each iteration is obtained from the stream reduction index array and is used as the offset into the memory block to obtain the memory starting point of the next iteration. Using the CUDA architecture, contour lines of five level sets are generated within 3-4 milliseconds for a 128×128 grid texture on the GPU.

Image denoising. Noisy volumetric data sets result in messy contour lines as shown in Fig. 10. To avoid this, the
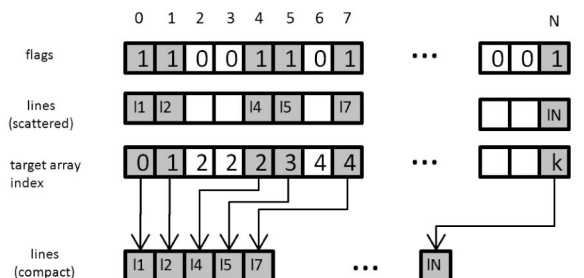
Fig. 12. A schematic diagram of the reduction for contour line generation. *flags* stores the flags that indicate each triangle has a line. *target array index* stores indices for lines to be stored in the compact array of lines. Finally, *line*, the mapped OpenGL VBO, stores compact lines.

(a)   (b)

(c)   (d)
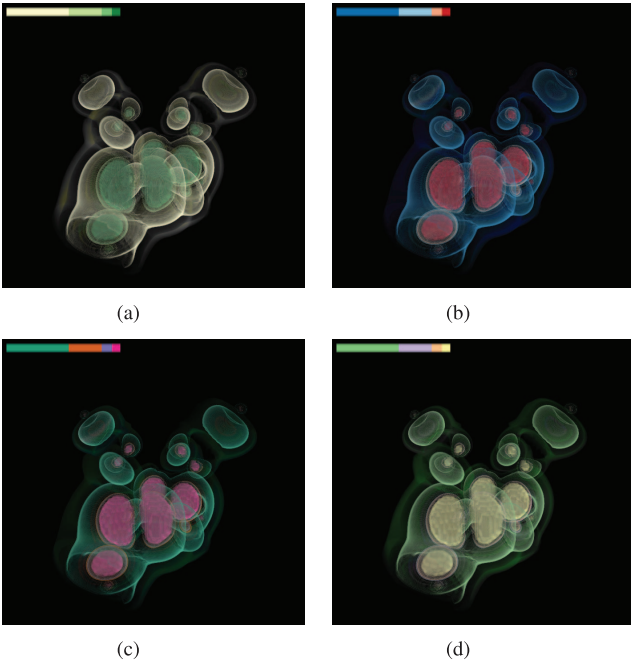
Fig. 13. Images from the *neghip* data set rendered using various color schemes provided by the Color Brewer [33]. (a) Uses the Yellow-Green sequential color scheme. (b) Uses the Red-Yellow-Blue diverging color scheme. (c) and (d) Use the Dark2 and the Accent qualitative color schemes, respectively.

image obtained from the slice view is filtered using a median filter or a bilateral filter. The median filter has been implemented based on the HDR Image Processing Library [41] and designed for the CUDA kernel. In our implementation, a 3×3 median filter is applied to the slice view image by default. However, users are given the option to change both the filter and its parameters.

5 RESULTS AND DISCUSSION

During the design and evaluation process of our work, we have collaborated with computational nanotechnology and flow simulation researchers. In fact, the need for a simplified interface for volume rendering on nanohub.org was the impetus for this work. Based on our initial success with computational nanotechnology data, we have applied

the pipeline to explore a variety of data sets including ratio and interval data sets (both with and without noise). In this section, we present a gallery of results highlighting various aspects of our data exploration pipeline and its ability to extract important features in various data sets as well as feedback from our various user constituents.

5.1 Nonnoisy Volumetric Data

Fig. 6a shows the results of applying the value profiling and data enhancement aspects of our pipeline to a *fuel simulation* data set. The value profile shows that important regions within the data are occluded by outer layers from a specific viewpoint. Our pipeline automatically extracts four layers based on the value profile and the peak value of each layer is enhanced. This enhancement, combined with the default diverging Color Brewer scheme, allows a user to easily identify important parts previously occluded within the data set. Based on the value profile, four layers are extracted and the peak values of each layer are enhanced. The number of colors is equal to the number of feature layers. Various Color Brewer schemes can be applied to a data set depending on its type. Fig. 13 shows different color schemes applied to the *neghip* data set ranging from a sequential color scheme Fig. 13a to a diverging scheme in Fig. 13b and qualitative color settings in Figs. 13c and 13d.

Automatic parameter specification in our pipeline is especially useful for scientists who are unfamiliar with volume rendering and visualization techniques. However, our pipeline also supports advanced visualization users by allowing them to tweak rendering parameters such as the color, opacity, and the number of colors.

Fig. 14 shows different rendering results using traditional 1D and 2D transfer function widgets and our widget. We worked with a nanoscientist who had previously been working with volumetric rendering tools for analyzing quantum dot simulations. This scientist was familiar with 1D and 2D transfer functions, and using these traditional tools, the user was able to visualize and highlight structures within the data set. In the 1D transfer function space, Fig. 14a, the scientist was able to explore the scalar data values and create a suitable transfer function; however, he reported that the automatic parameter settings provided a quicker and better rendering in terms of his

(a)   (b)   (c)   (d)
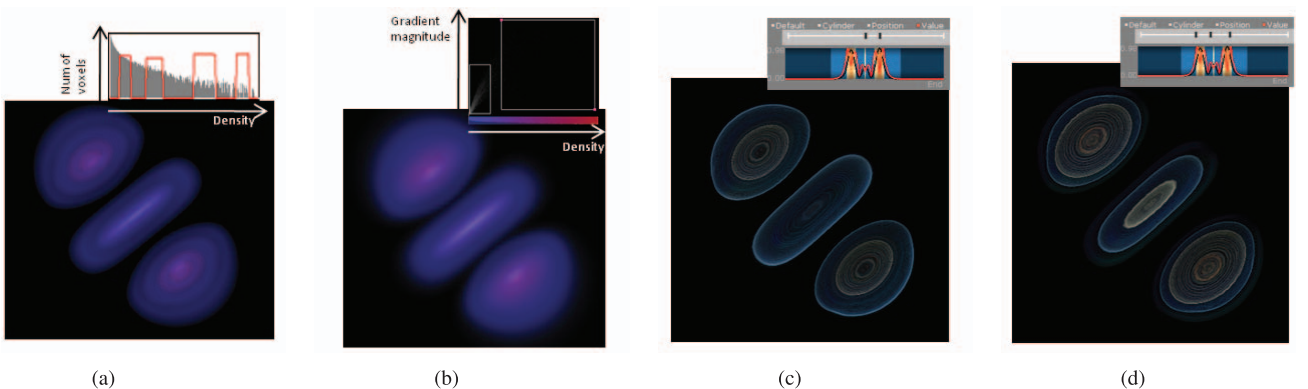
Fig. 14. Plots of a wave function using a 1D histogram widget, a 2D histogram widget, and our data exploration widget. (a) The result using the 1D histogram. (b) The result using the 2D histogram. Most of the values are binned within the left small box, but the area does not have any inner structure within it. (c) The result with the local maximum highlighted. (d) The result with the local maximum gradient magnitude highlighted.
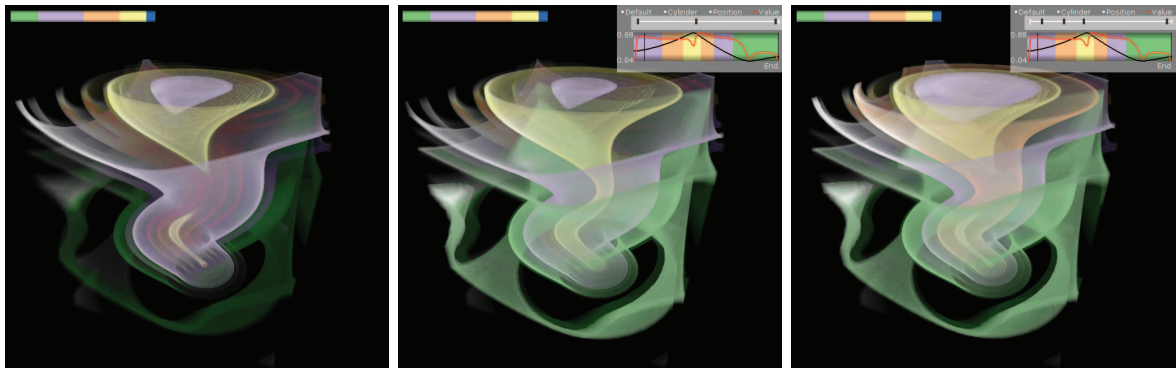
Fig. 15. Results of a *tornado* data set rendered by our data exploration method. The left image is a result of enhancing local peaks in the region of interest, the middle image shows a result from enhancing local maximum gradient magnitudes, and the right shows a resulting image from user-defined data enhancements.

analysis. In the 2D transfer function case, Fig. 14b, since the important structure within the simulation data is sparse, the upper right portion of the 2D transfer function histogram appears to contain no data; however, this region needs to be selected to create an appropriate visualization. In this case, the user had difficulty in determining the correct place to search for data and utilized a hunt and peck approach until finally settling on an image with inner structures of the data blurred out. By using our method, Figs. 14c and 14d, the scientist reported that he was able to reduce his search time and gain insights about the changes in wave functions within his data. According to the user, Fig. 14d is better than both Figs. 14a and 14b as the user was able to better highlight the inner structure of the data and gain an understanding of the electron potential fields within the data.

Figs. 14c and 14d show the difference between emphasizing local scalar maxima and local gradient magnitude maxima, respectively. In comparing Fig. 14c to 14d, the user found both images to be useful. Fig. 14c provided the scientist with a better understanding of the electron potential clouds, while Fig. 14d provided more details about the inner structures of the data.

Informal feedback from computational nanotechnology researchers and computational flow researchers has been very positive. These scientists find this system interface more effective and intuitive for exploring, analyzing, and understanding the features in their simulation data than existing interfaces. In terms of the usefulness, this tool helped them to better understand the distribution of the wave functions within the quantum dot area. From the line plot (value profile) that cuts through the center of their simulation data (e.g., quantum dot), they could see their simulation result (e.g., the wave function data) along critical directions in real space and directly highlight features of interest. In addition, they told us that this helped them in performing quantitative analysis which was absent in previously available visualization tools that only provide qualitative analysis. Previously, the scientist had developed scripting programs (e.g., MATLAB programs) to perform quantitative analysis by generating line plots in critical directions. Finally, our system also helped them calculate optical matrix elements in various directions for their simulation. While obtaining this informal feedback, we also

provided the end user with commonly used viewing directions (e.g., top, front, and $+-45$ rotation views) to help better extract the value profile. We found that the addition of these viewing angles were also a very popular feature and were able to further reduce the amount of time needed to analyze and visualize a data set.

The usefulness (to the scientist) was in the reduced amount of exploration time needed to generate the image (note that without the semiautomatic approach the transfer function needs 4 peaks (Fig. 14a). The addition of the ray profile tool provided them with a means to perform quantitative analysis that was previously ported to other software tools. Thus, the addition of the tools for the semiautomatic generation and quantitative analysis are able to both reduce the burden of transfer function creation on the user while enhancing their overall analytic capability.

As such, our data exploration method more closely couples the physics governing the data and techniques for mapping this data using a judiciously chosen transfer function. Allowing the tighter integration of the data with the transfer function provides a more intuitive interface for manipulating parameters present in the data itself and is one of the key strengths of the method. Because users are allowed to interactively manipulate the data and data gradients and see an illustration on the screen, they get a much more intuitive notion of how to best communicate and understand the physics under consideration.

5.2   Time-Varying Data

We further applied our methodology to the investigation of fluid flow. Fig. 15 shows a *tornado* data set and compares results obtained using the default and user-defined parameter settings. Fig. 16 shows the resultant images obtained by setting rendering parameters to extract interesting regions and track them in the *convection* data set. These images show a time advection of the flow in this data set. Informal feedback from a computational fluid dynamics expert also yielded positive results. This researcher indicated that by using our visualization techniques, the convection layers can clearly be identified. This allows the user to better understand the time-varying behavior of this complex data, allowing for better identification of salient structures. With respect to the convection in a box, Fig. 16, the user is able to clearly see the nested structure of the flow
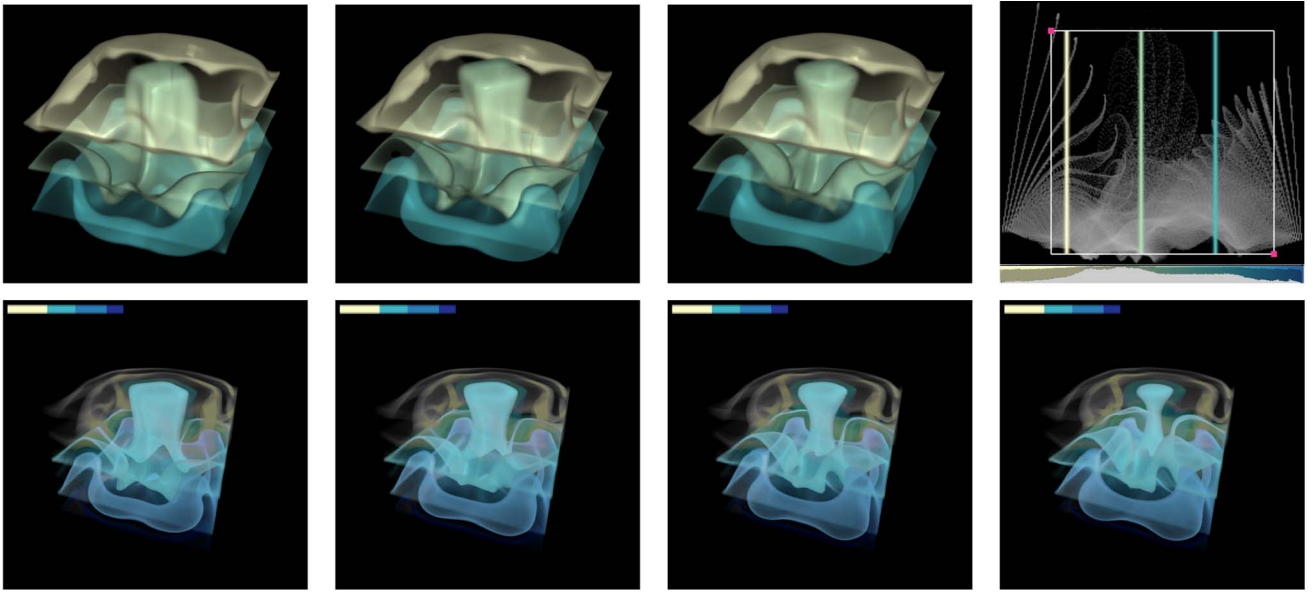
Fig. 16. Results of our data enhancement technique applied to visualize temporal advection in a convection simulation. The series of images (Top and Bottom) show every second sequential time step starting from 220. (Top) Using the sinusoidal transfer function method from Svakhine et al. [31]. (Bottom) Our semiautomatic transfer function generation procedure. Note that the layers in the bottom image visualize several different structures of the flow, particularly near the top portions of the image.

gradients. By allowing users to interactively investigate and pinpoint boundaries of interest, we can create visualizations that show global structures without occluding regions of interest that are limited to a more local region of interest.

We also applied our methods to the tornado data. Fig. 15 shows the nested structure of the flow and enables users to pinpoint the structure of the gradients. Visualizing flow in this manner is a powerful means for viewing these nested structures and how they change over time. Note that for all temporal data sets, a single time step was used to calculate the visual parameters. As such, a static transfer function is used for all time step renderings in order to keep coherency between the mapping of color and opacity to a particular scalar or gradient value. Depending on the time step chosen to generate the transfer function, occlusion of structures in future time steps may occur. By keeping the color mapping coherent (i.e., for all time steps, the same color maps to the same scalar value), as the data change, some structures may be lost. As such, the user could recolor the data based on the ray profile at any time step. Future work will focus on solutions to this issue of the tradeoff between color coherency and the occlusion of structures.

5.3 Noisy Data

While our previous examples showed smooth data sets, noisy data sets require further user intervention to generate useful results. Other parameters, such as the median window size, can be modified by the user to interactively select an optimal size that generates the best results. This is demonstrated in Fig. 10 where a noisy data set is rendered without any filtering in Fig. 10a. Application of the 3×3, 5×5, and 9×9 size median filter yields better results as shown in Figs. 10b to 10d. However, as we increase the window size, the rendered image starts losing important feature lines. Therefore, a default size of 3×3 is adopted in our implementation.

Fig. 17 shows another example with feature boundaries enhanced by our default rendering parameters. While these initial results are good at showing overall feature boundaries, in some cases, the default rendering parameters will not produce a clear visualization. This is illustrated in Fig. 18a. Here we show a rendering example using noisy data where the default rendering parameters are suboptimal. In this case,
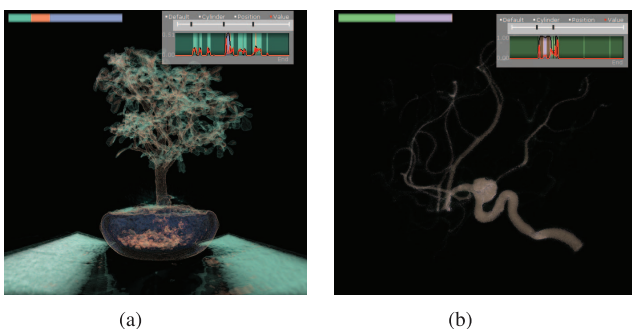
Fig. 17. Images of (a) *bonsai* data set. (b) *aneurism* data set rendered with default settings showing enhanced boundary areas.
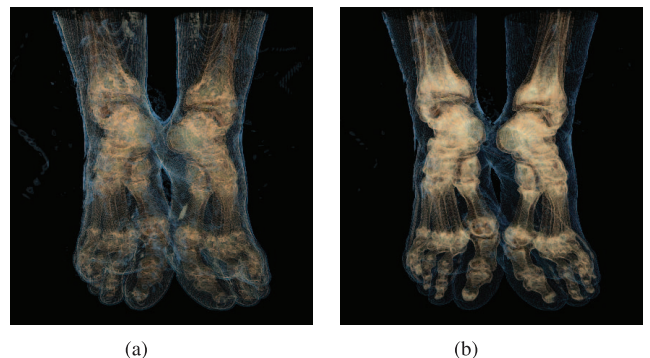
Fig. 18. Images of the *feet* data set rendered (a) with the default system-generated settings, and (b) after the user adjusts the emphasis points.

the user may edit the transfer function to obtain better results, as seen in Fig. 18b.

We explored both Median and Gaussian filters using guided exploration with expert users. Users expressed that they could tell no difference between the applied filters and were likely to simply use the default values in the cases where filtering is needed.

6 CONCLUSION AND FUTURE WORK

In this paper, we have presented a data exploration method for volumetric rendering using value profiles and contour lines. We devised methods to automatically provide default rendering parameter settings based on the features in the region of interest. This technique is applicable to interval and ratio data types with or without noise to support a variety of volumetric data sets, such as medical data and scientific simulation data. Based on the data types, we extract feature layers from the data based on a value profile and provide contour lines from the obtained feature layers. The extracted contour lines help to illustrate the shape of the data set in the region of interest.

To support data sets with noise, we denoise the value profile and the image captured on a slice view by filtering them with a feature preserving nonlinear filter (median or bilateral filter). We also modulate the opacities based on the feature layers to emphasize boundary features and maximum intensities in the region of interest. Moreover, we parameterized a weight map, the radius of an ROI, and a focus line.

We plan to include the other two data types (nominal and ordinal) [42] and add further automation to the system by providing users with an optimal viewpoint that can show the important data features by default [43]. Further, we also plan to extend our data exploration approach to support multidimensional transfer functions. Since a slice view has more sampled voxels, 2D ROI specification based on slices will be useful to generate a local feature-driven 2D transfer function using nonparametric clustering algorithms (e.g., kernel density estimation) [28]. Illustrative visualization techniques have been used to convey important features of data sets by abstracting away unnecessary details [18], [37]. We would like to incorporate some of these techniques to improve users' perception [44] of the extracted feature layers. Also, since scientific simulations typically return multivariate data sets, we plan to extend our technique to look at the ray profile across several variables at once. This will generate a variety of layers, and exploration will be done on ways to merge layers for enhanced visualization.

REFERENCES

[1] C. Catlett, W.E. Allcock, P. Andrews, and R. Aydt, "TeraGrid: Analysis of Organization, System Architecture, and Middleware Enabling New Types of Applications," *Proc. HPC and Grids in Action,* 2007.

[2] The Open Science Grid Consortium, http://www.openscience grid.org/, 22 Apr., 2011.

[3] nanoHUB.org, http://www.nanohub.org/ 22 Apr., 2011.

[4] G. Klimeck, S.S. Ahmed, N. Kharche, M. Korkusinski, M. Usman, M. Prada, and T. Boykin, "Atomistic Simulation of Realistically Sized Nanodevices Using Nemo 3D Part II: Applications," http://nanohub.org/resources/3825, 22 Apr., 2011, Jan. 2008.

[5] J. Kniss, G. Kindlmann, and C. Hansen, "Multidimensional Transfer Functions for Interactive Volume Rendering," *IEEE Trans. Visualization and Computer Graphics,* vol. 8, no. 3, pp. 270-285, July-Sept. 2002.

[6] NVIDIA Corporation, "NVIDIA CUDA Compute Unified Device Architecture," http://www.nvidia.com/object/cuda_home_new.html, 22 Apr., 2011, Nov. 2007.

[7] J. Kniss, G. Kindlmann, and C. Hansen, "Interactive Volume Rendering Using Multi-Dimensional Transfer Functions and Direct Manipulation Widgets," *Proc. Conf. Visualization,* pp. 255-562, 2001.

[8] J. Marks, B. Andalman, P.A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber, "Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation," *Computer Graphics,* vol. 31, pp. 389-400, 1997.

[9] M. Tory, S. Potts, and T. Möller, "A Parallel Coordinates Style Interface for Exploratory Volume Visualization," *IEEE Trans. Visualization and Computer Graphics,* vol. 11, no. 1, pp. 71-80, Jan./Feb. 2005.

[10] C. Rezk-Salama, M. Keller, and P. Kohlmann, "High-level User Interfaces for Transfer Function Design with Semantics," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 1021-1028, Sept./Oct. 2006.

[11] Y. Wu and H. Qu, "Interactive Transfer Function Design Based on Editing Direct Volume Rendered Images," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 5, pp. 1027-1040, Sept./Oct. 2007.

[12] S. Bruckner and M.E. Gröller, "Style Transfer Functions for Illustrative Volume Rendering," *Computer Graphics Forum,* vol. 26, no. 3, pp. 715-724, Sept. 2007.

[13] C.L. Bajaj, V. Pascucci, and D.R. Schikore, "The Contour Spectrum," *Proc. Conf. Visualization,* pp. 167-173, 1997.

[14] C. Lundström, A. Ynnerman, P. Ljung, A. Persson, and H. Knutsson, "The Alpha-Histogram: Using Spatial Coherence to Enhance Histograms and Transfer Function Design," *Proc. Eurographics/IEEE-VGTC Symp. Visualization,* pp. 227-234, 2006.

[15] C. Correa and K.-L. Ma, "Size-Based Transfer Functions: A New Volume Exploration Technique," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1380-1387, Nov./Dec. 2008.

[16] L. Wang, Y. Zhao, K. Mueller, and A. Kaufman, "The Magic Volume Lens: An Interactive Focus+Context Technique for Volume Rendering," *Proc. IEEE Conf. Visualization,* pp. 367-374, 2005.

[17] A. Lu, R. Maciejewski, and D.S. Ebert, "Volume Composition Using Eye Tracking Data," *Proc. Eurographics/IEEE-VGTC Symp. Visualization,* pp. 115-122, 2006.

[18] W. Li, L. Ritter, M. Agrawala, B. Curless, and D. Salesin, "Interactive Cutaway Illustrations of Complex 3D Models," *ACM Trans. Graphics,* vol. 26, no. 3, pp. 31-40, 2007.

[19] I. Viola, A. Kanitsar, and M.E. Gröller, "Importance-Driven Feature Enhancement in Volume Visualization," *IEEE Trans. Visualization and Computer Graphics,* vol. 11, no. 4, pp. 408-418, July-Aug. 2005.

[20] T. Ropinski, J.-S. Praßni, F. Steinicke, and K.H. Hinrichs, "Stroke-Based Transfer Function Design," *Proc. IEEE/EG Int'l Symp. Volume and Point-Based Graphics,* pp. 41-48, 2008.

[21] C. Rezk-Salama and A. Kolb, "Opacity Peeling for Direct Volume Rendering," *Computer Graphics Forum,* vol. 25, no. 3, pp. 597-606, 2006.

[22] M.M. Malik, T. Möller, and M.E. Gröller, "Feature Peeling," *Proc. Graphics Interface,* pp. 273-280, 2007.

[23] C.D. Correa and K.-L. Ma, "Visibility-Driven Transfer Functions," *Proc. IEEE Pacific Visualization Symp.,* pp. 177-184, 2009.

[24] P. Kohlmann, S. Bruckner, A. Kanitsar, and M.E. Gröller, "Contextual Picking of Volumetric Structures," *Proc. IEEE Pacific Visualization Symp.,* pp. 185-192, 2009.

[25] H. Akiba and K.-L. Ma, "A Tri-Space Visualization Interface for Analyzing Time-Varying Multivariate Volume Data," *Proc. Eurographics/IEEE-VGTC Symp. Visualization,* pp. 115-122, 2007.

[26] H. Akiba, K.-L. Ma, J.H. Chen, and E.R. Hawkes, "Visualizing Multivariate Volume Data from Turbulent Combustion Simulations," *Computing in Science and Eng.,* vol. 9, no. 2, pp. 76-83, Mar.-Apr. 2007.

[27] C. Muelder and K.-L. Ma, "Interactive Feature Extraction and Tracking by Utilizing Region Coherency," *Proc. IEEE-VGTC Pacific Visualization Symp.,* pp. 17-24, Apr. 2009.

[28] R. Maciejewski, I. Woo, W. Chen, and D. Ebert, "Structuring Feature Space: A Non-Parametric Method for Volumetric Transfer Function Generation," *IEEE Trans. Visualization and Computer Graphics,* vol. 15, no. 6, pp. 1473-1480, Nov./Dec. 2009.

[29] C. Ware, *Information Visualization: Perception for Design.* Morgan Kaufmann, 2004.

[30] S.S. Stevens, "On the Theory of Scales of Measurement," *Science,* vol. 103, no. 2684, pp. 677-680, 1946.

[31] N. Svakhine, Y. Jang, D. Ebert, and K. Gaither, "Illustration and Photography Inspired Visualization of Flows and Volumes," *Proc. IEEE Conf. Visualization,* pp. 687-694, 2005.

[32] J. Krüger, J. Schneider, and R. Westermann, "ClearView: An Interactive Context Preserving Hotspot Visualization Technique," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 941-948, Sep./Oct. 2006.

[33] M. Harrower and C.A. Brewer, "Colorbrewer.org: An Online Tool for Selecting Colour Schemes for Maps," *Cartographic J.,* vol. 40, no. 1, pp. 27-37, June 2003.

[34] I. Viola, A. Kanitsar, and M.E. Gröller, "Hardware-based Nonlinear Filtering and Segmentation Using High-level Shading Languages," *Proc. IEEE Conf. Visualization,* pp. 309-316, 2003.

[35] G. Kindlmann and J.W. Durkin, "Semi-Automatic Generation of Transfer Functions for Direct Volume Rendering," *Proc. IEEE Symp. Volume Visualization (VVS '98),* pp. 79-86, 1998.

[36] J. Zhou, M. Hinz, and K.D. Tönnies, "Focal Region-guided Feature-Based Volume Rendering," *Proc. First Int'l Symp. 3D Data Processing, Visualization, and Transmission,* pp. 87-90, 2002.

[37] D. Ebert and P. Rheingans, "Volume Illustration: Nonphotorealistic Rendering of Volume Models," *Proc. IEEE Conf. Visualization,* pp. 195-202, 2000.

[38] W.E. Lorensen and H.E. Cline, "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," *Proc. Conf. Computer Graphics and Interactive Techniques,* pp. 163-169, 1987.

[39] D. Horn, "Stream Reduction Operations for GPGPU Applications," *Proc. GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation,* pp. 573-583, 2007.

[40] M. Harris, S. Sengupta, and J.D. Owens, "Parallel Prefix Sum (scan) with CUDA," *GPU Gems 3,* H. Nguyen, ed. Addison Wesley, pp. 851-876, 2007.

[41] HDR Image Processing Library, http://courses.ece.uiuc.edu/ece498, 22 Apr., 2011.

[42] L.D. Bergman, B.E. Rogowitz, and L.A. Treinish, "A Rule-Based Tool for Assisting Colormap Selection," *Proc. Conf. Visualization,* pp. 118-125, 1995.

[43] S. Takahashi and Y. Takeshima, "A Feature-Driven Approach to Locating Optimal Viewpoints for Volume Visualization," *Proc. IEEE Conf. Visualization,* pp. 495-502, 2005.

[44] M.-Y. Chan, Y. Wu, W.-H. Mak, W. Chen, and H. Qu, "Perception-Based Transparency Optimization for Direct Volume Rendering," *IEEE Trans. Visualization and Computer Graphics,* vol. 15, no. 6, pp. 1283-1290, Nov./Dec. 2009.

Insoo Woo received the BS degree in computer engineering in 1998 from Dong-A University in South Korea. He is currently working toward the PhD degree in the School of Electrical and Computer Engineering at Purdue University and a research assistant in the Purdue University Rendering and Perception Lab. He worked as a software engineer from 1997 to 2006. His research interests include GPU-aided Techniques for Computer Graphics and Visualization. He is a member of the IEEE.

Ross Maciejewski received the PhD degree in electrical and computer engineering from Purdue University in December, 2009. He is currently an assistant professor at Arizona State University in the School of Computing, Informatics & Decision Systems Engineering. Prior to this, he served as a visiting assistant professor at Purdue University and worked at the Department of Homeland Security Center of Excellence for Command Control and Interoperability in the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) group. His research interests include geovisualization, visual analytics, and nonphotorealistic rendering. He is a member of the IEEE and the IEEE Computer Society.

Kelly P. Gaither received the masters and bachelors degree in computer science from Texas A&M University in 1992 and 1988, respectively, and the doctoral degree in computational engineering from Mississippi State University in May, 2000. While working toward the PhD degree, she worked full time at the Simulation and Design Center in the National Science Foundation Engineering Research Center as the leader of the visualization group. She is the director of Data & Information Analysis at the Texas Advanced Computing Center (TACC), is leading the scientific visualization, data management & collections, and data mining & statistics programs at TACC while conducting research in scientific visualization and data analysis. She is a research scientist, also serves as the area director for visualization in the National Science Foundation funded TeraGrid project. She has a number of refereed publications in fields ranging from Computational Mechanics to Supercomputing Applications to Scientific Visualization. She has given a number of invited talks. Over the past 10 years, she has actively participated in the IEEE Visualization conference and conducting the IEEE Visualization conference general chair in 2004. She is currently serving on the IEEE Visualization and Graphics Technical Committee. She is a member of the IEEE.

David S. Ebert received the PhD degree in computer science from Ohio State University. He is the silicon valley professor in the School of Electrical and Computer Engineering at Purdue University, a University Faculty Scholar, the director of the Purdue University Rendering and Perceptualization Lab, and the director of the Visual Analytics for Command, Control and Interoperability Environments Center of Excellence. His research interests include novel visualization techniques, visual analytics, volume rendering, information visualization, perceptually based visualization, illustrative visualization, and procedural abstraction of complex, massive data. He is a fellow of the IEEE and the IEEE Comptuer Society, and a member of the IEEE Computer Society's Board of Governors.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

Guest Editorial: Special Issue on Visualization and Visual Analytics

Aidong Lu*, David Ebert, Jinzhu Gao, Song Zhang, Alark Joshi

This special issue is devoted to the new research addressing challenges in the areas of visualization and visual analytics. Visualization and visual analytics are closely related research areas, both concentrating on developing visual techniques to reveal meaningful information out of various data in real-life applications. Visualization as a field has its roots in Computer Graphics and has become a popular research area over the years. The field of visual analytics is relatively young with a concentration on analytical reasoning facilitated by interactive visual interfaces. In general, visualization and visual analytics research is tightly connected with certain types of data or applications and researchers in both fields strive to discover known or unknown data patterns for domain users.

Since Tsinghua Science and Technology adjusted the scope to information technology in 2011, this is the second special issue on topics of Visualization and Visual Analytics. All the articles published in this special issue are selected through the open call. All the submissions have gone through the blind peer-reviewed process. We finally accepted eight research articles, five from the United States and three from China. The articles cover various topics on visualization and visual analytics. Applications include

---

● Aidong Lu is with the Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA. E-mail: aidonglu@gmail.com.

● David Ebert is with School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA.

● Jinzhu Gao is with the Department of Computer Science, University of the Pacific, Stockton, CA 95211, USA. E-mail: jgao@pacific.edu.

● Song Zhang is with the Department of Computer Science and Engineering, Mississippi State University, Starkville, MS 39762, USA.

● Alark Joshi is with the Department of Computer Science, Boise State University, Boise, ID 83725, USA.

∗ To whom correspondence should be addressed.

Manuscript received :2013-03-24; accepted : 2013-03-24

geo-spatial visualization, flow visualization, molecule visualization, online log visualization, microblogging visualization, image transition, edge-bundling, and insight management.

In the following, we roughly divide the eight accepted articles to three categories: scientific visualization, information visualization, and visual analytics. The readers may find that almost all the visualization articles involve components of data visualization, interactive exploration, and analytical reasoning components. The concentrations of each approach and the details related to visualization applications are different.

A brief overview of the accepted articles is given below.

Scientific Visualization

Climate research produces a wealth of multivariate data. In "An Interactive Visual Analytics Framework for Multi-Field Data in a Geo-Spatial Context", Zhang et al. present a framework for studying multi-field climate data. Several visualization and interaction techniques, such as fixed-window brushing and correlation-enhanced display, are presented and integrated with the Google Earth platform. The system has been tested by a team of climate researchers, who made a few important discoveries using it.

Turbulent flows are intrinsic to many processes in science and engineering, however the complex, non-linear interactions between individual eddies in these flows are hard to identify and quantify across multiple scales. In "Methods to Identify Individual Eddy Structures in Turbulent Flow", Wang et al. present several novel approaches for accurately segmenting individual eddy structures in turbulent flows. These methods can help quantify information of a flow at the level of individual structures and automatically track the evolution and interaction of large numbers of individual vortices in a complicated turbulent flow.

Atmospheric nucleation serves as a significantly

important role in many atmospheric and technological processes. In "Similarity-Based 3-D Atmospheric Nucleation Data Visualization and Analysis", Zhu et al. present a data visualization solution with a novel algorithm for calculating similarity between the 3-D molecular crystals to visualize and classify 3-D molecular crystals effectively. The overall performance of the visualization system has been further improved with GPU acceleration.

Information Visualization

Real-time data log visualization is challenge due to the data complexity: it is streaming, hierarchical, heterogeneous, and multi-sourced. In "An Online Visualization System for Streaming Log Data of Computing Clusters", Xia et al. present a two-stage streaming process to visualize the log data generated by computing clusters. The visualization supported by a visual computing processor consists of a set of multivariate and time variant visualization techniques. The effectiveness and scalability of the proposed system framework are demonstrated on a commodity cloud-computing platform.

Microblogging, similar to Twitter, has provided a popular communication scheme for Web users to share information and express opinions. In "Portraying User Life Status from Microblogging Posts", Tang et al. presented an interactive visualization, LifeCircle, to explore behaviors of microblog users. The approach tightly integrates interactive visualization with novel and state-of-the-art microblogging analytics. Data from Sina Weibo has been used in the case studies and the results demonstrate the approach provides a quick summary of user life status for potential personal users and commercial services.

Transition in many information visualization applications can help users perceive changes and understand the underlying data. In "A Study of Animated Transition in Similarity-Based Tiled Image Layout", Zhang et al. investigate the effectiveness of animated transition in a tiled image layout with spiral arrangement. Based on three aspects of animated transition, an integrated solution, called AniMap, is

presented for animating the transition between layouts during query processes. The effectiveness of the animated transition solution has been demonstrated by experimental results and a comparative user study.

Edge bundling has been a popular approach as edge is an important visual primitive for encoding data in information visualization research. In "Edge Bundling in Information Visualization", Zhou et al. first provide a survey on edge-bundling techniques for reducing visual clutter problem in visualization. They have reviewed the cost-based, geometry-based, and image-based edge-bundling methods designed for graphs, parallel coordinates, and flow maps. They also describe a number of visualization applications using edge-bundling techniques, discuss the evaluation studies on the effectiveness of edge-bundling methods, and point out some future research directions.

Visual Analytics

Significant progress has been made toward effective insights discovery in visual analytics systems, while managing large amounts of insights generated in visual analytics processes is also important. In "ManyInsights: A Visual Analytics Approach to Supporting Effective Insight Management", Chen and Yang present a multi-dimensional visual analytics prototype, ManyInsights, that integrates several insight management approaches, including insight annotation, browsing, retrieval, organization, and association. This paper also reports a longitudinal case study that has evaluated ManyInsights with a domain expert, realistic analytic tasks, and real datasets.

Acknowledgments

How Visualization Courses Have Changed over the Past 10 Years

G. Scott Owen
Georgia State University

Gitta Domik
University of Paderborn

David S. Ebert
Purdue University

Jörn Kohlhammer
Fraunhofer IGD Darmstadt

Holly Rushmeier
Yale University

Beatriz Sousa Santos
University of Aveiro

Daniel Weiskopf
University of Stuttgart

The past 10 years have seen profound changes in visualization algorithms, techniques, methodologies, and applications. For example, we're seeing

- extensive use of GPUs,
- improved algorithms for flow or volume visualization,
- emphasis on highly interactive visual interfaces,
- the advent and increasing importance of visual analytics,
- an increase in nontechnical students in our courses,
- greater need for professional use of visualization in the workplace, and
- evaluation frameworks for effective visualization.

All this forces alterations to our visualization courses, especially what, how, or whom we teach. A basic problem has always been that we couldn't rely on standard textbooks to frame the mandatory knowledge in this field. This situation is unlike that of computer graphics, in which the community widely acknowledges several standard textbooks. Visualization curricula suggestions—for example, ACM Siggraph's Education Committee recommendations (www.upb.de/cs/vis) or the Visual Analytics Digital Library (http://vadl.cc.gatech.edu)—are partly outdated or incomplete. Computer science curricula guidelines, such as

from the IEEE and ACM, also lag in their recommendations of content for this novel, dynamic knowledge area.

Outdated course content recommendations, together with profound changes in the underlying technology and methodology, produce an unstable ground for educators at a time when visual representations have gained great importance in economics, science, and many other areas of society.

To address this issue, under the auspices of the ACM Siggraph Education Committee, we held meetings or workshops at Siggraph 2011 and 2012 and a panel and workshop at Eurographics 2012. At the panel, called "The Changes We Have Made to our Visualization Courses over the Last 10 Years," Holly Rushmeier, Jörn Kohlhammer, David Ebert, Beatriz Sousa Santos, and Daniel Weiskopf discussed how they've changed their courses to reflect current problems and practical solutions.[1] (Slides are at www.upb.de/cs/vis.) Each panelist has many years' experience teaching courses covering topics such as scientific visualization, data visualization, information visualization, visualization techniques, and visual analytics.

Here, we examine the insights gathered at the panel, workshops, and meetings.

Visualization in a Liberal Education

Rushmeier teaches the course Visualization: Data, Pixels, and Ideas in the context of Yale's liberal

arts education. Liberal education is, according to the Association of American Colleges and Universities, "a philosophy of education that empowers individuals with broad knowledge and transferable skills, and a strong sense of value, ethics, and civic engagement."[2] Consequently, Rushmeier has a mixed audience with technical and nontechnical backgrounds for which she must design meaningful course content and assignments. Because of her students' mixed background, her course requires no programming or advanced mathematics.

The students' goals are to

- understand visualization's basic components,
- understand computer graphics tools for producing visualizations,
- recognize bad visualizations, and
- use visualization effectively in discovery and communication.

The course covers

- spatial visualization and projections (3D to 2D);
- motion;
- interaction;
- communication best practices—for example, Edward Tufte's principles; and
- scientific and information visualization.

(The courses of the other educators we'll be looking at also include many or most of these topics.)

Class assignments focus on the students' future work life needs. This means Rushmeier must find engaging, manageable datasets on political, historical, or other issues in the humanities. It also means having the students experiment with tools that allow flexibility in designing visualizations without requiring programming skills. Because there's no programming, the course uses tools such as Excel, Matlab, and a VRML (Virtual Reality Modeling Language) viewer.

Geographic information systems are another tool the students learn to use. These systems are used increasingly in the humanities, including in subjects such as comparative literature (see Figure 1).

A Perspective between Research and Business Careers

Kohlhammer sees an increase in students' motivation to learn visualization aspects for their business careers. As in Rushmeier's course, assignments and projects in his Information Visualization and Visual Analytics course at Technische Universität Darmstadt focus on students' later work life by providing hands-on experience with
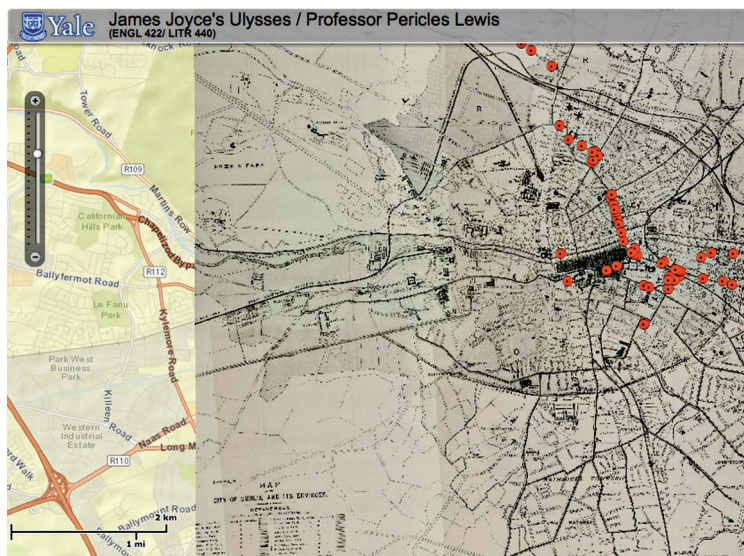
Figure 1. Students in Pericles Lewis's *Ulysses* seminar at Yale University mapped major events in the novel using the addresses in Google, cross-referenced with a map of Dublin in the time of James Joyce.

real-world datasets and data types. His courses have a mix of computer science, business informatics, and mathematics students, with some psychology and engineering students. They all have solid programming skills and an affinity for computer graphics.

His course has evolved to include more practical exercises and a strong connection to industry, dealing with areas such as business intelligence, finance (for example, risk analysis), and security. He encourages and supports student involvement in using real-world datasets in global competitions such as the VAST Challenge (Visual Analytics Science and Technology; http://vacommunity.org/VAST+Challenge+2013).

Figure 2 shows a Web-based visual search system for time-oriented research data that Kohlhammer's students developed.[3]

Teaching Visual Analytics: Leveraging Multidisciplinarity

Ebert and Elmqvist teach Introduction to Visual Analytics to students with diverse backgrounds, so the course requires no programming expertise.[4] The students are expected to have a knowledge of one or more of these areas: data analysis, knowledge management, statistics, computer graphics, or visualization. The course consists of group discussions of papers, lectures by the instructors (the course is team-taught), projects, and student presentations of papers.

The projects, which might be individual or group, are particularly important. At least five projects have resulted in conference submissions. Figure 3 shows an example project.
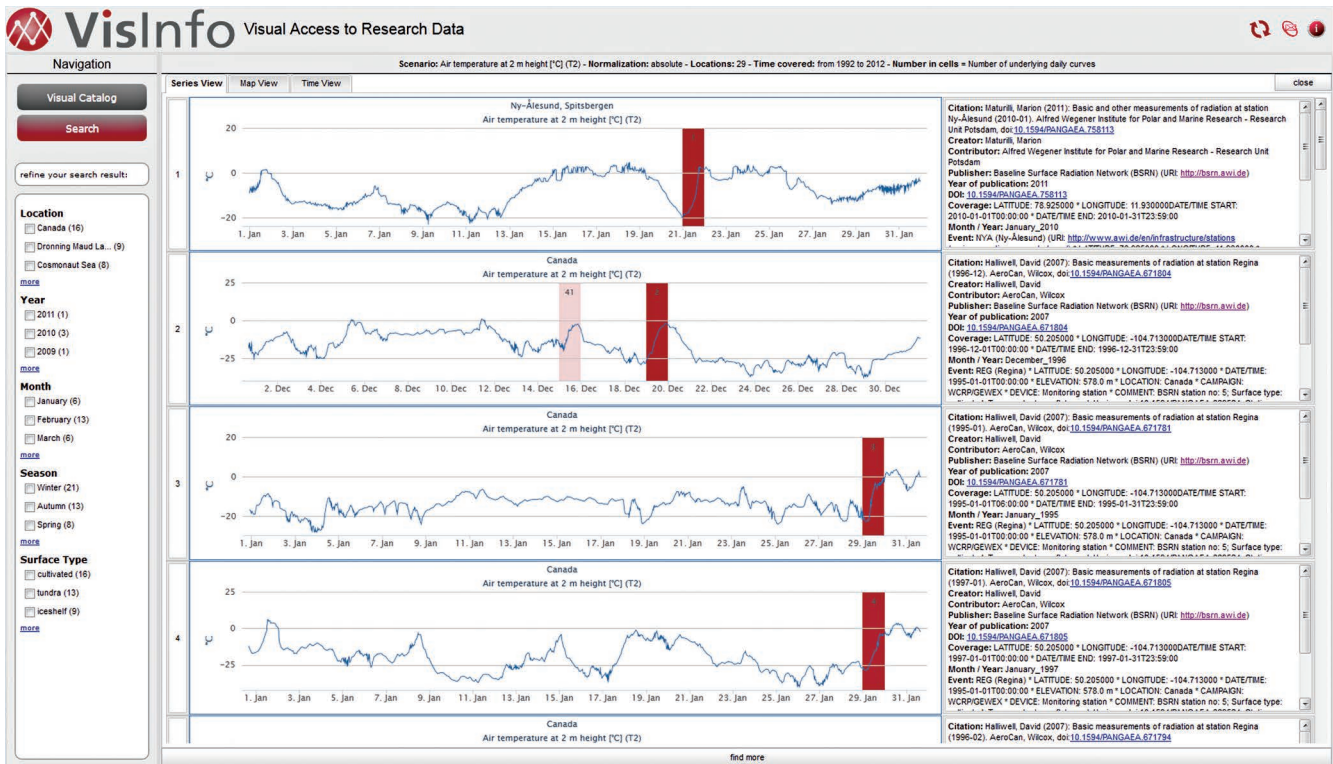
Figure 2. Jörn Kohlhammer's students at Technische Universität Darmstadt developed a Web-based visual search system for time-oriented research data. First, they created a visual catalog of daily temperature patterns based on the self-organizing-maps algorithm (not shown). Then, they presented a results list of documents based on a detailed selection. Metadata facets allowed interactive drill-downs in the results set.

This course, the only one in this article that focuses on visual analytics, has become a model for courses at many universities. Ebert finds a challenge in the fact that field of visualization has become too broad to cover in 15 weeks. So, the university complements this course with other courses—for example, Visualization Techniques, which covers in detail such topics as volume and flow visualization.

Changes in Beatriz Sousa Santos's Visualization Courses

Over the past decade, Sousa Santos has added more material on human characteristics (beyond visual perception), distributed and collaborative visualization, and displays. She also now teaches information visualization courses. Her students read more research papers and perform more evaluation experiments. Her courses include both undergraduate and graduate students with backgrounds in computer science, engineering, and management information systems. They do practical assignments using the Visualization Toolkit.[5] Sousa Santos and her colleagues have also integrated user studies into their courses.[6]

Teaching Visualization at the University of Stuttgart

The University of Stuttgart, where Weiskopf teaches, offers a variety of courses in computer graphics, geometric modeling, image synthesis, and visualization. In particular, there are courses on scientific visualization and information visualization. Additionally, the university uses a two-semester visualization-centered project to teach software engineering.[7] The university also offers the outreach course Introduction to Visualization in Science and Engineering. Students in that course have limited programming experience, so the course is heavily tool based.

The program aims to generate a common basis for computer graphics, visualization, and computer vision and to complement computer science students' typical mathematical and theoretical education. A challenge is the growing need for background knowledge from diverse fields such as mathematics, computer science, human-computer interaction, psychology, data mining, machine learning, and application-specific domains.

The Emerging Areas

From the panel, workshops, and meetings, three distinct areas emerged:

- *scientific or data visualization*, in which the data dimensions usually coincide with physical di-

mensions, such as in medical or remote-sensing scalar or flow data;

- *information visualization*, with typically multidimensional data, such as in finance, business intelligence, or large databases; and
- *visual analytics*, with massive, multisource, multiscale, heterogeneous, and streaming data.

Data preprocessing (for example, filtering, normalizing, and linguistic analysis) and subsequent visual presentations (for example, line graphs, line-integral-convolution images, and cone trees), which both depend on data syntax and semantics, might be different for these areas but also overlap considerably.

All three areas share some learning objectives. Students should be able to

- understand visualization techniques;
- recognize good versus misleading visualizations;
- select appropriate visualization techniques and visual attributes on the basis of the data and task;
- explain selected algorithms underlying visualization techniques for, for example, 2D data, 3D scalar or vector data, time-dependent data, multivariate data, hierarchically structured data, graphs and networks, or data with other structures;
- discuss the handling of unstructured data;
- understand the appropriate manipulation of data before mapping (which differs between data visualization, information visualization, and visual analytics);
- understand the limitations and capacity of human information processing;
- group and describe visualization techniques by some order (for example, by domain, data characteristics, or tasks);
- discuss how scaling (of the data or display) influences visualization techniques; and
- understand the theory and application of evaluation techniques to prove a visualization or interaction technique's success.

Although the three areas have distinct data domains, courses in them must cover the following themes.

The User
This theme includes human information-processing limitations and capabilities as well as an understanding of the tasks users bring to visualization problems.

The Design Stage
This stage describes a careful mapping of data
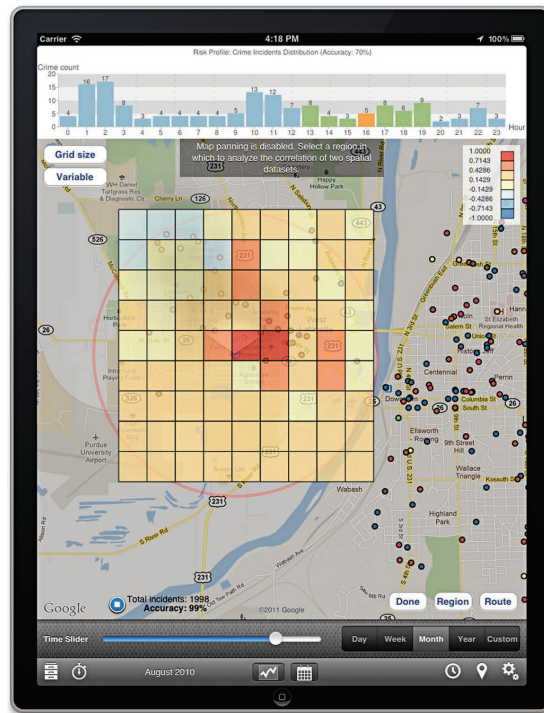
Figure 3. This project, by Ahmad Razip, a student in Niklas Elmqvist and David Ebert's Introduction to Visual Analytics course at Purdue University, correlated bus stops and crime incidence distribution.

components to visual attributes and the interactivity between users and data as well as between users and visualizations.

Visual Presentation
This includes a wealth of visualization solutions, sorted by data characteristics, application domain, or task and described by their various parameters. Instructors can present this theme at the breadth level by showing and discussing (interactive) visualizations. They can provide breadth-level training by using the available tools, in-depth training by developing interactive visualization techniques on a GPU, or training at any stage in between, depending on the students' qualifications.

Interaction Techniques
Interaction techniques are a requirement for visual analytics. They're also becoming increasingly necessary for data and information visualization, in which GPU techniques can reach the necessary processing speed.

Communication
Visual analytics in particular has stressed production, presentation, and dissemination as part of the visualization process. However, these topics are also important for data visualization and information visualization.

Collaboration

Interactivity aids collaboration, especially in synchronous and local situations. However, collaboration among stakeholders can also involve aspects that are asynchronous and distributed, such as Web-based collaboration technologies.

Evaluation

Evaluation is continuous. It starts with requirements analysis of the visualization problem. It continues with the human-in-the-loop's constant awareness of the software processes proceeding toward the visualization goal. It ends with evaluation to ensure reaching the goal for the specific visualization problem.

> One challenge is that more nontechnical students are interested in the courses because of visualization's increasing use in business and industry.

Displays

The variety of different displays' capabilities (size and spatial and tonal resolution) poses problems for visualization techniques, interactivity, and communication. These capabilities must be addressed at least at the mapping or design stage.

Challenges Identified

One challenge is that more nontechnical students are interested in the courses because of visualization's increasing use in business and industry. This necessitates the use of tools and real-world cases and datasets. If a course teaches both technical and nontechnical students, this potential difficulty could actually be an opportunity to approach a visualization problem from the viewpoints of multiple disciplines.[8] For computer science students, the courses should add newer developments such as shader programming and computer vision. One benefit of this is that some student projects might be worthy of conference publication, as has been the case with Elmqvist and Ebert's visual-analytics course.

Another challenge, as Ebert and Weiskopf stated in the panel discussion, is the need for instructors to update their own knowledge in diverse background fields ranging from math, to human-computer interaction and perception, to shader programming. So, this article's references include a few textbooks we use.[9–12]

At the panel, both Rushmeier and Ebert remarked that visualization has become too broad of a field to cover in one semester in suitable depth. So, instructors must decide between the breadth and depth of topics or offer one or more complementary visualization courses.

To help educators respond to the changes occurring in visualization courses, we've compiled a set of materials they can use to update their courses:

- the complete set of slides of the panelists and coauthors in this article,
- previously published articles by Ebert, Weiskopf, and Sousa Santos on their visualization courses (all from *IEEE CG&A*'s Education department), and
- related articles from *IEEE CG&A*'s Education department and other sources.

Links to these materials are at www.upb.de/cs/vis.

References

1. G. Domik et al., "Visualization Curriculum Panel—or the Changes We Have Made to Our Visualization Courses over the Last 10 Years," *Eurographics 2012—Education Papers*, 2012; www.cs.uni-paderborn.de/fileadmin/Informatik/AG-Domik/VisCurriculum/folien/eg2012-panel-Domik-2.pdf.
2. "Liberal Education," Assoc. of Am. Colleges and Universities, 2013; www.aacu.org/resources/liberaleducation/index.cfm.
3. J. Bernard et al., "Irina: A Visual Digital Library Approach for Time-Oriented Scientific Primary Data," *Int'l J. Digital Libraries*, vol. 11, no. 2, 2011, pp. 111–123.
4. N. Elmqvist and D.S. Ebert, "Leveraging Multidisciplinarity in a Visual Analytics Graduate Course," *IEEE Computer Graphics and Applications*, vol. 32, no. 3, 2012, pp. 84–87.
5. P. Dias, J. Madeira, and B. Sousa Santos, "Education: Teaching 3D Modelling and Visualization Using VTK," *Computers and Graphics*, vol. 32, no. 3, 2008, pp. 363–370.
6. B. Sousa Santos et al., "Integrating User Studies

into Computer Graphics-Related Courses," *IEEE Computer Graphics and Applications*, vol. 31, no. 5, 2011, pp. 94–96.

7. C. Müller et al., "Large-Scale Visualization Projects for Teaching Software Engineering," *IEEE Computer Graphics and Applications*, vol. 32, no. 4, 2012, pp. 14–19.

8. G. Domik, "Fostering Collaboration and Self-Motivated Learning: Best Practices in a One-Semester Visualization Course," *IEEE Computer Graphics and Applications*, vol. 32, no. 1, 2012, pp. 87–91.

9. C. Ware, *Information Visualization: Perception for Design*, 3rd ed., Morgan Kaufmann, 2012.

10. M.O. Ward, G. Grinstein, and D. Keim, *Interactive Data Visualization*, AK Peters, 2010.

11. J.J. Thomas and K.A. Cook, eds., *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE, 2005.

12. M. Bailey and S. Cunningham, *Graphics Shaders: Theory and Practice*, 2nd ed., AK Peters, 2011.

G. Scott Owen is professor emeritus at Georgia State University's Department of Computer Science. Contact him at sowen@gsu.edu.

Gitta Domik is a professor at the University of Paderborn's Faculty for Electrical Engineering, Computer Science, and Mathematics. Contact her at domik@uni-paderborn.de.

David S. Ebert is the Silicon Valley Professor of Electrical and Computer Engineering at Purdue University's School of Electrical and Computer Engineering. Contact him at ebertd@purdue.edu.

Jörn Kohlhammer heads the Competence Center for Information Visualization and Visual Analytics at Fraunhofer IGD Darmstadt and is a member of the Interactive Graphics Systems Group at Technische Universität Darmstadt. Contact him at joern.kohlhammer@igd.fraunhofer.de.

Holly Rushmeier is a professor and the chair of computer science at Yale University. Contact her at holly@acm.org.

Beatriz Sousa Santos is an associate professor in the University of Aveiro's Department of Electronics, Telecommunications, and Informatics. Contact her at bss@det.ua.pt.

Daniel Weiskopf is a professor of computer science at the University of Stuttgart. Contact him at weiskopf@visus.uni-stuttgart.de.

Contact department editors Gitta Domik at domik@uni-paderborn.de and Scott Owen at sowen@gsu.edu.

Leveraging Multidisciplinarity in a Visual Analytics Graduate Course

Niklas Elmqvist and David S. Ebert
Purdue University

The emerging field of visual analytics is defined as the science of analytical reasoning aided by visual interfaces, and was coined as a scientific term as late as 2005.[1] VA is a rapidly growing area in several branches of society. For example, in academia, the annual IEEE Conference on Visual Analytics Science and Technology (VAST) is in its seventh year and has been steadily growing. In government, an increasing number of agencies across the globe are adopting and funding VA. In industry, self-proclaimed VA companies, such as Tableau, Oculus, and i2, are successfully selling data analysis software to an expanding market of companies in such varied areas as business, health, and manufacturing.

However, VA's increasing popularity is a two-edged sword. Whereas stakeholders in academia, government, and industry alike are expanding and, therefore, desperately seeking recruits with VA expertise, higher-education institutions are struggling to respond to this sudden need. This problem is compounded by the multidisciplinary nature of VA, which leads to three main challenges.

First, the subject matter of what constitutes VA is wide and encompasses much more material than a standard graduate course can comfortably cover (see Figure 1). Yet, the burden of teaching all this material usually falls onto the VA course because traditional academic departments offer few complementary courses. Second, students interested in taking a VA course typically come from across the academic spectrum and therefore have diverse, nonoverlapping backgrounds. Finally, the requirements for teaching such a disparate set of topics falls to the instructor, who can't realistically be expected to be an expert on the full subject matter.

Here, we report on best practices we've discovered while teaching an experimental graduate VA course at Purdue University's School of Electrical and Computer Engineering. These practices transform the multidisciplinary challenges we just described into strengths that will both produce competent VA researchers and attract new students to the field.

ECE 695D

The course, ECE 695D (Introduction to Visual Analytics), has been taught twice: in 2009 (12 students) and 2011 (14 students). It's a standard three-credit, 15-week course open to any Purdue graduate student. The only prerequisite is experience in a topic such as statistics, data analysis, or visualization. Even though the School of Electrical and Computer Engineering offers the course, we don't require that students have programming expertise. That's because we wish to attract a multidisciplinary array of students who want to apply VA to their own research.

ECE 695D centers on a semester-long research project. We teach it in the Fall semester (in odd years) so that students with strong projects can polish and improve them for submission to one of the IEEE VisWeek conferences, which have deadlines at the end of March. Because we designed the project to teach every step of conducting a successful research project, the course includes regularly spaced deadlines for inception, literature survey, alpha and beta release, final paper, and peer review. These projects can be done individually or in teams of up to three students. We, as instructors, give continual feedback on the submitted material, ensuring that students stay on track and make progress in a worthwhile direction. In general, student feedback on the projects has been very positive; they seem to particularly enjoy working on novel research projects that could result in a paper in a peer-reviewed publication.

Besides the project, we ask students to perform an analytical exercise with the Atlantic Storm dataset (approximately 50 police and intelligence documents containing a hidden threat)[2] during the course's first two weeks. In the first week, students try to identify the hidden threat using pen, paper, and standard software tools; in the second, they do the same using the Jigsaw VA tool.[3] This exercise aims to give them first-hand exposure, as early as possible, to the challenges of investigative analysis. We based it on the exercise used to enlighten the panel who created the VA research agenda.[1] In particular, it often reveals confirmation bias; students tend to use Jigsaw simply to confirm what they thought they knew from the pen-and-paper phase. Students seem to find this lesson instructive.

Multidisciplinary Topics

A VA course by necessity becomes something of an umbrella course that introduces the basics of many different topics (see Figure 1). Yet, it can't examine any of these topics in depth. This is unfortunately a fact of life. The courses that would teach each of these topics in depth are spread across multiple traditional academic departments such as computer science, statistics, psychology, informatics, and engineering. Few students can take all these courses separately. In some cases, a university might not even offer all the required graduate courses.

So, designing, organizing, and teaching a VA course becomes a problem of describing each topic's bare essentials in an order that makes sense to the students, while charting the VA landscape to inspire future exploration by the students. Table 1 shows our current 15-week schedule. The first five weeks heavily emphasize the human analytical process; most of this material is new to the
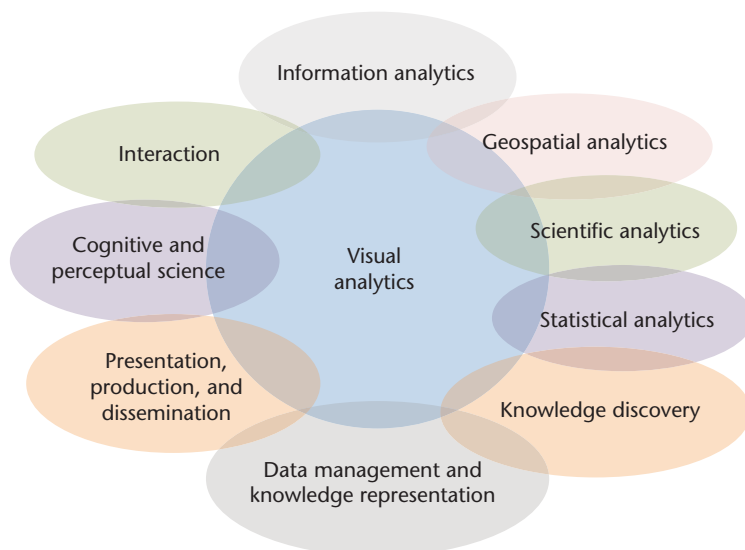
Figure 1. The multidisciplinary scope of visual analytics (created and refined by Jim Thomas, Daniel Keim, David Ebert, Niklas Elmqvist, and others). This topic encompasses much more material than a standard 15-week graduate course can comfortably cover.

students yet introduces them to topics central to designing effective VA systems.

Certainly, a VA course such as this would be even better if it could rely on an array of advanced graduate courses that went into detail on each topic, letting students specialize in a particular area. At Purdue, this is possible only for statistics, cognitive science, human factors, and visualization. Most other graduate courses in other departments don't cover the subject matter that ECE 695D introduces. This naturally puts additional strain on instructor expertise, as we discuss later.

We designed ECE 695D to complement our graduate course on visualization, and we expect virtually all graduate students who take one of the courses to also take the other (although in an unspecified order). So, ECE 695D doesn't spend much

Table 1. The 15-week syllabus for ECE 695D (Introduction to Visual Analytics).

Week	Topic	Content
1	Introduction	Analytical exercise
2–3	Analytical reasoning	The analysis process, critical thinking, sensemaking, and situation awareness
4	Perception	Human perception, preattentiveness, color, shape, and texture
5	Cognition	Cognitive theory
6–7	Data management	Representations, transformations, and statistics (temporal and spatial)
8–9	Visual representations	Visualization techniques
10–11	Interaction	Interaction techniques
12	Communication	Production, presentation, and dissemination
13	Collaboration	Collaborative VA
14	Evaluation	Evaluating VA
15	Advanced topics	Conducting VA research, novel computing platforms, and mobile VA

time on visual representations, and our visualization course doesn't spend much time on VA. We recommend that all universities that teach one of these courses should also teach the other. However, if a choice must be made, VA is preferable owing to its wider scope, applicability, and general interest. Students can always individually read up on the visualization techniques relevant to their projects, provided that instructors give them the foundation and references.

> *We explicitly market the course as one to take if you have a data analysis problem to solve for your own research, regardless of the field.*

Diverse Student Backgrounds

In our two offerings of ECE 695D, we've taught an astoundingly diverse pool of students, including students from not only traditional engineering departments such as computer, mechanical, and industrial engineering but also forestry, psychology, physics, geomatics, education, economics, and technology. We explicitly market the course as one to take if you have a data analysis problem to solve for your own research, regardless of the field.

Because our students aren't required to have software development expertise, the course project could pose a significant challenge. Contributions to peer-reviewed VA conferences typically involve a nontrivial amount of software development to build the implementations of the novel techniques presented in the papers. However, such software development is clearly out of reach for most students with no CS or CE background. Also, asking them to learn not only VA but also programming during a single semester is clearly unfair and unrealistic.

We've dealt with this challenge in two ways that leverage the diversity of student backgrounds instead of treating it as a problem.

First, we employ multiple project types. The IEEE VisWeek call for papers explicitly identifies contribution types (evaluations, design studies, and models[4]) that don't necessarily have to involve building new software systems. Accordingly, we let students work on any of these project types, enabling them to harness their expertise in a specific field toward their project.

Second, we form multidisciplinary teams. In several instances, we've been able to form teams consisting of both technical and nontechnical students, thereby providing complementary expertise that significantly multiplied the teams' capabilities. In one case, a team consisted of an industrial engineer with human-factors expertise and two computer engineers with strong software development expertise. Their paper is currently undergoing review for a prestigious visualization journal.

Demands on Instructor Expertise

Because instructors can't be intimately familiar with all the topics a VA course should contain, they can easily get overwhelmed when preparing for and teaching those topics.

We've found that the best remedy is to share this burden between a team of instructors. We're both co-instructors for ECE 695D and have been closely involved in lecturing, writing notes, evaluating, grading, and giving feedback for this course's different components. We have similar yet complementary expertise in the different course topics and have taken turns lecturing on them. This has also helped us expand our VA expertise. Furthermore, we've brought in guest lecturers from our colleagues across campus, from the pool of senior students in our research groups, and even from different universities (typically in conjunction with research visits). Several times we've recommended that students attend relevant seminars elsewhere in lieu of traditionally scheduled lectures. These guest lecturers and seminars have been able to offer a different perspective and different expertise on aspects of the course. We feel that this can only widen the students' horizons on VA, which is the course's whole purpose.

Unfortunately, no comprehensive VA textbook exists that could serve as the course's backbone, which is a major barrier against adopting VA as a topic in graduate curricula. Visualization textbooks are fairly common—such as those by Matthew Ward and his colleagues,[5] Colin Ware,[6] and Robert Spence.[7] However, they tend to focus on visual representations only and therefore generally fail to cover VA's entire scope. Until someone writes a VA textbook, we hope that the course schedule in Table 1 will serve as a guideline on which topics to cover. Furthermore, the Visual Analytics Digital Library (VADL; http://vadl.cc.gatech.edu) contains additional resources, syllabi, slides, lecture notes, and schedules that VA instructors can adopt and extend.

Student Results and Success

Several ECE 695D projects have become fully fledged VA papers published in peer-reviewed

conference proceedings (Figure 2 shows two examples). Several other student papers are undergoing review for publication. In some cases, the published papers were the first ones those students had published in a VA conference proceedings. This suggests that ECE 695D has been a stepping stone for students to enter the VA scientific community.

Designing and teaching a VA course can be challenging, but our experience is that instructors can harness the field's broad scope to make the course even more engaging and compelling. So far, our course has received top ratings in course surveys and seems popular with the graduate students. We hope that some of the best practices we've reported here will help other instructors improve and refine their VA education programs.

Furthermore, we'd love to hear about your experiences teaching VA; see our bios at the end of this article for contact information. The ECE 695D website is at https://engineering.purdue.edu/~elm/teaching/ece695d.

References

1. J.J. Thomas and K.A. Cook., eds., *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE CS, 2005.
2. F.J. Hughes, "All Fall Down," *Discovery, Proof, Choice: The Art and Science of the Process of Intelligence Analysis*, tech. report, 2005, case study 6.
3. J. Stasko et al., "Jigsaw: Supporting Investigative Analysis through Interactive Visualization," *Proc. 2007 IEEE Symp. Visual Analytics Science and Technology* (VAST 07), IEEE, 2007, pp. 131–138.
4. T. Munzner. "Process and Pitfalls in Writing Information Visualization Research Papers," *Information Visualization: Human-Centered Issues and Perspectives*, A. Kerren et al., eds., LNCS 4950, Springer, 2008, pp. 134–153.
5. M.O. Ward, G. Grinstein, and D. Keim, *Interactive Data Visualization: Foundations, Techniques, and Applications*, A K Peters, 2010.
6. C. Ware, *Information Visualization: Perception for Design*, Morgan Kaufmann, 2004.
7. R. Spence, *Information Visualization: Design for Interaction*, Prentice Hall, 2007.
8. K. Kim et al., "WordBridge: Using Composite Tag Clouds in Node-Link Diagrams for Visualizing Content and Relations in Text Corpora," *Proc. 44th Hawaii Int'l Conf. System Sciences*, IEEE, 2011; pp. 1–8.
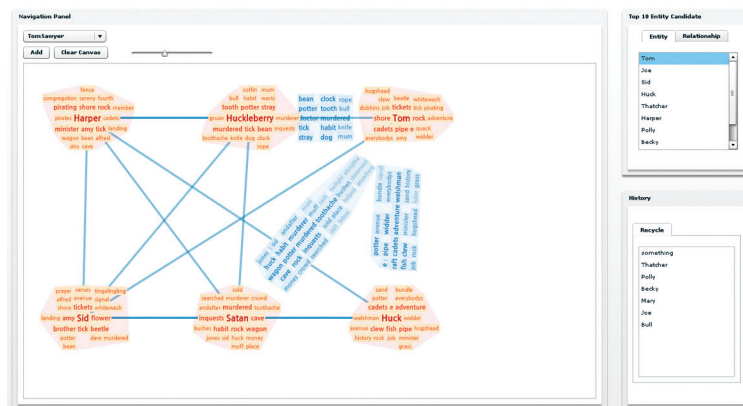
(a)
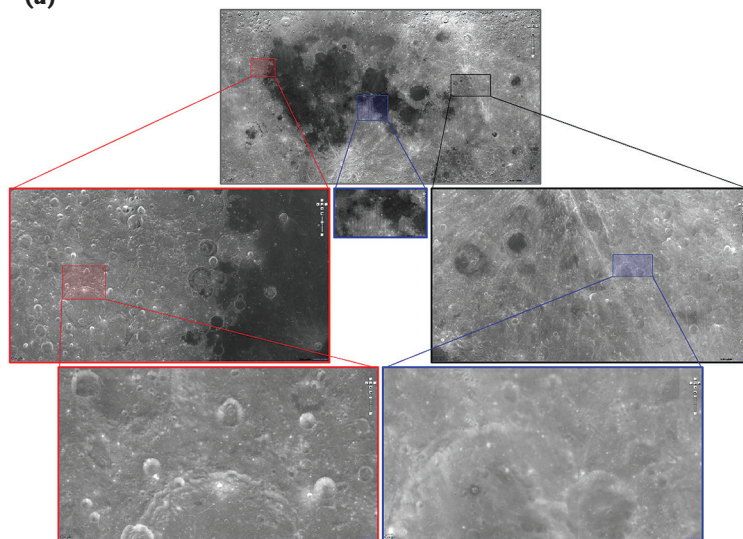
(b)

Figure 2. Two student projects that have led to papers published in peer-reviewed publications. (a) A WordBridge text analytics visualization showing relations in document collections.[8] (b) Using the PolyZoom multifocus technique to explore a lunar map dataset.[9]

9. W. Javed, S. Ghani, and N. Elmqvist, "PolyZoom: Multiscale and Multifocus Exploration in 2D Visual Spaces," to be published in *Proc. ACM CHI Conf. Human Factors in Computing Systems*, ACM, 2012.

Niklas Elmqvist *is an assistant professor in Purdue University's School of Electrical and Computer Engineering. Contact him at elm@purdue.edu.*

David S. Ebert *is the Silicon Valley Professor in Purdue University's School of Electrical and Computer Engineering. Contact him at ebertd@purdue.edu.*

Contact department editors Gitta Domik at domik@uni-paderborn.de and Scott Owen at sowen@gsu.edu.

MarketAnalyzer: An Interactive Visual Analytics System for Analyzing Competitive Advantage Using Point of Sale Data

S. Ko[†1], R. Maciejewski[‡2], Y. Jang[§3], and D. S. Ebert[¶1]

[1]Purdue University
[2]Arizona State University
[3]Sejong University

Abstract
Competitive intelligence is a systematic approach for gathering, analyzing, and managing information to make informed business decisions. Many companies use competitive intelligence to identify risks and opportunities within markets. Point of sale data that retailers share with vendors is of critical importance in developing competitive intelligence. However, existing tools do not easily enable the analysis of such large and complex data. therefore, new approaches are needed in order to facilitate better analysis and decision making. In this paper, we present MarketAnalyzer, an interactive visual analytics system designed to allow vendors to increase their competitive intelligence. MarketAnalyzer utilizes pixel-based matrices to present sale data, trends, and market share growths of products of the entire market within a single display. These matrices are augmented by advanced underlying analytical methods to enable the quick evaluation of growth and risk within market sectors. Furthermore, our system enables the aggregation of point of sale data in geographical views that provide analysts with the ability to explore the impact of regional demographics and trends. Additionally, overview and detailed information is provided through a series of coordinated multiple views. In order to demonstrate the effectiveness of our system, we provide two use-case scenarios as well as feedback from market analysts.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Pixel-based visualization, Linked Views, Geospatial, Temporal, Multi-variate, Business Intelligence, Competitive Intelligence, Market Analysis—

1. Introduction

The underlying goal of a business is to increase (or at least maintain) its current market share and to maximize its profits within the market. In order to pursue this goal, analysts must constantly explore and analyze market share data changes that are relevant to their current business sector. Their goal is to forecast changes in the market as a means of controlling and expanding the company's current market share. This exploration, analysis, and prediction of the market share is termed *competitive intelligence (CI)* [Kah98]. Companies use CI to compare themselves to other companies, to identify market risks and opportunities and to evaluate the potential impact of new sales strategies.

In order to generate intelligence reports, many companies extract information from a variety of sources using various methods of data collection and analysis (e.g., networking with company rivals, examining security filings, patent application analysis). One key data source is point of sale data that retailers share with vendors. This point of sale data is temporal, multivariate, and spatial in nature; therefore, it is well suited for analysis in a visual analytics environment. However, it is difficult to find systems that manage the characteristics of point of sale data effectively. In this paper, we present MarketAnalyzer, a visual analytics system for exploring, comparing, analyzing, and predicting trends of point of sale data. We have worked directly with analysts to provide proper and accurate analysis of their point of sale data (e.g., 288 stores with 36 different products) to increase their understanding and improve their market insight. We use an enhanced pixel-based visualization approach [KK94, Kei00] in MarketAnalyzer to efficiently utilize limited screen space for the large store and product

---

† ko@purdue.edu
‡ rmacieje@asu.edu
§ jangy@sejong.edu: corresponding author
¶ ebertd@purdue.edu

information. Our system allows analysts to explore current sales volume, trend, and temporal market share growth rates using a series of linked views including pixel-based visualization matrices, line graphs, stacked bar graphs, and choropleth maps.

MarketAnalyzer has several benefits compared to other tools for performing market analysis tasks. MarketAnalyzer enables analysts to investigate the status of the market by observing all the characteristics of point of sale data at the same time. In addition, the status of the competition in point of sales, trends and growth rates is projected onto a map for regional market analysis. MarketAnalyzer provides forecasts for both individual products and different stores utilizing statistical models, such as linear trend estimation [DS98] and ARIMA (Auto-Regressive Integrated Moving Average) [BJ76]. In order to reduce the perceptual difficulties inherent in pixel-based visualizations, a local magnification lens is also provided for focus + context analysis. Additionally, CUSUM [Pag54] and normalized trend filtering are provided for data filtering. For evaluation, we provide two case studies that describe how sales data can be analyzed with MarketAnalyzer. Although the case studies are business domain specific, it is easy to extend our system to other multivariate, spatial, temporal datasets such as property sales, crime and disease data to provide comparisons, insight, and new intuition.

2. Related Work

Traditionally various tasks, such as discovering market trends and predicting future prices of assets have been addressed with charts and line graphs in the financial data domain [Mur99, EM01]. While charts and line graphs provide useful visualizations of univariate data, they quickly become clutter as new dimensions are added. Analysts often use tree map visualizations to represent the market [Sma, VvWvdL06]. Unfortunately, these maps only provide a snapshot of the current market value whereas analysts often wish to explore short or long term trajectories within the market [TA03, STKF07].

In order to display the maximum amount of data relative to the screen space, Keim et al. introduce pixel-based [OJS*11] or pixel-oriented [KK94, Kei00] visualization techniques. In these techniques, each data element is assigned to a pixel. Then, a predefined color map is used to shade the pixel to represent the range of the data attribute. Thus, the amount of information in the visualization is theoretically limited only by the resolution of the screen. In the context of our work, pixel-based visualizations are visualizations that utilize small areas of the screen to encode one data item. Note that the areas may not necessarily be pixels, as the use of small rectangles also falls under the accepted classification of pixel-based visualizations [OJS*11].

Borgo et al. [BPC*10] present how the usability of the pixel-based visualization varies over different tasks and block resolutions. Oelke et al. [OJS*11] studies visual boosting techniques for pixel-based visualization such as halos and distortion. Ziegler et al. [ZNK08] presents how the pixel-based visualization helps analysts gain insight for long-term investments.

Many systems have been developed for visually exploring multivariate data (e.g., Xmdv [War94], Spotfire [Ahl96], XGobi [SCB98], GGobi [SLBC03], Comvis [MFGH08], Polaris [STH08], Tableau [Tab]). Common amongst these systems is the extensive use of interactive techniques (brushing, linking, zooming, filtering) to refine the user's queries. However, such systems often do not support market forecasting or geographical analysis. In discussing design strategies with our market analysts, it was noted that forecasting future trends and understanding outperforming geographical locations are important in market analysis. Of the systems previously listed, Tableau software [Tab] allows analysts to easily access and analyze their data by offering flexible operations. Although multivariate and time-series data analysis is possible in the tool, comparison between multivariate attributes with geographical information is not well supported by Tableau. In MarketAnalyzer, all attributes of the data is visualized in multiple linked views for simultaneous comparison, and analysts can investigate future sale trends based on statistical models and market share growth rates.

3. Visual Analytic Environment

Analysts tend to easily understand competition within a market and quickly draw conclusions when maximal information is presented. In order to present the most information for analysis, coordinated and multiple linked views have been used in various applications [SFOL04, WFR*07, CGK*07, Rob07, SGL08]. In this work, we also employ coordinated multiple linked views to visualize attributes from point of sale data. Figure 1 shows how MarketAnalyzer provides complete information in multiple linked views. Note that all color maps are chosen to fit perceptually with the data being analyzed. Both sequential and divergent color maps from ColorBrewer (which have been previously tested and evaluated [HB03]) are used. Sequential color scales are chosen to show ordered data, while divergent color maps are chosen to show differences in data values with respect to some point of interest.

Companies are displayed in a selectable list in (a), stores in (b), and products in (c). Note that in window (b), there are two selectable lists. The leftmost store list is used to select multiple stores for computing the sales average, while the rightmost list is used to select a single store, whose sales will be compared to the computed sales average. This single store selection is the anchor for view (h) and (i) that shows the sales of the primary company (h) or competitor company (i). In (h) and (i), the selected store's sales of the products chosen in the list view (c) are plotted in green, and the average sales of the products across a group of stores selected

Figure 1: *The MarketAnalyzer interface. MarketAnalyzer consists of multiple coordinated views linked with interactive filters: (a) Company filter, (b) Store filter, (c) Products filter, (d) Legend view, (e) (Sorted) Matrix view for sales, trends, and growth rates. (f) Stacked bar view, (g) Geographical view, (h) and (i) Line graph small multiples views, (j) and (k) Time slider widgets and aggregation tools for temporal comparison. (l) Tooltip. (m) Filter. In the legends, the blue indicates positive and the red represents negative measurements in sales, trends, or growth rates.*

in the leftmost list of (b) are plotted in purple. At the end of the line graph, a four week forecast for sales based on an ARIMA model is plotted as a blue line bounded by two red dotted lines representing the upper and lower error bounds. Note that the ARIMA forecasts are calculated in *R* [R D06], which is integrated directly into our system.

3.1. Pixel-Oriented Display Matrix

A fundamental challenge in the visualization of large multivariate data is that screen space limits the amount of information that can be simultaneously presented to a user. This scalability problem causes various difficulties in analysis, such as inefficiency in comparison and tedious jumping back and forth to adjust different parameters. In order to alleviate this problem, we incorporate a pixel-based visualization [KK94, Kei00] that is effective when the screen space is limited. Sales, trends, and market share growth rates for stores and products (e.g., 288 stores, 36 products) are effectively presented in MarketAnalyzer as shown in Figure 1 (e). Note that we place different stores and products in different columns and rows, respectively. The matrix view in (e) consists of three views: sales, trends, and growth views. Each view has two matrices for a primary company and its

competitor that are chosen in Figure 1 (a) by users. We place small squares side-by-side in each matrix with all matrices positioned vertically. The rationale behind this arrangement is that its conceptual simplicity makes the comparison and discovery of trends easier. Also comparing data side-by-side is more efficient than jumping back and forth and memorizing previously shown data, based on the principle of small multiples [Tuf90].

In the sales view, we use Equation 1 to define the sales ($S_{i,j}$) for each square as a sum of the sales during the user-selected time interval.

$$S_{i,j} = \sum_{t=m}^{n} Sales(t,i,j), \qquad (1)$$

where i and j indicate the i_{th} row (product) and the j_{th} column (store) while m and n are the first and the last month in the time interval. The darker the blue, the more units are being sold. One frequent question that decision-makers might have is "Are sales increasing in this specific time period?". In order to answer this question, we present the user with a trend view in which the slope (variable b in Equation 2) of our linear trend estimation is visualized using a divergent color scale. Positive slopes (indicating that sales are trend-

Feb-May 2009

Feb-May 2010
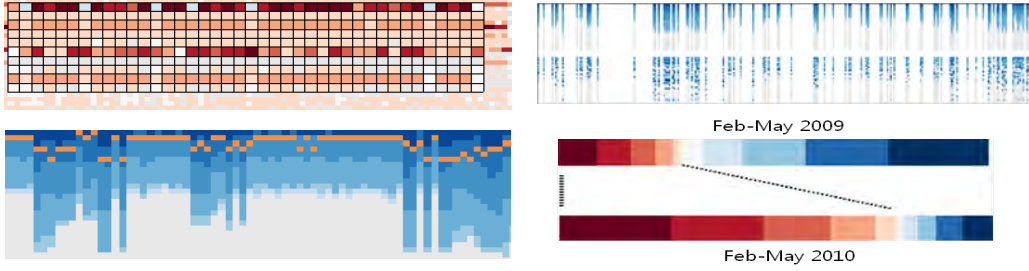
Figure 2: *(top-left) Magnification is applied for detailed comparison, (top-right) the CUSUM filtering method with the strict option is applied on Feb 2010, (bottom-left) SimulSort is applied to the sales view, (bottom-right) Proportional legend.*

ing upwards) are mapped towards the blue values, negative slopes (indicating sales are trending downward) towards the red values, with a slope of zero being white (indicating that sales are stable).

$$y = a + bx, \qquad (2)$$

where x, y are variables, a is the intercept point of the regression line and the y axis, b is the slope from the linear trend estimation for a certain time interval. b is computed in Equation 3,

$$b = \frac{n(\sum_{m=1}^{n} X_m Y_m) - (\sum_{m=1}^{n} X_m)(\sum_{m=1}^{n} Y_m)}{n(\sum_{m=1}^{n} X_m^2) - (\sum_{m=1}^{n} Y_m)^2} \qquad (3)$$

Here, n represents the number of months in the specified time interval, Y_m is the index of each month (e.g., 1, 2, ..., n), and X_m is the sales for the month.

Along with sales trends, we also define the growth rate of a business using Equation 4.

$$G_{i,j} = \frac{S_{i,j,cur} - S_{i,j,past}}{S_{i,j,past}} \times 100, \qquad (4)$$

where $S_{i,j,cur}$ is defined as the sum of the sales between the first (M_f) and the last (M_l) month using the red time slider (Figure 1-(k)), as shown in Equation 5.

$$S_{i,j,cur} = \sum_{T_{red}=M_f}^{M_l} Sales(T_{red}, i, j). \qquad (5)$$

$S_{i,j,past}$ is similarly defined for a past time interval (T_{blue}) from the blue time slider (Figure 1-(j)). A divergent color scale is applied to represent the growth rate. When $S_{i,j,past}$ is zero, the growth rate is set to the maximum rate in the range. This indicates a case where a product was not supplied at the selected time but was provided later at the time specified by the red time slider.

We calculate the competitive advantage for the primary company ($C_{P_{i,j}}$) using Equation 6 when the analyst activates the comparative mode, as shown in Figure 5.

$$C_{P_{i,j}} = M_{P_{i,j}} - M_{C_{i,j}}, \qquad (6)$$

where $M_{P_{i,j}}$ is a measurement of the i_{th} product from the j_{th} store for the primary company and $M_{C_{i,j}}$ is that of the competitor. The measurement can be any input data attribute, such as sales, trends, and growth rates used in these examples. This calculates a new matrix representing the competitive advantage of the primary company. Analysts should be alerted about a red row or a red column because it explicitly indicates the primary company is losing sales in comparison to its competitor as shown in Figure 5.

In order to provide more detailed information, we use a tooltip and magnification lens. A tooltip, as shown in Figure 1 (l), provides numerical information of the product in the store where a mouse is hovering. Squares are gray when no product is supplied to the stores. In order to help recognition and comparison, MarketAnalyzer provides a magnification lens as shown in Figure 2 (top-left).

We reorder rows in the matrix to emphasize the importance of data where squares near the top-left locations are more important than squares near the bottom-right. For the sales matrix, rows and columns are sorted based on sales sums. Products (rows) are sorted by the sales sum across the entire stores and then stores (columns) are re-sorted by the sales sum across the entire products. The topmost row indicates the top-selling product, and the leftmost column represents the top-selling store in the market as shown in Figure 1 (e). Note that matrices in the trend and growth views are also sorted corresponding to the sorting of the sales matrix. Negative measurements in trends and growth rates represent adverse situations and analysts want to find out the adverseness promptly in their analysis. In this case we use adverseness for the importance.

We also incorporate SimulSort [HY09] to sort columns independently as an auxiliary approach. In each store, products are sorted by sales. Thus, squares with higher sales tend to go upward in matrices. This enables analysts to quickly evaluate sales performance of products in a store. In order to see the performance of a product across all stores, MarketAnalyzer highlights the product in orange for all stores as presented in Figure 2 (bottom-left) where the sales performance of the product is good in most stores with some minor variations.

(a) Stacked bar view   (b) Geographical view
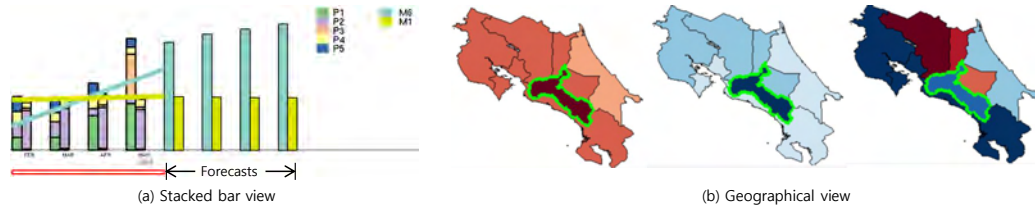
Figure 3: *(a) The stacked bars represent trends in individual products. Analysts can see sudden increase in sales of product P3 in May. Note that M1 and M6 indicate company names. (b) Competitive advantages in February–May 2010 are linked on the maps to represent regional competitions for sales (left), trends (middle), and growth rates (right) compared to the past time interval, March–June 2009.*

3.2. Filtering Methods

MarketAnalyzer helps analysts filter out uninteresting stores by providing two filtering methods: the cumulative summation method (CUSUM) [Pag54] applied to sales and the trend filtering method based on the computed trends. The CUSUM filtering method has two modes: default and strict. In the default mode, the CUSUM method in MarketAnalyzer filters out stores during the visualization process that have a negative $S_{st,n}$ from Equation 7.

$$S_{st,n} = \sum_{m=1}^{n} (X_{st,m} - \mu_{st}) \qquad (7)$$

Here, $S_{st,n}$ is the CUSUM of the st store in the n_{th} month (where the n_{th} month is the first month of the red time slider). $X_{st,m}$ is the sum of all products of store st at the m_{th} month (where the m_{th} month is started from the first month of our data set–January 2009). μ_{st} is the mean sales of st over the time period n-m. The CUSUM method can also be changed from the default parameter to a strict mode. In the strict mode, stores are filtered out only when the CUSUM of the n_{th} month and the $n-1_{th}$ month are negative. If the CUSUM of the n_{th} month is negative but that of the $n-1_{th}$ month is positive, then the store will still be visualized.

When the trend filtering method is used instead of CUSUM, MarketAnlayzer maps trends from -1 to 1. Then, MarketAnalyzer visualizes stores whose trends are below the threshold specified in the filter (Figure 1 (m)). The normalized trend N_{st} for store st is computed in Equation 8.

$$N_{st} = \frac{T_{st} - T_{mean}}{\sqrt{T_{var}}} \qquad (8)$$

Here, T_{st} is the sum of the trends of all products in the store st, T_{mean} and T_{var} are the mean and variance of trends across all stores. Then, N_{st} is mapped to M_{st} in Equation 9.

$$M_{st} = \left(\frac{N_{st} - N_{min}}{N_{max} - N_{min}} - 0.5 \right) * 2, \qquad (9)$$

where N_{max} and N_{min} are the maximum and minimum in normalized trends.

3.3. Proportional Legends

The legend view in Figure 1 (d) provides numeric and color map information for each view. The scale in MarketAna-

lyzer can be adjusted to have the denominator represent either measurements (sales, trends, and growth rates) at the local level (comparing measurements within a company) or at the global level (comparing measurements across the primary company and its competitor). In the local scale mode, six legends (two for each view) are provided, while three legends (one for each view) are provided in the global scale mode.

Generally a legend consists of evenly divided intervals. However, analysts often need to walk through a huge volume of data within a relatively short period to check for abnormalities. For the market analysis, the legend can be as important as the data itself [Cle94, Wil05, TLH10]. In order to enhance understanding abnormalities of the data, the intervals for the legends in MarketAnalyzer can be nonlinearly mapped according to importance as shown in Figure 2 (bottom-right). Analysts need to be alerted when red intervals are much wider than blue intervals. For instance, the change of the width for negative measurements during investigation explicitly presents the worst growth rates.

3.4. Geographical View

The geographical view as shown in Figure 1 (g) supports the visual analysis of spatiotemporal patterns. The regional status of competition is a measure of the difference between two selected companies within the selected geographical area and is represented by colors. For instance, analysts are able to see that the primary company has lower sales (red) compared to its competitor in the left most image in Figure 3 (b). However, the market looks optimistic (blue) for the primary company because the sales trend is increasing (middle image in Figure 3 (b)). Note that when a mouse hovers on a store in the matrix view, the corresponding region on the map is highlighted.

Dynamic querying for specifying time intervals also plays an important role in geographic analysis because sales patterns evolve over time. This querying helps analysts identify complicated spatiotemporal patterns in the competition and facilitate appropriate strategies. For instance, analysts can easily find the time when the leading competitor started losing its competitive advantage. Then, the analyst can start investigating reasons for the loss by investigating informa-
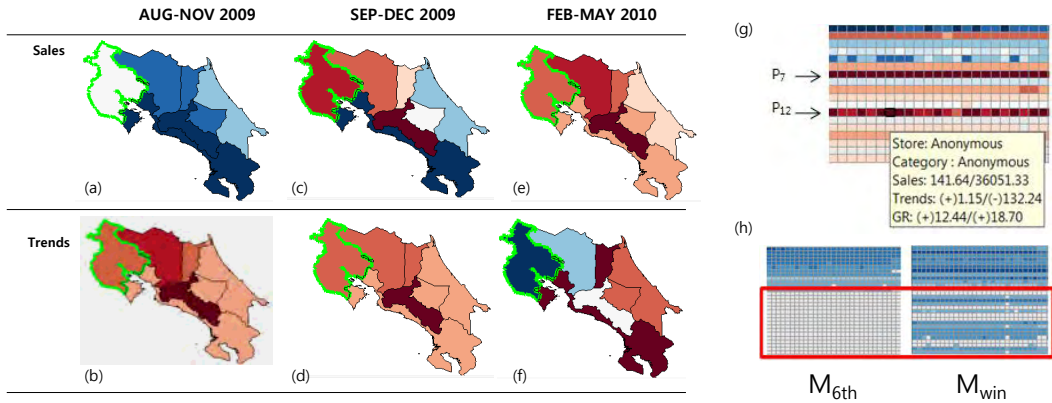
Figure 4: *(a)–(f) Analysis from the geographical view. The sales row shows the process of losing competitive advantage while the trends row presents forecasts for each column interval (August 2009–May 2010). The blue color represents good performance while the red color represents bad performance for the primary company compared to its competitor. (g) The trend view helps an analyst design short term tactics such as promotions. This example shows a decreasing overall sales trend of the competitor in some stores. The analyst notes that the competitor has the worst downward trend in products P_7 and P_{12}. (h) The red box represents possible new markets for the new company M_{6th} but its competitor M_{win} has already started its business in various products.*

tion on the matrices and other views. These processes during analysis are described in Section 4.

3.5. Stacked Bar View and Time Sliders

Well-designed stacked graphs are popular because of their aesthetics and ease of perception [HHWN02, BW08]. Since it is important to verify combinatorial trends of multiple products, we employ stacked bar graphs to investigate these trends as shown in Figure 1 (f) whose products and stores are chosen from (b) and (c). When a person buys two products together, there could be various assumptions to explain why these two are chosen. It may be because of complementary relationship between two products (e.g., ketchup and mustard), similar purchase cycles (e.g., beer and diapers), impulse purchases of a product (e.g., candy bar), or undiscovered reasons, such as happenstance [MAG99]. Figure 3 (a) shows an example where one can see that as the sales of P_1 increases and reaches a certain level, the sales of P_3 rapidly increase as well.

Draggable and length-changeable time sliders are used to select time intervals for the stacked bar view, as shown in Figure 1 (j) and (k), respectively. The red slider (k) is for specifying current time interval. For the growth rate computation, a past time interval is required that is selected using the blue slider (j). When the slider is dragged or the range of the slider is changed, the time interval in the analysis also changes, updating all other views with the new time interval.

4. Case Studies

We describe two scenarios with anonymous manufacturers, stores, and products for privacy. The first scenario presents the process of designing a strategy for a young company that wants to increase its market share. The second scenario illustrates the analysis of a defending champion that needs to verify and understand its competitive advantage compared to 94 other competitors in the market.

4.1. Analysis to Step into a New Market

An analyst in a two year old company (M_{6th}) is asked to look for potential areas in which the leading competitor (M_{win}) may currently be showing signs of weakness. By finding these weaknesses, the new company can begin making inroads into the market as it expands its base. The exploration begins in the pixel-based matrix views, as shown in Figure 1 (e). In the top matrix view, the y-axis represents all 36 products that are currently sold in the market. The x-axis represents all 288 stores that sell its products. The analyst chooses to sort the matrix by sales. Here, the darker the blue, the more units are being sold. In the local mode, the analyst can see which stores and products are the top sellers for M_{6th} and M_{win}. On the other hand, in the global mode, as shown in Figure 4 (h), the analyst sees that the competitor has more blue squares, meaning it is outperforming M_{6th}. In the red box in Figure 4 (h), the analyst easily verifies the products and stores in which the new company is not supplying any product while the competitor has been earning additional profit.

Before deciding on which market opportunities might be the most profitable, the analyst needs to understand the weakness and trends within the market. In the trend view, the analyst sees the rate of growth or decline over the last sales period. First, the sales period is selected using one time slider (e.g., February–May 2010) in the lower portion of the
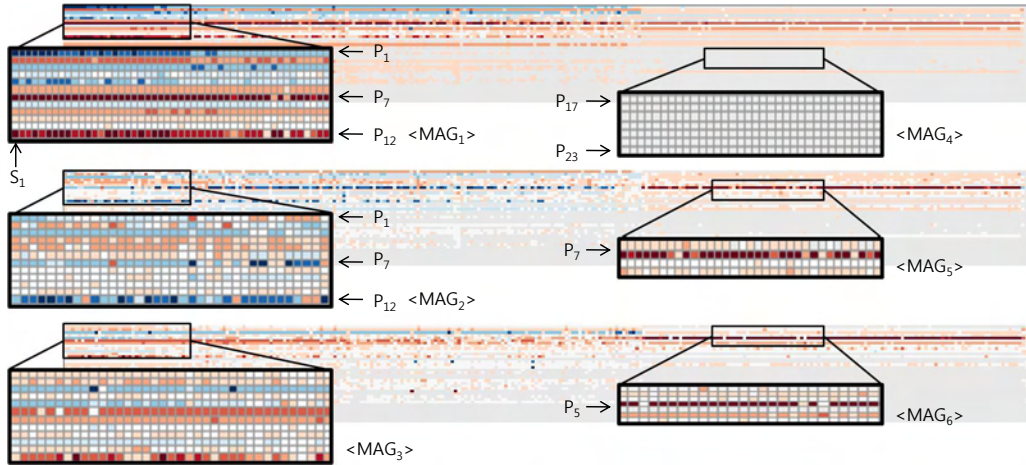
Figure 5: *Our pixel-based matrix views using the comparative mode. Each row of the matrix represents a product, and each column represents the store selling the product. The three views present three types of information (from top to bottom): sales, trend, and growth rate comparisons between two selected companies. Note that gray color indicates zero. The blue color presents positive and the red color represents negative in difference of the two measurements.*

stacked bar view as shown in Figure 1 (k) (red slider). Here, the analyst sees that many of the stores that the new company supplies are showing a downward growth trend. While alarming, it is important to note that the leading competitor is also displaying negative trends for similar products during this time period. Looking at the trend analysis, the analyst notes that the P_7 and P_{12} products of the competitor have the worst trends, which might indicate they are becoming less popular. In that case, they could be the targets for the new company to take the market share as shown in Figure 4 (g).

Of primary interest to the analyst is the matrix view under the comparative mode, where the analyst can explore how the current product sales is performing with respect to M_{win}. If a square is blue, then M_{6th} is outperforming M_{win}, and if red, then M_{win} is still outperforming the new company. Variations in the red and blue hue show the degree of sales performance. The analyst notices the following from Figure 5.

- P_1 is the best-selling product, outperforming its competitor (MAG_1),
- P_7 records the worst sales performance in almost all stores (MAG_1). The sales of the P_7 in two thirds of stores is expected to increase (MAG_2) while those in other stores will still keep decreasing (MAG_5),
- P_5 has not been sold much in one third of the stores (MAG_6),

In Figure 3 (a), the analyst sees that the sales of P_3 suddenly increased in May 2010 while those of P_1 has grown gradually. Thanks to increasing sales in these two products, M_{6th} has outperformed its competitor since May 2010 in those selected products and it is predicted to outperform the competitor with those products in June 2010 according to the ARIMA forecast.

The analyst further wants to explore the distribution of its

performance across stores and chooses to use the geographical view for regional competition analysis. Competition in the last quarter (February–May 2010), as presented in Figure 3, can be summarized as follows:

- Sales (Figure 3 (b) Left): M_{6th} has difficulty in overcoming its competitor in all regions. It is notable that the company is losing in the most important region including the capital city that is colored in the darkest red.
- Trends (Figure 3 (b) Middle): The trends are optimistic in all regions and the most important region has the highest upward trend.
- Growth rates (Figure 3 (b) Right): Sales in some regions have grown compared to sales in March–June 2009 although there are regions where growth rates are negative.

Therefore, the analyst will propose attacking the weaker points of the leading competitor (M_{win}). For instance, P_7 and P_{12} of the competitor in the trend view in Figure 4 (g) are expected to keep losing competitive advantage. Thus, they might be proper targets for launching additional new products, to take over the market. Conversely, to keep or increase the sales, a local promotion focusing on the capital city or a global advertisement should be designed for the P_7 product, as shown in MAG_1 in Figure 5. Giving up P_5 in MAG_6 might not be a good strategy since it still remains in 5th place in sales for the company. Watching other companies' strategies when they launch a new product, could be an effective solution at the moment. In addition, the analyst notices that there are many gray products where M_{win} does not supply its products. These would be the secondary locations where M_{6th} could launch new products for earning more profit. Lastly, finding reasons for the abnormal sales pattern between P_1 and P_3 in Figure 3 (a) is important as well since they are in the 1st and 3rd places in sales.
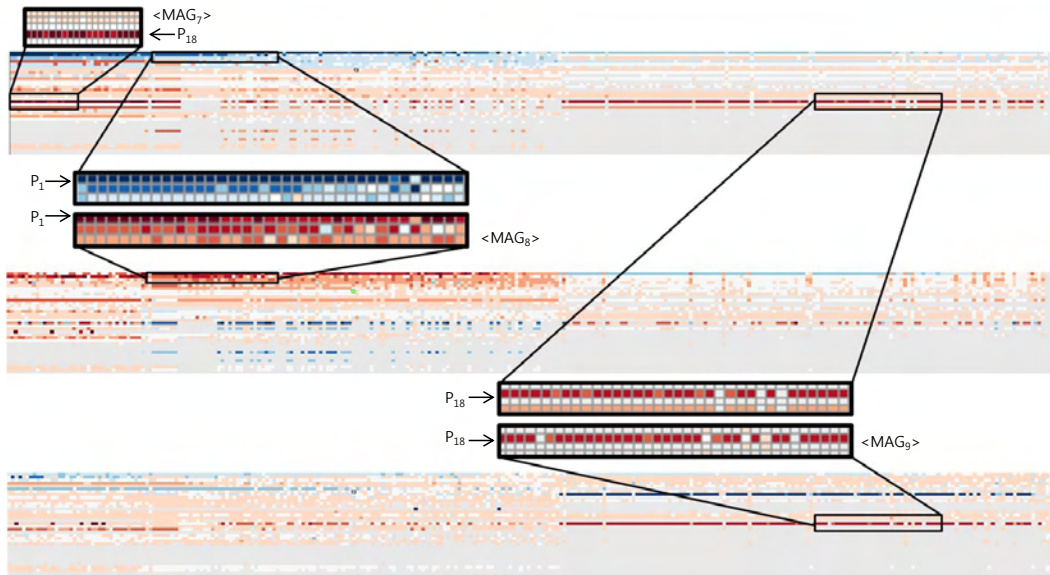
Figure 6: *The matrix view under comparative mode with sorting provides direct information for the competition. These matrices imply a possible strategic failure that could cause the loss of a big market.*

4.2. Analysis of a Defending Champion

Often, the leading competitor's analyst (company M_{win}) is asked to analyze market share competition that requires a performance comparison of all companies in all regions. The analyst might begin exploration in the geographical view that quickly shows the regional competition status for a given time interval. In Figure 4 (a), the analyst sees that M_{win} had market share higher than 50% before November 2009 but the trend during the period decreased in all regions (Figure 4 (b)). The analyst notes that the most important region, colored in the darkest red, was expected to have the most severe decrease in sales. From December 2009, as shown in Figure 4 (b), M_{win} started losing competitive advantage in three regions but still had three winning regions (Figure 4 (c)). It is notable that the regions that had moderate downward trends (Figure 4 (b)) still kept a competitive advantage, while others with darker red colors were losing the competitive advantage. At the same time, forecasts kept warning of downward trends in all regions (Figure 4 (d)). In the end, M_{win} did not have any winning region during February–May 2010 (Figure 4 (e)). However, Figure 4 (f) shows two regions having upward trends in sales (darker blue colors). During the analysis, the analyst sees that the proportional legends keep changing. For instance, the red area representing an adverse situation, only occupied a small portion of the growth legends in the February–May 2009 time period but it ends up taking almost 80% of the area in the February–May 2010 period, as shown in Figure 2 (bottom-right). The *strict* CUSUM filtering also suggests deeper investigation in February 2010 because the number of visualized stores is largest (93 stores) while 28 stores are visualized in January 2010. This implies the largest number of stores turned

profit into loss in cumulated sales in February 2010 as shown in Figure 2 (top-right). The default CUSUM filtering also effectively reduces the number of stores, which, for example, visualizes 150 stores on average during October 2009–January 2010 (48% reduction).

In order to find the causes for the loss of competitive advantage, a view of primary interest to the analyst is the matrix view under the comparative mode, as shown in Figure 6. In this view, the analyst sees one of the problematic products. From October 2009, the company started losing competitive advantage in P_{18}. The effect of the loss was not significant at that time. However, during its sharpest decrease between December 2009 and March 2010, the sales of product P_{18} decreased in January–April 2009. Even worse, P_{18} had the worst growth rate in about 40% of the stores severely (MAG_9). This implies a strategic failure that could cause the loss of a big market. On the other hand, the analyst sees additional important information in MAG_8 meaning that the three top-selling products (P_1–P_3) will severely lose competitive advantage in the next month. Through the analysis, the analyst verified how the leading company has been losing its market share due to the decreasing sales performance. The analyst also verified that as a product became less popular in the market, the trend, growth, and proportional legend view reflected the adverse situation while the geographical view highlighted the region where the sales performance is lowest. Although discoveries should be interpreted from various perspectives in management strategies, we believe that the two use-case examples show the effectiveness of the system for competitive advantage analysis using point of sale data.

For our analysis, we collaborated with two groups of pro-

Figure 7: *288x36 resized pixels with assumptions of 63 (left), 126 (middle), and 252 (right) products in 1000 stores.*

fessionally employed analysts. The first group consisted of four analysts that had not previously used automated tools for analyzing CI. Working in a supervised learning environment in our lab, they described that they required tools for report generation and trend discovery. They noted that they usually compare singular data aspects one at a time using conventional tools such as Excel. The second group consisted of three analysts who were working with custom-made tools for CI analysis. Their tools utilized line charts, bar graphs and data aggregation schemes. The analysts in group 2 noted that the support of geographical visualization and the coupling of advanced analytical methods were a marked improvement over their current tools. Both groups utilized MarketAnalyzer and reported on their increased ability to assess sales performance against their competitors. The greatest benefit in the analysis is that analysts can recognize the overall competition status at a glance without tedious selections. This enables time-efficient investigation and easy comparison in various dimensions that can lead to discovery of unexpected trends. The system is also very helpful for people who are responsible for analyzing a competitive advantage in wide areas such as countries. In addition, the system is easy to understand and use. A manager from the company's finance department immediately gained new insight without difficulty by using the system for few minutes.

5. Discussion

In this work, we discussed a relatively small data set using 288 stores and 36 products. In this section we discuss some of the potential limitations of our chosen techniques when scaling to larger datasets. One key scaling issue is the limitation of the pixel-based visualization. Figure 7 illustrates the effects of resizing the matrix. We show results using 63, 126 and 252 products (from left to right) in 1000 stores. Here, we see that we are still able to distinguish between blue and red stores (P_1– P_3 and P_{18}) as well as discrete patterns of light red pixels (left, middle). As the number of products increases, the ability to distinguish between such groupings becomes more difficult. As such, intelligent filter controls are necessary to help show the most important aspects of the data (thus the motivation for the CUSUM and trend filtering). In our current data set, of the 288 stores, approximately 157 stores (std: 4.74) would be removed utilizing the *default* mode, and 229 stores (std: 23.86) would be removed utilizing the *strict* mode. Future work will focus on other importance metrics for guiding the data analysis. Another method for overcoming the scaling issue is to use focus+context interaction methods. Our current system provides a single-level zoom magnification lens. Future itera-

tions of this tool will employ further interaction techniques to better facilitate problems of scale. Finally, the stacked bar chart method employed also becomes less effective as the data size increases. However, the purpose of this particular view is to look at a small combination of products. Furthermore, future iterations will utilize more statistical methods and machine learning algorithms for directing users to particular products of interest. In ARIMA, blue and red are chosen due to their maximum distance in a diverging color scheme [HB03] to contrast with each other.

6. Conclusions and Future Work

We have introduced a new visual analytics tool for market and competitive business analysis incorporating multiple tightly integrated interactive visualizations with integrated trend analytics. Our zoomable and sortable matrix displays presents sales, trends, and growth rates with enhanced pixel-based visualization, while linked line graph views and stacked bar views aid analysis and awareness acquisition of global and specific product/store information. The linked choropleth maps enable geospatial, temporal, and regional competition analysis. We presented two use-case examples using real point of sale data to illustrate the use and potential of the system. Our system can be easily applied to analysis with any other multivariate spatiotemporal data. In the future, we plan to improve the geographical view with advanced selecting and filtering to investigate correlations between point of sale data, demography, and geolocations for more advanced business analysis. To find strength and weakness, a user study comparing to other alternative techniques such as small multiple horizon graphs is needed. We also plan to deploy our system with our corporate partner and start a longitudinal study.

Acknowledgements

References

[Ahl96]   AHLBERG C.: Spotfire: An information exploration environment. *ACM Special Interest Group on Management of Data Record 25*, 4 (1996), 25–29. 2

[BJ76]   BOX G., JENKINS G.: *Time Series Analysis: Forecasting and Control.* Holden–Day Press, San Francisco, U.S.A., 1976. 2

[BPC*10] BORGO R., PROCTOR K., CHEN M., JANICKE H., MURRAY T., THORNTON I.: Evaluating the impact of task demands and block resolution on the effectiveness of pixel-based visualization. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 963–972. 2

[BW08] BYRON L., WATTENBERG M.: Stacked graphs - geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1245–1252. 6

[CGK*07] CHANG R., GHONIEM M., KOSARA R., RIBARSKY W., YANG J., SUMA E., ZIEMKIEWICZ C., KERN D., SUD-JIANTO A.: WireVis: Visualization of categorical, time-varying data from financial transactions. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology* (2007), pp. 155–162. 2

[Cle94] CLEVELAND W. S.: *Visualizing Data*. Hobart Press, New Jersey, U.S.A., 1994. 5

[DS98] DRAPER N. R., SMITH H.: *Applied Regression Analysis*. Wiley-Interscience, New Jersey, U.S.A., 1998. 2

[EM01] EDWARDS R., MAGEE J.: *Technical Analysis of Stock Trends*. Amacom Press, New York, U.S.A., 2001. 2

[HB03] HARROWER M. A., BREWER C. A.: Colorbrewer.org: An online tool for selecting color schemes for maps. *Cartographic Journal 40*, 1 (2003), 27–37. 2, 9

[HHWN02] HAVRE S., HETZLER E., WHITNEY P., NOWELL L.: ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics 8*, 1 (Jan. 2002), 9–20. 6

[HY09] HUR I., YI J. S.: Simulsort: Multivariate data exploration through an enhanced sorting technique. In *Human-Computer Interaction. Novel Interaction Methods and Techniques, 13th International Conference, HCI International 2009, San Diego, CA, USA, July 19-24, 2009, Proceedings, Part II* (2009), vol. 5611 of *Lecture Notes in Computer Science*, Springer, pp. 684–693. 4

[Kah98] KAHANER L.: *Competitive Intelligence: How to Gather Analyze and Use Information to Move Your Business to the Top*. Touchstone Press, New York, U.S.A., 1998. 1

[Kei00] KEIM D. A.: Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics 6*, 1 (2000), 59–78. 1, 2, 3

[KK94] KEIM D. A., KRIEGEL H.: VisDB: Database exploration using multidimensional visualization. *IEEE Computer Graphics and Applications 14*, 5 (1994), 40–49. 1, 2, 3

[MAG99] MANCHANDA P., ANSARI A., GUPTA S.: The shopping basket: A model for multicategory purchase incidence decisions. *Journal of Marketing Science 18*, 2 (1999), 95–114. 6

[MFGH08] MATKOVIC K., FREILER W., GRACANIN D., HAUSER H.: ComVis: A coordinated multiple views system for prototyping new visualization technology. In *Proceedings of 13th International Conference on Information Visualisation* (2008), pp. 215–220. 2

[Mur99] MURPHY J. J.: *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. Prentice Hall Press, New Jersey, U.S.A., 1999. 2

[OJS*11] OELKE D., JANETZKO H., SIMON S., NEUHAUS K., KEIM D. A.: Visual boosting in pixel-based visualizations. *Computer Graphics Forum 30*, 3 (2011), 871–880. 2

[Pag54] PAGE E. S.: Continuous inspection scheme. *Biometrika 41* (1954), 100–115. 2, 5

[R D06] R DEVELOPMENT CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0. URL: http://www.R-project.org. 3

[Rob07] ROBERTS J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In *Proceedings of Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization* (2007), pp. 61–71. 2

[SCB98] SWAYNE D. F., COOK D., BUJA A.: XGobi: Interactive dynamic data visualization in the X window system. *Journal of Computational and Graphical Statistics 7*, 1 (1998), 113–130. 2

[SFOL04] SHIMABUKURO M., FLORES E., OLIVEIRA M., LEVKOWITZ H.: Coordinated views to assist exploration of spatio-temporal data: A case study. In *Proceedings of Second International Conference on Coordinated and Multiple Views in Exploratory Visualization* (2004), pp. 107–117. 2

[SGL08] STASKO J. T., GÖRG C., LIU Z.: Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization 7*, 2 (2008), 118–132. 2

[SLBC03] SWAYNE D. F., LANG D. T., BUJA A., COOK D.: GGobi: Evolving from XGobi into an extensible framework for interactive data visualization. *Journal of Computational Statistics & Data Analysis 43*, 4 (2003), 423–444. 2

[Sma] SmartMoney. http://www.smartmoney.com/marketmap/. Accessed by 12 Aug 2011. 2

[STH08] STOLTE C., TANG D., HANRAHAN P.: Polaris: A system for query, analysis, and visualization of multidimensional databases. *ACM Communications 51*, 11 (2008), 75–84. 2

[STKF07] SCHRECK T., TEKUSOVA T., KOHLHAMMER J., FELLNER D. W.: Trajectory-based visual analysis of large financial time series data. *ACM Special Interest Group on Knowledge Discovery and Data Mining Explorer Newsletter 9*, 2 (2007), 30–37. 2

[TA03] TASKAYA T., AHMAD K.: Bimodal visualisation: A financial trading case study. In *Proceedings of Seventh International Conference on Information Visualization* (2003), pp. 320–326. 2

[Tab] Tableau. http://www.tableausoftware.com, Accessed on Aug 2011. 2

[TLH10] TALBOT J., LIN S., HANRAHAN P.: An extension of Wilkinson's algorithm for positioning tick labels on axes. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1036–1043. 5

[Tuf90] TUFTE E. R.: *Envisioning Information*. Graphics Press, CT, U.S.A., 1990. 3

[VvWvdL06] VLIEGEN R., VAN WIJK J. J., VAN DER LINDEN E.-J.: Visualizing business data with generalized treemaps. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 789–796. 2

[War94] WARD M. O.: XmdvTool: Integrating multiple methods for visualizing multivariate data. In *Proceedings of IEEE conference on Visualization* (1994), pp. 326–333. 2

[WFR*07] WEAVER C., FYFE D., ROBINSON A., HOLDSWORTH D., PEUQUET D., MACEACHREN A. M.: Visual exploration and analysis of historic hotel visits. *Information Visualization 6*, 1 (2007), 89–103. 2

[Wil05] WILKINSON L.: *The Grammar of Graphics (Statistics and Computing)*. Springer-Verlag, New Jersey, U.S.A., 2005. 5

[ZNK08] ZIEGLER H., NIETZSCHMANN T., KEIM D. A.: Visual analytics on the financial market: Pixel-based analysis and comparison of long-term investments. In *Proceedings of International Conference on Information Visualisation* (2008), IEEE Computer Society, pp. 287–295. 2

Novel Technologies for Assessing Dietary Intake: Evaluating the Usability of a Mobile Telephone Food Record Among Adults and Adolescents

Bethany L Daugherty, RD, MS,1 TusaRebecca E Schap, RD, MSc,1 Reynolette Ettienne-Gittens, RD, MS, PhD,2 Fengqing M Zhu, MS, PhD,3 Marc Bosch, MS,3 Edward J Delp, PhD,3 David S Ebert, PhD,3 Deborah A Kerr, PhD,4 and Carol J Boushey, RD, MPH, PhD1,2

1Department of Nutrition, Purdue University, West Lafayatte, IN, United States

2Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, United States

3School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, United States

4School of Public Health, Curtin University, Perth, Australia

Carol J Boushey, Epidemiology Program, University of Hawaii Cancer Center, 1236 Lauhala Street, Honolulu, HI, 96813, United States, Phone: 1 808 564 5915, Fax: 1 808 586 2982, Email: cjboushey@cc.hawaii.edu .

Corresponding author.

Abstract

Background

The development of a mobile telephone food record has the potential to ameliorate much of the burden associated with current methods of dietary assessment. When using the mobile telephone food record, respondents capture an image of their foods and beverages before and after eating. Methods of image analysis and volume estimation allow for automatic identification and volume estimation of foods. To obtain a suitable image, all foods and beverages and a fiducial marker must be included in the image.

Objective

To evaluate a defined set of skills among adolescents and adults when using the mobile telephone food record to capture images and to compare the perceptions and preferences between adults and adolescents regarding their use of the mobile telephone food record.

Methods

We recruited 135 volunteers (78 adolescents, 57 adults) to use the mobile telephone food record for one or two meals under controlled conditions. Volunteers received instruction for using the mobile telephone food record prior to their first meal, captured images of foods and beverages before and after eating, and participated in a feedback session. We used chi-square for comparisons of the set of skills, preferences, and perceptions between the adults and adolescents, and McNemar test for comparisons within the adolescents and adults.

Results

Adults were more likely than adolescents to include all foods and beverages in the before and after images, but both age groups had difficulty including the entire fiducial marker. Compared with adolescents, significantly more adults had to capture more than one image before (38% vs 58%, P = .03) and after (25% vs 50%, P = .008) meal session 1 to obtain a suitable image. Despite being less efficient when using the mobile telephone food record, adults were more likely than adolescents to perceive remembering to capture images as easy (P < .001).

Conclusions

A majority of both age groups were able to follow the defined set of skills; however, adults were less efficient when using the mobile telephone food record. Additional interactive training will likely be necessary for all users to provide extra practice in capturing images before entering a free-living

situation. These results will inform age-specific development of the mobile telephone food record that may translate to a more accurate method of dietary assessment.

Introduction

Dietary intake is an important environmental exposure to consider when evaluating an individual's or population's risk for chronic disease. A link between diet and the development of certain cancers, cardiovascular disease, liver disease, and type 2 diabetes has been established. However, scientific evidence linking diet and genetics to these diseases continues to emerge [1]. The development of genome-wide association studies has led to the identification of genetic variations associated with risk for diseases such as type 2 diabetes [2], atherosclerosis [3], and Crohn disease [4]. Diet and genetics may play a shared role in the etiology of or protection from many diseases. Methodological issues with dietary assessment, however, have limited the ability to identify gene–nutrient interactions.

Dietary assessment is difficult due to the increasing complexity of the food supply and day-to-day variability in a person's diet [5]. Traditional self-report methods of dietary assessment, including the 24-hour dietary recall, food record, and food frequency questionnaire [6], rely on the respondent's memory and ability to estimate portion sizes. Both adults and adolescents tend to underreport total energy intake by as much as 30% [7-12]. Developing diet assessment methods that can be incorporated into the lifestyle of adolescents is especially difficult. Adolescents are in a rapid phase of growth requiring increased energy, eat more frequently, and have more unstructured eating events outside of the home [13]. There is much day-to-day variability in the composition and timing of their eating occasions, leading to forgetfulness and lack of compliance in recording their dietary intake [14]. Adolescents also report becoming irritated with their parents reminding them to complete their food records [15]. Adults, on the other hand, follow a more regular routine than adolescents. Senior adults may have more consistent meal times, while working adults may be more influenced by the demands and characteristics of their working environment. However, all adults may encounter occasions where their more structured routines are disrupted by events that make accurate recording via the current assessment methods more difficult. In addition to being burdensome to the respondent, these methods can be expensive and labor intensive for the researcher. The Genes, Environment and Health Initiative of the National Institutes of Health in the United States is attempting to address many of these shortcomings by supporting the development of novel methods to assess diet and of high-throughput methods to assess genetic profiles in individuals and populations [16].

Researchers have been striving to harness the potential of new digital technologies to improve the effectiveness of their work, and researchers in the field of dietary assessment are no different. The past 10 to 15 years has seen steadily increasing usage of mobile communication devices [17]. Significant advances in the capabilities of these devices have coincided with mobile phones achieving the status of an essential communication tool, so that mobile computing devices, such as mobile telephones with cameras known as smart phones, are now poised to realize their potential as a computing device with specific health applications. Personal digital assistants (PDAs) were the first generation of mobile computers used for data collection [18,19]. However, some of the initial studies using PDAs were not promising [20], as earlier PDAs used technology that lacked user-friendly options, and backlit screens made their content difficult to see. As a result of these limitations, early investigators concluded that the technology was a barrier to collecting accurate information.

However, with the rapid advancement in the capabilities of mobile devices, researchers are now pursuing image-based methods as a way of addressing the limitations of traditional dietary assessment methods [21-23]. The use of mobile applications to assist in the monitoring of diabetes, physical activity, and smoking cessation has previously been discussed in the literature [24-26] and has informed the use of these tools for new diet assessment methods. The development of a mobile telephone food record for adults and adolescents for use in a new, image-based dietary assessment method, partially supported by the Genes, Environment and Health Initiative, was the subject of this study.

The design of the mobile telephone food record has been described previously [27]. For all users, the task of recording images of their food should be relatively quick and easy for it to be acceptable. Briefly, participants would use the mobile telephone food record application to capture images of their foods and beverages before and after eating. Methods of image analysis [28,29] are used to automatically identify the food in the image. With the inclusion of a fiducial marker, an object of known dimension and size, the volume of consumption can be estimated. The information from image analysis and volume estimation can be linked to a nutrient database to compute the energy and nutrients consumed, so this method will not have to rely on the respondent's memory and ability to estimate portion sizes. Additionally, real-time data collection eliminates the need for researchers to enter and code food records. Ideally, the ease of use of mobile telephone food record will result in an accurate dietary assessment tool for both adults and adolescents.

There are challenges related to using smart phones in this new dietary assessment method. For example, for adolescents to use the device, school administrators must accept its use on the school campus, as young people are in school most days of the week. Adults are often less facile than adolescents with using new technology. Therefore, the mobile telephone food record design needs to address these concerns.

Evidence-based development is a crucial step in designing the mobile telephone food record for use by both adults and adolescents [30]. The form of evidence-based development of the mobile telephone food record is an interaction design, which is the discipline of defining the characteristics of products that a user can interact with in their everyday and working lives [30]. The mobile telephone food record design process, when applying interaction design, is an iterative cycle of usability testing in which the user feedback is applied to the next version of the mobile telephone food record, which is tested again [27]. Using this process has allowed the design of the mobile telephone food record to evolve from the perspective of the user or client, resulting in a more positive experience for the user.

The objectives of this study were to evaluate a defined set of user skills for both adults and adolescents—that is, successful image capturing of an eating occasion, while using the mobile telephone food record—and to compare the perceptions and preferences between adults and adolescents regarding their use of the mobile telephone food record. A priori, our hypothesis was that statistically significant differences between adults and adolescents would emerge that would need to be translated into different mobile telephone food record designs to accommodate lifestyles and abilities to use a new technology.

Methods

Study Design and Participant Recruitment

We collected data from two samples of adolescent participants [27] and one sample of adult participants. The data collected from the adolescent samples are combined in this analysis (n = 78). The study methods for all samples were approved by the Purdue University Institutional Review Board. Informed assent and consent were obtained from the adolescent participants and their parents, respectively. The adults completed informed consent prior to participation.

The first adolescent sample was drawn from summer camps for adolescents, ages 11–18 years, taking place on the campus of Purdue University in 2008. A total of 63 participants from these camps used the mobile telephone food record for meal session 1, and 55 (87%) returned for meal session 2 the following day. After using the mobile telephone food record for meal session 1, participants provided feedback and received additional training during the postmeal 1 session. During this session, the participants responded to a series of statements regarding their perceptions of the mobile telephone food record and preferences when using the mobile telephone food record. The advanced interactive instruction included activities in which the participants practiced taking images in potentially problematic snacking scenarios.

The second adolescent sample was a convenience sample drawn from the local community [31]. A total of 15 participants, ages 11–18 years, received all meals and snacks for a 24-hour period while being monitored under controlled conditions. These participants also took part in the feedback and advanced interactive instruction session after using the mobile telephone food record for meal session 1. Data from their first two meal sessions during the 24-hour period are included in this analysis. Figure 1 shows the data collection flow for the two samples of adolescents.

The adult sample was a convenience sample drawn from the campus of Purdue University and the local community during the fall of 2008. A total of 57 participants, ages 21–65 years, used the mobile telephone food record for meal session 1, and 24 (42%) returned for meal session 2 on a subsequent day (Figure 1). During the premeal session, the participants provided feedback regarding their perceptions of the mobile telephone food record, as well as their current use of mobile telephones and digital cameras, by responding to a series of statements and questions. After using the mobile telephone food record in meal session 1, the participants provided additional feedback during the postmeal 1 session, during which they responded to a series of statements regarding their perceptions of the mobile telephone food record and preferences when using the mobile telephone food record.

Meal Sessions

The menus served to the adolescents have been described previously [27]. For the adults, one breakfast menu and four dinner menus were cycled between the sessions. Figure 2 shows examples of meals served to adults and adolescents. Participants received instruction for using the mobile telephone food record during the premeal session (Figure 1). Use of the mobile telephone food record involves recording images of a meal before and after eating. Participants were instructed to include two items in each image: (1) all food and beverage, and (2) the entire fiducial marker (Figure 2). A fiducial marker is an object of known dimensions and markings, which serves as a size reference and must be included in the image [27]. The only instruction provided to participants for placement of the fiducial marker was to avoid placing it near beverages to prevent damage to the object. The meal environment was set up to mimic a restaurant dining atmosphere; however, participants were instructed not to mix or share their foods. The participants took an image of their meal prior to eating, saved the image, took an image of their meal after eating, and saved the image. Participants ate to satiation and, if they requested more, were served a second meal. At three of the four adult dinner meals, dessert was offered as a separate course. The process of capturing images was repeated for these desserts and any additional portions served.

We used HTC p4351 mobile telephones (HTC Corporation, Taoyuan, Taiwan) running Windows Mobile 6.0 Professional (2007; Microsoft Corporation, Redmond, WA, USA). The software, described previously [28,29], guided the user to select the meal occasion and capture an image of foods and beverages. After capturing the image, the user was prompted to review the image and was then given a choice to either retake or save the image. Once the user was satisfied with the image, the mobile telephone prompted the user to eat before proceeding to the next screen. At the next screen, the user was prompted to take an image of the place setting regardless of whether foods and beverages remained. The final screen showed the before and after images prior to exiting the program. If questions arose, the participants were assisted during meals by trained nutrition students. These students also recorded the number of image-capturing attempts before and after meals, as well as the number of images taken by each participant before capturing a satisfactory image and whether participants sat or stood to take images. Participants in the first adolescent sample were compensated US $5 per meal. The second adolescent sample participated for a 24-hour period and they were compensated US $85 for their time. Participants in the adult sample were compensated US $5 for the first meal session and US $15 for the second meal session.

Image Evaluation: Skill Set

To assess the two skills of including all foods and beverages and the entire fiducial marker in the image, the before and after meal images were evaluated for the inclusion of these two required items. When evaluating the inclusion of all foods and beverages, the images were coded as yes if all of the foods and beverages were visible in the image, no if any of the food or beverage was not visible, or software programming error if the image was unavailable due to software malfunction. When evaluating for the inclusion of the entire fiducial marker, the images were coded as yes if the entire fiducial marker was visible in the image, no if a portion of the fiducial marker was cut off, or software programming error. To evaluate the skill of efficiently taking only one image, the number of images taken by each participant before and after meal sessions was coded as one image or greater than one image.

Feedback Session: Perceptions and Preferences

During the feedback sessions, we showed statements regarding possible perceptions of the mobile telephone food record and preferences when using the mobile telephone food record using PowerPoint (Office 2007, PowerPoint 2007; Microsoft Corporation). The participants responded to these statements using a 5-category ordinal response scale (ie, strongly agree, agree, neutral, disagree, and strongly disagree).

We showed the following five statements regarding perceptions to all participants: (1) I think it would be easy to remember to take an image before meals, (2) I think it would be easy to remember to take an

image after meals, (3) I think it would be easy to remember to take an image before snacks, (4) I think it would be easy to remember to take an image after snacks, and (5) the software was easy to use.

We showed the following four statements regarding preferences to all participants: (1) I think it would be easy to carry and use a credit card-sized fiducial marker, (2) I think it would be easy to carry and use a USB-sized fiducial marker (to denote size, this was defined to participants as USB flash drive, USB memory stick, USB jump drive, or USB thumb drive), (3) I prefer to stand while taking an image, and (4) I prefer to sit while taking an image.

The adolescents responded to these nine statements at the start of the postmeal 1 session, followed by the advanced interactive instruction. The adolescents' responses were collected with the eInstruction Classroom Performance System (eInstruction, Cincinnati, OH, USA).

During the premeal session, participants in the adult sample responded to perception statements 1–4. We asked the adults questions to assess their previous experience capturing images with digital cameras and mobile telephones. These were (1) Do you own a digital camera? (2) How often have you taken pictures with a digital camera? and (3) How often have you taken pictures with a mobile telephone?

The response choices for these latter questions were frequently, occasionally, and never or rarely. The adult participants responded to the nine statements above in the postmeal 1 session. The adults recorded their responses on a paper form.

Statistical Analysis

We used data that we collected using the same methods among the adults and the adolescents for statistical comparisons. To further delineate differences by age, we divided the adolescent sample into early and late adolescence: 11–14 years and 15–18 years, respectively. The adult sample was divided into early and middle adulthood: 21–40 years and 41–65 years, respectively. Descriptive analysis included frequencies and percentages. Within both the adolescent and the adult samples, we analyzed differences in age groups and gender using chi-square. McNemar test was used for comparisons of the set of skills for capturing images within the adolescents and within the adults. For those comparisons, each skill (eg, all foods being in image) was classified as yes (demonstrating the skill) or no (not demonstrating the skill). Chi-square was used for comparisons of the skill set between the adults and adolescents; for these comparisons, we grouped no and software programming error together. The 5-category ordinal response scales used by the participants to provide their preferences and perceptions

were recoded as agree, neutral, or disagree. We compared perceptions and preferences between adults and adolescents using chi-square. For comparisons with an expected cell count of less than 5, limiting the comparison to agree and disagree eliminated the inadequate cell counts. We used SPSS 17.0 (IBM Corporation, Somers, NY, USA) for all statistical analyses.

Results

A total of 135 participants (78 adolescents, 57 adults) used the mobile telephone food record for meal session 1, and 94 (70 adolescents, 24 adults) returned to use the mobile telephone food record for meal session 2. The descriptive characteristics of both samples are in Table 1. The average meal duration was 14 minutes for adolescents and 20 minutes for adults. The participants were of diverse ethnic backgrounds. Among the adults, 87% (39/45) claimed to own a digital camera and almost half (22/45, 49%) frequently used it to take pictures. All of the adult participants owned a mobile telephone, but only 16% (7/45) frequently took pictures with their mobile telephone.

Software programming errors occurred when saving the image on the mobile telephone food record, making them unavailable for the analysis. These errors resulted in partial loss of images, either a before or an after image; however, no images were available for only one adult participant, leaving 56 adults for this analysis. Changes to the software were made after testing it with the adolescents, which likely accounted for the reduction in programming errors experienced by the adults. Table 2 shows an evaluation of the participants' ability to follow a defined set of skills when capturing images with the mobile telephone food record. The majority of adults (53/56, 95%) were able to include all foods and beverages in both the before and after images for meal session 1, while 96% (23/24) were able to do the same for meal session 2. A statistically significantly lower proportion of adolescents than adults were able to include all foods and beverages in both the before and after images for meal session 1 (P = .008). This proportion improved to being similar to that of the adults for meal session 2, as Table 2 shows.

For both adults and adolescents, inclusion of the fiducial marker in the image was more problematic than inclusion of all of the foods and beverages (Table 2). There were no significant differences between the adolescents and the adults. Among the adult participants self-selecting desserts, the inclusion of all the dessert and the fiducial marker was very high (Table 2). A significantly higher proportion of adults than of adolescents had to capture more than one image before (P = .03) and after (P = .008) meal session 1 to obtain an image suitable for image analysis (Table 3). This was also the case before and after meal session 2 between adults and adolescents, although this difference was not statistically significant. The adolescents significantly improved their efficiency with capturing suitable images from meal session 1 to meal session 2 (P = .04).

Table 4 shows perceptions of the mobile telephone food record and preferences when using the mobile telephone food record. The majority of both age groups (52/57, 91% of adults; 55/78, 71% of adolescents) agreed that the software was easy to use. Although the adults needed to take more images, they still perceived that capturing images with the mobile telephone food record was easy; however, the proportion was not significantly different from that of the adolescents. Compared with adults, adolescents were less likely to agree that it would be easy to take images before and after meals (P < .001).

Adolescents had a stronger preference than adults for the size of the fiducial marker that they would be willing to use (Table 4). The majority of adolescents (55/71, 77%) and adults (52/57, 91%) reported being willing to use a credit card-sized fiducial marker, but adolescents were less likely to prefer a USB-sized fiducial marker (P = .002). Adolescents reported they would prefer to stand while taking images (P < .001) while adults preferred to sit (P = .002) while taking images. For all analyses regarding skills, preferences, and perceptions, there were no significant differences by gender, early adolescence and middle adolescence, or early adulthood and middle adulthood.

Discussion

This is the first study to systematically evaluate the abilities of adolescents and adults to provide accurate images of an eating occasion. A priori, we assumed that huge differences in skills with technology between adolescents and adults would emerge; however, other than number of images captured, nothing else became obvious. The adolescents were more efficient: they took fewer images than the adults. By the second meal, the adolescents became even more efficient, whereas the adults made insignificant gains. Also, by the second meal, the inclusion of all foods in the images was the same between adults and adolescents, whereas inclusion of the important nonedible item (ie, the fiducial marker) was more problematic for both adolescents and adults. These results support that the fiducial marker was too large. As such, it was difficult to include in images without being partially covered by a plate or utensil. Evaluation of the images for the placement of the fiducial marker revealed that the participants placed the fiducial marker in various locations in the meal setting. Thus, work to reduce the size of the fiducial marker is justified. For both age groups, a notification from the device that the entire fiducial marker is not in the camera's field of view may be helpful in reminding participants to include the entire fiducial marker when capturing images. Clear instruction on the desired placement of the fiducial marker may prevent the participant from spending time deciding where to locate it in the meal setting, which might reduce the burden of this task and translate to better cooperation with this step.

We have also established that the perceptions and preferences of adolescents and adults regarding use of the mobile telephone food record were more disparate than their skill set. In particular, adolescents were less likely than adults to agree that capturing images of meals before and after would be easy.

Adolescents were more opinionated about preferring a credit card-sized fiducial marker. The adolescents may have preferred a credit card-sized fiducial marker because it could be easily carried in a wallet. Finally, adolescents stated a preference to stand while using the mobile telephone food record and adults preferred sitting. This preference for standing is consistent with irregular eating patterns and selecting snacks that are easily portable and often eaten while standing [32,33].

Adolescents are typically the earlier and more eager adopters of new technology [17]. Previous dietary assessment research on adolescents showed that they preferred methods using technology over typical paper or pencil methods [32]. In the current study, adults were noticeably less confident than the adolescents in using this new technology. Whereas the adolescents were eager to use the mobile telephone food record and quickly started taking images, the adults were much more cautious and asked more questions prior to taking an image. This could explain the adults being more likely to include all foods and beverages in the image. However, the adolescents' skills matched the adults' after extensive training, a phenomenon previously documented by Six and colleagues [27]. In all cases, it is impossible to separate the participants' skill in using the mobile telephone food record from motivation to follow the instructions given.

Despite the adolescents being observed as more confident and comfortable when using the mobile telephone food record, they were less likely to agree that it was easy to capture pre and post meal images. This could be a result of differences in daily schedules between the two age groups and may reflect adolescents having more irregular meal times than adults [14,15]. Alternatively, the adolescents may have higher expectations of and demands from the technologies they use [17].Therefore, for adolescents, improvements to the mobile telephone food record might include more reminders throughout the day to ensure that they capture both before and after meal images. For all users, a reminder system, such as an alarm or pop-up message, will likely be needed to remind participants to record their snacks.

Based on the length of time between the before and after images, the average meal duration was shorter for adolescents than for adults. This information provides a basis for programming age-specific software to start timing after the first meal image is captured to initiate a reminder for taking the after image. Next steps include testing the mobile telephone food record with participants in a free-living environment to ascertain the true level of burden, duration of cooperation, and accuracy of recorded energy intake using a biomarker for energy, such as doubly labeled water. There were minimal differences regarding preferences, which will simplify the design process for the mobile telephone food record for adults and adolescents.

Conclusions

The results of these studies will translate to minimal design differences of the mobile telephone food record between adolescents and adults. The majority of both adults and adolescents were able to follow the defined set of skills when capturing before and after images of their meals; however, these results do provide evidence for the need for some age-specific development of the mobile telephone food record, such as reminder programming. The adults were more cautious than the adolescents when taking images and as a result were more likely to include all food and the fiducial marker, which are necessary to capture an image suitable for image analysis. However, adults had to take more images than adolescents before capturing satisfactory ones. Although they were less efficient, the adults perceived that remembering to capture images with the mobile telephone food record would be easy. Additional use of the mobile telephone food record improved adolescents' perceptions and set of skills when capturing images. Additional interactive training will likely be necessary for all users to provide extra practice in taking images before entering a free-living situation. The adolescents had a stronger opinion about the size of the fiducial marker than the adults, suggesting that the fiducial marker design needs to accommodate adolescents over adults. Software improvements between the adolescent and adult meal sessions greatly reduced the number of software programming errors. Some problems will likely never be entirely eliminated due to low battery power and other software-related difficulties, but advances in technology will ensure that these errors will become less frequent.

A more accurate method of dietary assessment will help strengthen the ability of researchers to identify diet–disease and diet–gene relationships. The data generated from a tool such as the mobile telephone food record could be combined with measures of the built environment to inform public policy and assist in the development of nutrition interventions. Further, novel dietary assessment methods will contribute to the growth of mobile applications to enhance self-monitoring for diabetes, weight control, and other diet-related diseases.

Abbreviations

mpFR mobile telephone food record PDA personal digital assistant

Footnotes

Contributed by

All authors contributed equally to this work.

Conflicts of Interest:

None declared.

References

1. Seedorf U, Schulte H, Assmann G. Genes, diet and public health. Genes Nutr. 2007 Oct;2(1):75–80. doi: 10.1007/s12263-007-0001-1. [PMCID: PMC2474917] [PubMed: 18850146]

2. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007 Feb 22;445(7130):881–5. doi: 10.1038/nature05616.nature05616 [PubMed: 17293876]

3. O'Donnell CJ, Cupples LA, D'Agostino RB, Fox CS, Hoffmann U, Hwang SJ, Ingellson E, Liu C, Murabito JM, Polak JF, Wolf PA, Demissie S. Genome-wide association study for subclinical atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study. BMC Med Genet. 2007;8 Suppl 1:S4. doi: 10.1186/1471-2350-8-S1-S4. http://www.biomedcentral.com/1471-2350/8%20Suppl%201/S4.1471-2350-8-S1-S4 [PMCID: PMC1995605] [PubMed: 17903303]

4. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet. 2007 May;39(5):596–604. doi: 10.1038/ng2032.ng2032 [PMCID: PMC2757939] [PubMed: 17435756]

5. Favé G, Beckmann ME, Draper JH, Mathers JC. Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? Genes Nutr. 2009 Jun;4(2):135–41. doi: 10.1007/s12263-009-0120-y. [PMCID: PMC2690728] [PubMed: 19340473]

6. Tucker KL. Assessment of usual dietary intake in population studies of gene-diet interaction. Nutr Metab Cardiovasc Dis. 2007 Feb;17(2):74–81. doi: 10.1016/j.numecd.2006.07.010.S0939-4753(06)00166-9 [PubMed: 17046222]

7. Mahabir S, Baer DJ, Giffen C, Subar A, Campbell W, Hartman TJ, Clevidence B, Albanes D, Taylor PR. Calorie intake misreporting by diet record and food frequency questionnaire compared to doubly labeled water among postmenopausal women. Eur J Clin Nutr. 2006 Apr;60(4):561–5. doi: 10.1038/sj.ejcn.1602359.1602359 [PubMed: 16391574]

8. Blanton CA, Moshfegh AJ, Baer DJ, Kretsch MJ. The USDA Automated Multiple-Pass Method accurately estimates group total energy and nutrient intake. J Nutr. 2006 Oct;136(10):2594–9. http://jn.nutrition.org/cgi/pmidlookup?view=long&pmid=16988132.136/10/2594 [PubMed: 16988132]

9. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, Sharbaugh CO, Trabulsi J, Runswick S, Ballard-Barbash R, Sunshine J, Schatzkin A. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol. 2003 Jul 1;158(1):1–13. http://aje.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=12835280. [PubMed: 12835280]

10. Champagne CM, Bray GA, Kurtz AA, Monteiro JB, Tucker E, Volaufova J, Delany JP. Energy intake and energy expenditure: a controlled study comparing dietitians and non-dietitians. J Am Diet Assoc. 2002 Oct;102(10):1428–32. [PubMed: 12396160]

11. Champagne CM, Baker NB, DeLany JP, Harsha DW, Bray GA. Assessment of energy intake underreporting by doubly labeled water and observations on reported nutrient intakes in children. J Am Diet Assoc. 1998 Apr;98(4):426–33. doi: 10.1016/S0002-8223(98)00097-2.S0002-8223(98)00097-2 [PubMed: 9550166]

12. Bandini LG, Must A, Cyr H, Anderson SE, Spadano JL, Dietz WH. Longitudinal changes in the accuracy of reported energy intake in girls 10-15 y of age. Am J Clin Nutr. 2003 Sep;78(3):480–4. http://www.ajcn.org/cgi/pmidlookup?view=long&pmid=12936932. [PubMed: 12936932]

13. Neumark-Sztainer D, Story M, Perry C, Casey MA. Factors influencing food choices of adolescents: findings from focus-group discussions with adolescents. J Am Diet Assoc. 1999 Aug;99(8):929–37. doi: 10.1016/S0002-8223(99)00222-9.S0002-8223(99)00222-9 [PubMed: 10450307]

14. Livingstone MB, Robson PJ, Wallace JM. Issues in dietary intake assessment of children and adolescents. Br J Nutr. 2004 Oct;92 Suppl 2:S213–22.S0007114504002326 [PubMed: 15522159]

15. Goodwin RA, Brulé D, Junkins EA, Dubois S, Beer-Borst S. Development of a food and activity record and a portion-size model booklet for use by 6- to 17-year olds: a review of focus-group testing. J Am Diet Assoc. 2001 Aug;101(8):926–8. doi: 10.1016/S0002-8223(01)00229-2.S0002-8223(01)00229-2 [PubMed: 11501871]

16. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc. 2010 Jan;110(1):48–51. doi: 10.1016/j.jada.2009.10.008.S0002-8223(09)01684-8 [PMCID: PMC2823476] [PubMed: 20102826]

17. Lenhart A, Ling R, Campbell S, Purcell K. Teens and mobile phones. Washington, DC: Pew Internet & American Life Project; 2010. Apr 20, [2012-04-05]. webcite http://www.pewinternet.org/~/media//Files/Reports/2010/PIP-Teens-and-Mobile-2010-with-topline.pdf.

18. Wang DH, Kogashiwa M, Ohta S, Kira S. Validity and reliability of a dietary assessment method: the application of a digital camera with a mobile phone card attachment. J Nutr Sci Vitaminol (Tokyo) 2002 Dec;48(6):498–504. [PubMed: 12775117]

19. Kretsch MJ, Blanton CA, Baer D, Saples R, Horn WF, Keim NL. Measuring energy expenditure with simple low-cost tools. J Am Diet Assoc. 2004 Aug;104(Suppl 2):A-13.

20. Yon BA, Johnson RK, Harvey-Berino J, Gold BC. The use of a personal digital assistant for dietary self-monitoring does not improve the validity of self-reports of energy intake. J Am Diet Assoc. 2006 Aug;106(8):1256–9. doi: 10.1016/j.jada.2006.05.004.S0002-8223(06)00880-7 [PubMed: 16863723]

21. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. Br J Nutr. 2009 Feb;101(3):446–56. doi: 10.1017/S0007114508027438.S0007114508027438 [PMCID: PMC2626133] [PubMed: 18616837]

22. Sun M, Fernstrom JD, Jia W, Hackworth SA, Yao N, Li Y, Li C, Fernstrom MH, Sclabassi RJ. A wearable electronic system for objective dietary assessment. J Am Diet Assoc. 2010 Jan;110(1):45–7. doi: 10.1016/j.jada.2009.10.013.S0002-8223(09)01689-7 [PMCID: PMC2813220] [PubMed: 20102825]

23. Weiss R, Stumbo PJ, Divakaran A. Automatic food documentation and volume computation using digital imaging and electronic transmission. J Am Diet Assoc. 2010 Jan;110(1):42–4. doi: 10.1016/j.jada.2009.10.011.S0002-8223(09)01687-3 [PMCID: PMC2813222] [PubMed: 20102824]

24. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res. 2011;13(3):e65. doi: 10.2196/jmir.1874. http://www.jmir.org/2011/3/e65/v13i3e65 [PMCID: PMC3222161] [PubMed: 21979293]

25. Bexelius C, Löf M, Sandin S, Trolle Lagerros Y, Forsum E, Litton JE. Measures of physical activity using cell phones: validation using criterion methods. J Med Internet Res. 2010 Jan;12(1):e2. doi: 10.2196/jmir.1298. http://www.jmir.org/2010/1/e2/v12i1e2 [PMCID: PMC2821583] [PubMed: 20118036]

26. Whittaker R, Maddison R, McRobbie H, Bullen C, Denny S, Dorey E, Ellis-Pegler M, van Rooyen J, Rodgers A. A multimedia mobile phone-based youth smoking cessation intervention: findings from

content development and piloting studies. J Med Internet Res. 2008 Nov;10(5):e49. doi: 10.2196/jmir.1007. http://www.jmir.org/2008/5/e49/v10i5e49 [PMCID: PMC2630843] [PubMed: 19033148]

27. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. Evidence-based development of a mobile telephone food record. J Am Diet Assoc. 2010 Jan;110(1):74–9. doi: 10.1016/j.jada.2009.10.010.S0002-8223(09)01686-1 [PMCID: PMC3042797] [PubMed: 20102830]

28. Zhu F, Mariappan A, Boushey CJ, Kerr D, Lutes KD, Ebert DS, Delp EJ. Technology-Assisted Dietary Assessment. Proc SPIE. 2008 Mar 20;6814:681411. doi: 10.1117/12.778616. [PMCID: PMC3224859] [PubMed: 22128303]

29. Mariappan A, Bosch M, Zhu F, Boushey CJ, Kerr DA, Ebert DS, Delp EJ. Personal dietary assessment using mobile devices. Proc SPIE. 2009 Jan 1;7246 doi: 10.1117/12.813556. [PMCID: PMC3109913]

30. Hertzum M, Simonsen J. Evidence-based development: a viable approach?. Proceedings; Third Nordic Conference on Human-Computer Interaction; Oct 23-27, 2004; Tampere, Finland. New York, NY: ACM Press; 2004.

31. Schap TE, Six BL, Delp EJ, Ebert DS, Kerr DA, Boushey CJ. Adolescents in the United States can identify familiar foods at the time of consumption and when prompted with an image 14 h postprandial, but poorly estimate portions. Public Health Nutr. 2011 Jul;14(7):1184–91. doi: 10.1017/S1368980010003794.S1368980010003794 [PMCID: PMC3114201] [PubMed: 21324224]

32. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. Eur J Clin Nutr. 2009 Feb;63 Suppl 1:S50–7. doi: 10.1038/ejcn.2008.65.ejcn200865 [PMCID: PMC2830089] [PubMed: 19190645]

33. Laska MN, Graham D, Moe SG, Lytle L, Fulkerson J. Situational characteristics of young adults' eating occasions: a real-time data collection using Personal Digital Assistants. Public Health Nutr. 2010 Dec 8;:1–8. doi: 10.1017/S1368980010003186.S1368980010003186 [PMCID: PMC3516625] [PubMed: 21138611]

Figures and Tables

Figure 1

Study design, activities, and measures of participants using the mobile telephone food record. For 15 of the adolescent participants, meal session 2 was later in the same day. For the remainder of participants, meal session 2 occurred on a different day. Adult participants were offered dessert as a separate course. For meal session 1, 39 selected dessert, and for meal session 2, 15 selected dessert.

Figure 2

Images that demonstrate meeting two skills required for using the mobile telephone food record: included in the image are all foods and beverages and the entire fiducial marker (checkerboard square).

Table 1

Characteristics of adults and adolescents testing the usability of the mobile telephone food record.

Characteristic	Adolescents (n = 78), n (%)	Adults (n = 57) n (%)
Gender		
Male	26 (33%)	18 (32%)
Female	52 (67%)	39 (68%)
Age group (years)		
11–14	45 (58%)	NAa
15–18	33 (42%)	NA
21–40	NA	27 (47%)
41–65	NA	30 (53%)
Ethnic group		
Asian	1 (1%)	4 (7%)
Hispanic	7 (9%)	0 (0%)
Non-Hispanic white	55 (70%)	45 (79%)
Black/African American	10 (13%)	2 (4%)

Multiple 5 (6%) 6 (11%)

View it in a separate window

a Not applicable.

Table 2

Evaluation of participants' set of skills when capturing images with the mobile telephone food record.

Skill	Adolescents (n = 78)			Adults (n = 56)a		
	Yes, n (%)	No, n (%)	Software errorb, n (%)	Yes, n (%)	No, n (%)	Software error, n (%)
All foods and beverages included in image						
Meal session 1c	61 (78%)	7 (9%)	10 (13%)	53 (95%)	0 (0%)	3 (5%)
Meal session 2	59 (84%)	9 (13%)	2 (3%)	23 (96%)	0 (0%)	1 (4%)
Dessert session 1d	NAe	NA	NA	39 (100%)	0 (0%)	0 (0%)
Dessert session 2d	NA	NA	NA	14 (93%)	0 (0%)	1 (7%)
Entire fiducial marker included in image						
Meal session 1	54 (69%)	14 (18%)	10 (13%)	44 (79%)	9 (16%)	3 (5%)

Meal session 2 53 (76%) 15 (21%) 2 (3%) 18 (75%) 5 (21%) 1 (4%)

Dessert session 1d NA NA NA 37 (95%) 2 (5%) 0 (0%)

Dessert session 2d NA NA NA 11 (73%) 3 (20%) 1 (7%)

View it in a separate window

a Due to software programming error, n = 56 instead of 57.

b Paired images unavailable due to software programming errors.

c P = .008 using chi-square and comparing adolescents versus adults.

d Dessert was served as a separate course for adult participants. For meal session 1, 39 selected dessert, and for meal session 2, 15 selected dessert.

e Not applicable.

Table 3

Comparisons between and within adolescents and adults of the number of images acquired prior to obtaining a suitable image.

Group Adolescents
(n = 63 meal session 1,

n = 55 meal session 2) Adults

(n = 56a meal session 1,

n = 24 meal session 2)

1 image,

1 image, n (%) >1 image, n (%) Data recording errorb, n 1 image, n (%) >1 image, n (%) Data recording errorb, n

	1 image, n (%)	>1 image, n (%)	Data recording errorb, n	1 image, n (%)	>1 image, n (%)	Data recording errorb, n
All participants						
Meal session 1						
Before imagec,d	38 (62%)	23 (38%)	2	21 (42%)	29 (58%)	6
After imagec,e	44 (75%)	15 (25%)	4	25 (50%)	25 (50%)	6
Meal session 2						
Before image	39 (77%)	12 (24%)	4	13 (59%)	9 (41%)	2
After image	40 (78%)	11 (22%)	4	16 (73%)	6 (27%)	0
Matched participantsf						
Meal session 1						
Before image	28 (58%)	20 (42%)g,h	NAi	9 (45%)	11 (55%)	NA
After imagec,j	36 (75%)	12 (25%)	NA	7 (35%)	13 (65%)	NA
Meal session 2						
Before image	38 (79%)	10 (21%)g,h	NA	12 (60%)	8 (40%)	NA
After image	37 (77%)	11 (23%)	NA	14 (70%)	6 (30%)	NA

View it in a separate window

a Due to software programming errors, n = 56 instead of 57.

b Data recording error on the part of staff; therefore, numbers not included in percentages, which represent only users' abilities.

c Comparison between adolescents and adults.

d P = .03.

e P = .008.

f Number of before and after meal images these participants took was recorded for both meal session 1 and meal session 2 (n = 48 session pairs for adolescents; n = 20 session pairs for adults).

g Comparison between meal session 1 (before) and meal session 2 (before) within adolescents.

h P = .04

i Not applicable.

j P = .002.

Table 4

Comparison of perceptions and preferences between adolescents and adults regarding use of the mobile telephone food recorda.

Perceptions and preferences Adolescents (n = 78)b Adults (n = 57)

Agree,

n (%) Neutral,

n (%) Disagree,

	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)
Perceptions						
The software was easy to use	55 (71%)	9 (13%)	6 (9%)	52 (91%)	1 (2%)	4 (7%)
I think it would be easy to remember to take an image before meals[c]	26 (37%)	22 (31%)	22 (31%)	47 (83%)	5 (9%)	5 (9%)
I think it would be easy to remember to take an image after meals[c]	29 (41%)	27 (38%)	15 (21%)	42 (74%)	8 (14%)	7 (12%)
I think it would be easy to remember to take an image before snacks	8 (11%)	16 (23%)	46 (66%)	15 (26%)	12 (21%)	30 (53%)
I think it would be easy to remember to take an image after snacks	15 (21%)	19 (27%)	37 (52%)	19 (33%)	13 (23%)	25 (44%)
Preferences						
I think it would be easy to carry and use a credit card-sized fiducial marker	55 (77%)	10 (14%)	6 (8%)	52 (91%)	4 (7%)	1 (2%)
I think it would be easy to carry and use a USB-sized fiducial marker[d]	30 (42%)	19 (27%)	22 (31%)	38 (67%)	15 (26%)	4 (7%)
I prefer to stand while taking an image[c]	43 (63%)	14 (21%)	11 (16%)	13 (23%)	12 (21%)	32 (56%)
I prefer to sit while taking an image[d]	25 (36%)	21 (30%)	23 (33%)	39 (68%)	8 (14%)	10 (18%)

View it in a separate window

a Percentages do not add to 100% due to rounding.

b Missing values due to a malfunction of the eInstruction Classroom Performance System.

c P < .001 using chi-square and comparing adolescents versus adults.

d P = .002 using chi-square and comparing adolescents versus adults.

**BMC
Bioinformatics**

RESEARCH

Open Access

OmicsVis: an interactive tool for visually analyzing metabolomics data

Philip Livengood[1], Ross Maciejewski[2*], Wei Chen[3], David S Ebert[1]

Abstract

When analyzing metabolomics data, cancer care researchers are searching for differences between known healthy samples and unhealthy samples. By analyzing and understanding these differences, researchers hope to identify cancer biomarkers. Due to the size and complexity of the data produced, however, analysis can still be very slow and time consuming. This is further complicated by the fact that datasets obtained will exhibit incidental differences in intensity and retention time, not related to actual chemical differences in the samples being evaluated. Additionally, automated tools to correct these errors do not always produce reliable results. This work presents a new analytics system that enables interactive comparative visualization and analytics of metabolomics data obtained by two-dimensional gas chromatography-mass spectrometry (GC × GC-MS). The key features of this system are the ability to produce visualizations of multiple GC × GC-MS data sets, and to explore those data sets interactively, allowing a user to discover differences and features in real time. The system provides statistical support in the form of difference, standard deviation, and kernel density estimation calculations to aid users in identifying meaningful differences between samples. These are combined with novel transfer functions and multiform, linked visualizations in order to provide researchers with a powerful new tool for GC × GC-MS exploration and bio-marker discovery.

Introduction

In recent years, GCxGC-MS has become an invaluable laboratory analysis tool. However, this procedure produces large (gigabytes of data per sample), four dimensional datasets (retention time one, retention time two, mass and intensity). Such data is cumbersome, and researchers must spend time formatting and processing the data in order to remove acquisition artifacts, and quantify and identify chemical compounds [1]. Furthermore, while statistical analysis has played an important role in this work (because of the need to reduce the thousands of acquired spectral features to a more manageable size), the large data size, inherent biological variability and measurement noise makes the identification of bio-markers through purely statistical processes extremely difficult and time consuming.

This work presents a visual analytics environment for analysis and visualization of metabolomic data obtained by GCxGC-MS. The system is in the form of a comparative visualization tool kit that allows users to quickly perform comparisons across and between sample sets, in order to enable advanced analytic exploration of GCxGC-MS data sets. The guiding principle is to utilize the ability of the visual perception system to cluster, detect outliers, and find patterns much more efficiently than the most advanced computerized algorithms. By balancing the statistical analysis with advanced visualization techniques in order to analyze and understand differences between samples, OMIC researchers are able to target analysis and re ne algorithms in order to begin to identify cancer biomarkers. These biomarkers are the key to identifying differences in susceptibility and treatment response for patient groups.

In this paper, we describe a new, efficient visual analytics suite of tools for GC × GC-MS data sets to be used in metabolomics for a variety of applications including

* Correspondence: rmacieje@asu.edu
[2]School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA
Full list of author information is available at the end of the article

cancer signatures detection. Our work is being developed in collaboration with analytical chemists and biology researchers, and is designed to provide an interactive, integrated visual and statistical analysis environment with multiple functionalities. Compared to our previous work [2] that processes a GC × GC-TOF dataset as a 2D (TIC) image, our expanded system presents and handles mass spectral data as three-dimensional distribution data. We employ volume visualization to depict and compare spectral distributions in a portion or a set of data, showing the individual spectral distribution for each pair of retention times, instead of the integrated sum. We perform comparative visualization on selected regions and potential bio-markers for identifying distinctive metabolic differences among a set of samples. System features include:

1. A comparative visualization system that allows multiple samples (and multiple views of individual samples) to be displayed and explored simultaneously,
2. Data exploration tools for exploring mass spectra and filtering and comparing TIC images in real-time,
3. Grouping and rendering of samples and calculation of group means for comparison and difference calculation,
4. Comparative volume rendering across samples.

Related work

In order to develop a system that is both effective and efficient at finding differences across samples, this system builds on previous work from several areas. In particular, previous work from GCxGC-MS analysis and data visualization were examined. With regards to visualization, areas of relevance include focusing and linking techniques, comparative visualization, multi-dimensional visualization, user interface techniques and design, and the use of color as it relates to data and difference visualization.

Two-dimensional gas chromatography-mass spectrometry

A traditional gas chromatograph consists of a carrier gas, injector port, column, detector and a recorder. The injector is set to be hotter than the boiling point of the sample so that the sample will be vaporized. The carrier gas flows through the system and pushes the gaseous components of the sample into the column. Generally, an inert gas, such as helium, is used. As the carrier/sample mixture moves through the column, the different components of the sample interact with the column and these interactions cause the samples to pass through the column at different rates. This chemical separation is used to classify the compounds within the sample, as each compound will exhibit a characteristic retention time. A detector resides at the end of the column, quantizing the output. The most common type of detector is a mass spectrometer that ionizes the gas as it elutes from the column, producing additional

separation in the mass spectral domain [3]. This type of setup is referred to as Gas Chromatography-Mass Spectrometry, or GC-MS. GC-MS has proven to be very effective for simple mixtures. However, as the complexity of the samples increases, it becomes difficult to achieve adequate chemical separation. Different components will exit the column at the same or nearly the same time, a problem known as coelution. This results in partial or full overlap of peaks in the mass spectral domain. To help overcome this problem, two-dimensional gas chromatography-mass spectrometry (GCxGC-MS) has become a common chemical analysis tool.

There are two columns in the two-dimensional gas chromatography-mass spectrometry. Rather than going to a detector, the output of the first column goes into a modulator that collects the sample and periodically injects it into a second column with different chemical properties. Due to the different properties of this second column, compounds will have a different retention time in the second column, resulting in two levels of chemical separation. The detector is positioned at the output of the second column. This process results in a four-dimensional dataset with two time axes (retention time one (RT1) from the first column and retention time two (RT2) from the second column), mass, and intensity. For each RT1/RT2 coordinate pair, the mass and intensity form a mass spectrum.

An understanding of the structure, strengths, and shortcomings of GCxGC-MS data is essential in order to effectively apply visualization techniques. GCxGC-MS is known for its large peak capacity, an order of magnitude increase in chemical separation ability compared to GC-MS [4], and its improved speed [3]. However, GCxGC-MS still has several complexities that make it a challenging process to use effectively. Inconsistencies in sample amounts will lead to differences in peak intensities. Peak retention times and shapes experience slight differences that are uncontrollable, but are not related to actual chemical differences in the samples [4]. Data samples exhibit background noise that can vary from sample to sample, and make accurate peak quantification difficult. Finally, despite the improved separation ability over one-dimensional GC-MS, GCxGC-MS still exhibits some peak overlap due to coelution.

Algorithms that attempt to correct these deficiencies have been developed with promising results, but this is still an active area of research. To correct for differences in sample amounts, normalization can be performed. As GCxGC intensities are fairly linear with respect to sample amounts, a multiplicative scaling factor can be used to normalize samples as long as at least one consistent peak can be found to normalize against [4]. While fairly straightforward, background levels and peak co-elution can hinder the accuracy.

Several peak alignment techniques have been developed to correct the problem of peak shift in chromatography, such as the piecewise automated beam search algorithm developed by Yao et al. [5], or the registration technique described by Hollingsworth et al. [4]. Reichenbach et al. [6] found that the background level was mainly the sum of two slowly changing functions: a steady-state standing-current offset and a temperature-induced column bleed. They found that this background level varies slowly compared to the characteristic peak widths, and has the statistical characteristics of random white noise. With this knowledge they were able to develop an efficient algorithm for removing the background from a total ion count image, resulting in a near-zero mean background level.

Very often one would like to identify, or search for, a particular analyte of interest within a GCxGC-MS sample. These identification algorithms generally involve and correlation with samples from a known mass spectral library, such as in [7]. These algorithms, however, are complicated by coelution. In order to remove the overlap in the mass spectral domain that results when multiple components elute at the same time, spectral peak deconvolution algorithms have been developed, such as [8]. In general, however, these algorithms are dependent on there being at least some separation between the centers of the overlapping peaks. Due to such issues, the visual inspection and explorations of samples plays a vital role in the analysis process.

GCxGC-MS visualization techniques

Several visualization techniques have been developed for exploring GCxGC-MS data and are used almost universally. Typical methods include: two-dimensional total ion count rendering [3,4], two-dimensional contour plots, three-dimensional height rendering, or displaying the data in tabular form [3]. Additionally, subsets of the data can be displayed, such as mass-spectrum views for a given retention time coordinate pair, or visualizing a single mass or range of masses using any of the previously mentioned methods. The scientific community has recently begun taking advantage of modern graphics hardware for more detailed, complex and realistic visualizations for all sorts of complex datasets, and mass spectrometry is no exception. Corral et al. [9] develop some basic hardware-accelerated rendering techniques for very large Liquid Chromatography-Mass Spectrometry (LCMS) datasets, adapting terrain-rendering techniques to suit the LCMS data. This height rendering technique is also used by Linsen et al. for differential protein expression analysis [10]. Their work also includes data resampling, as well interactive modification of the color scheme and material properties. Other research com-

bines mathematical and statistical methods with visualization in order to provide more insight into data. Recent work by Wiklund et al. [11] makes use of an orthogonal partial least-square (OPLS) model-based approach for definition of statistically and potentially biochemically significant compounds. Wiklund et al. present the S-plot, visualizing both the covariance and correlation between metabolites and class designation, and an extension dubbed the SUS-plot (shared and unique structure) to compare the outcome of multiple classification models compared to a common reference.

However, little research has gone into interactive comparative visualization techniques for this type of data. Hollingsworth et al. utilized image processing based techniques based on the total ion count image [4]. This basically amounts to a flattening of the data from four dimensions to three by aggregating the individual intensities at each retention time coordinate in order to obtain the total ion count. Prior to visualization and comparison each image must undergo background removal and peak detection. Peak alignment is performed on the reference image in order to remove incidental differences with the analyzed image. Finally, normalization is performed so that differences in sample amounts will not result in false positives. Once this pre-processing is complete, a difference image is computed. This difference can be visualized using tabular data, a 2D image, or 3D height field visualization in either grayscale or color. Their main contribution, however, is the calculation of a fuzzy difference that compares each pixel value in one image with a small neighborhood of pixels in the other, rather than doing a pixel-by-pixel comparison. This technique helps to reduce the incidental differences in peak shape and retention time that may still exist even after alignment. While these techniques represent a good first step, they are ill-suited for biomarker detection as they rely solely on images produced from the total ion count image. Additionally, since only the magnitude of the difference is considered, the results obtained can be misleading. Large, yet statistically insignificant differences may be emphasized while small, yet significant differences may not be noticeable.

Visual analysis system

We present and handle mass spectral data as both two and three-dimensional distributions through the use of novel interactive techniques. We allow users to explore their data using coordinated multiple views as shown in our previous work [2]. Furthermore, we have also expanded these tools to include dynamic filtering and selection, as well as three dimensional volume rendering of the data. In this section, we illustrate various components of our visual analysis system.

Total Ion Count (TIC) visualization

The first component that we show is the TIC visualization. The total ion count is a reduction of the data set from four dimensions down to three. For each retention time coordinate, the intensities of the entire mass spectrum at that point are summed together to obtain the total intensity, or total ion count. Once the TIC data has been computed, it can be visualized either in 2D, or as a 3D height rendering.

2D TIC

The two-dimensional total ion count image is one of the most common visualization techniques for GCxGC-MS data. Ion count values are mapped to a color using the specified transfer function and rendered to the screen as a flat, two-dimensional image, as in Figure 1.

3D TIC

A total ion count height rendering is nearly identical to the two-dimensional total ion count image. For the height rendering, the intensity is used as the z-coordinate in a polygonal mesh, and can be scaled linearly or logarithmically to fit within a reasonable dimension. The x and y coordinates are evenly spaced points corresponding to retention time 1 and retention time 2.

When used in conjunction with a normal color mapping, this does not actually convey any more information than a two-dimensional TIC image (Figure 2, left). However, this is still a useful technique as data can often be portrayed more effectively by mapping the data values to multiple display parameters, in this case color and height. Not only do the two parameters serve to reinforce each other, but one may overcome deficiencies in the other.

Additionally, as a new application, the height field can be used with alternative color mapping schemes, similar to work done by Linsen et al. with LC-MS data [12]. In this case, the height of the peaks is a useful method for communicating peak intensity compared with the color mapped attribute. As an alternative to color mapping, we also provide a 'high contrast' rendering option for the height field. For this technique we enable OpenGL lighting and create a single light source positioned along the positive z-axis with ambient, diffuse, and specular components. We apply diffuse, specular, and shininess material properties to the polygons. Vertex normals are calculated at each vertex in the mesh, corresponding to each RT1/RT2 coordinate. The end result is a high-gloss, metallic looking rendering with high contrast. With this technique, even small peaks are highlighted and readily noticeable, as seen in Figure 2 (right). Background noise is also highly visible in this view, as it produces a large number of small peaks and valleys.

TIC color maps

The total ion count visualizations both support color mapping based on intensity, difference, and standard deviations away from a mean. Each of these methods can be configured to use either a continuous or discrete color scheme. For the continuous color scheme, the system uses a set of three curves that allow independent control of the hue, saturation, and brightness.

For the discrete color scheme, the system presents the user with a histogram that displays bin colors and data distribution. The user can modify the number of bins, data range for the bins, and bin colors interactively. The color for each bin can be specified by the user, or the system can automatically generate a color mapping. In each case, values are initially mapped to a logarithmic scale where a larger color range is used to represent small intensities, and the scale of large peaks is greatly reduced.

We provide three different types of visualization modes for the TIC color maps: intensity, difference and standard deviation.

Intensity

The intensity mode is a simple mapping of the peak intensity to color. This technique is useful for providing a high-level view of the data. It reveals the location and relative intensity of peaks, and can be useful in helping a user identify any samples that may contain data collection errors. An example of the intensity mode mapping is shown in Figure 3 (left).

Difference

The difference mode calculates the difference between two samples. The result is then displayed by simply mapping the difference to a color. The system uses a separate set of HSV curves for positive and negative differences. By default, the hue for negative differences is set to pure green, and the hue for positive differences is set to pure red. An example of the difference mode is shown in Figure 3 (middle). Here, the user can quickly find areas of high positive or negative differences between two samples.

Standard deviation

The standard deviation mode also calculates the difference between samples and renders an image based on that difference. However, a color mapping based solely on the magnitude of the difference in intensities may not be what is most interesting. Even using the mean of two sets of samples, the difference in intensity between two large peaks could be relatively high in magnitude compared to two smaller peaks, but this does not necessarily mean that difference is meaningful. By analyzing the standard deviation within user specified groups of samples, differences can be visualized in more certain terms. We create a standard deviation color mapping from a sample group by first calculating a mean TIC for all the samples within that group. Note that the samples are chosen by the user such that they have been pre-normalized as input to the system.

Figure 1 GCxGC-MS data sets. GCxGC-MS data sets rendered as 2D TIC images (top). Our system provides visual analytic tools to help researchers identify meaningful differences (bottom). Here we show a canine cancer sample (left) compared with a healthy canine sample (right).
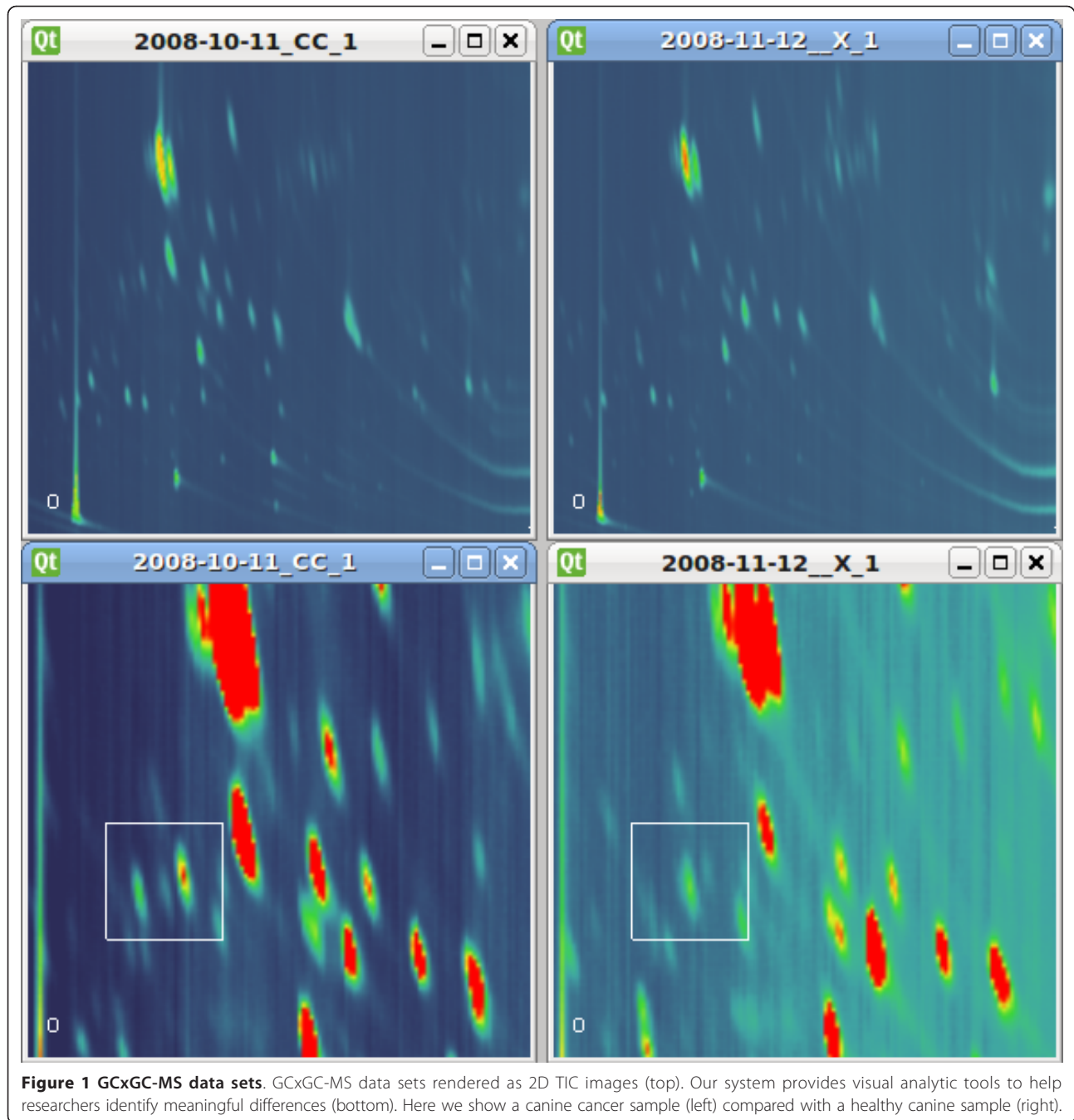
Next a standard deviation TIC is calculated as:

$$\sigma_A = \sqrt{\frac{1}{n_A} \sum_{i=1}^{n} (x_i - \mu_A)^2} \tag{1}$$

Here, n_A is the number of samples in group A, μ_A is the mean TIC of group A, and the x_i are the TICs of the ith sample. Once we have computed the standard deviation TIC, it is stored to use for color mapping.

This color mapping can then be applied to a sample visualization. Generally, it would be applied to a sample that is part of another group. In order to determine the color at a particular retention time coordinate, the system calculates the corresponding z-value for each point b in the new sample, as shown in Equation 2. The z-value is simply how many standard deviations different a value is than the calculated mean. We then use that difference to determine the appropriate color.
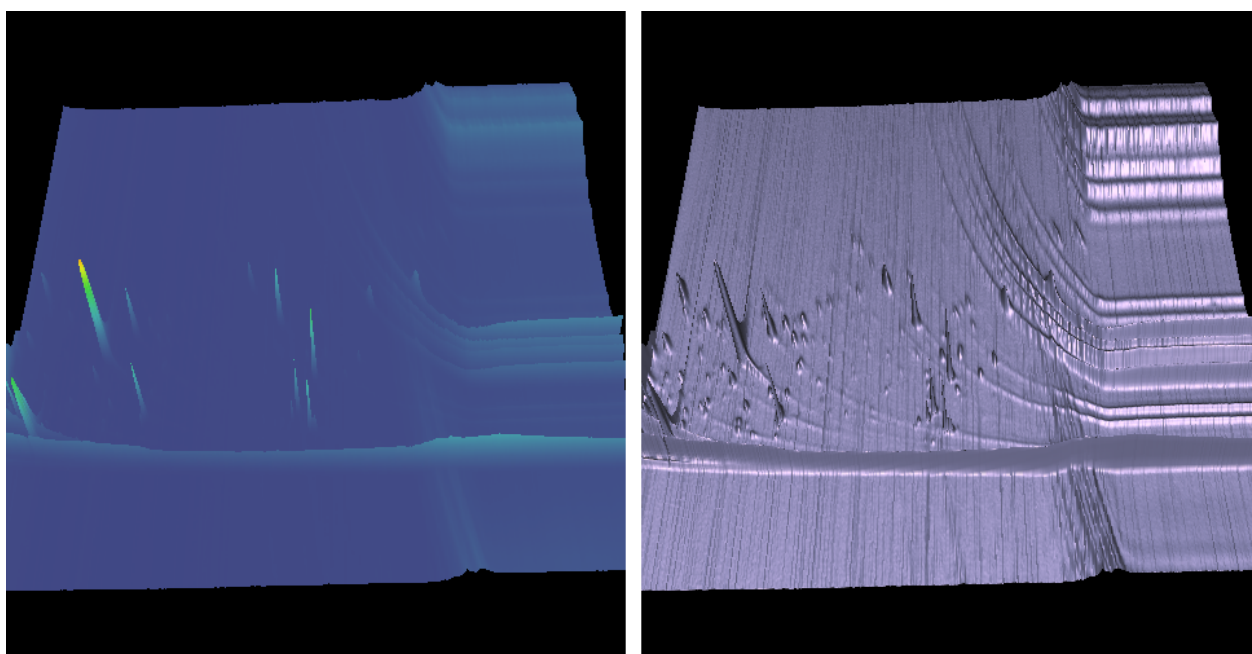
Figure 2 A color Mapped Height Field. A color mapped height field (left). Using the high contrast option for height field rendering (right), small peaks and background noise, which might otherwise be hidden, are easily seen.

$$z = \frac{\mu_A - b}{\sigma_A} \qquad (2)$$

This can help a user to determine whether an observed difference is truly meaningful. Additionally, this technique may effectively reveal areas of difference in smaller peaks that are significant in terms of standard deviations, but were not previously noticed simply because the peaks themselves are smaller. An example is shown in Figure 3 (right), note the green streaks that are not seen in Figure 3 (middle). This helps the user explore regions in the image that are statistically different in a sample when compared to a group of samples.

Mass spectrum visualization

A mass spectrum view is simply a plot of intensity on the y-axis vs. the mass-to-charge ratio on the x-axis, as seen in Figure 4 (bottom). The user is given the option to plot this spectrum as a bar graph, or as a connected line. In order to help a user form a hypothesis about particular mass values that constitute an observed difference, this system provides a novel technique for visualizing the mass spectral difference between two samples. The technique is similar to rendering a normal mass spectrum. In this case, the zero intensity baseline is drawn across the middle of the viewing window, with bars for positive differences rising upward, and bars for

Figure 3 Color mapping examples. Examples of the different color mappings that can be applied to total ion count images. Intensity (left), difference (middle), and standard deviations (right).
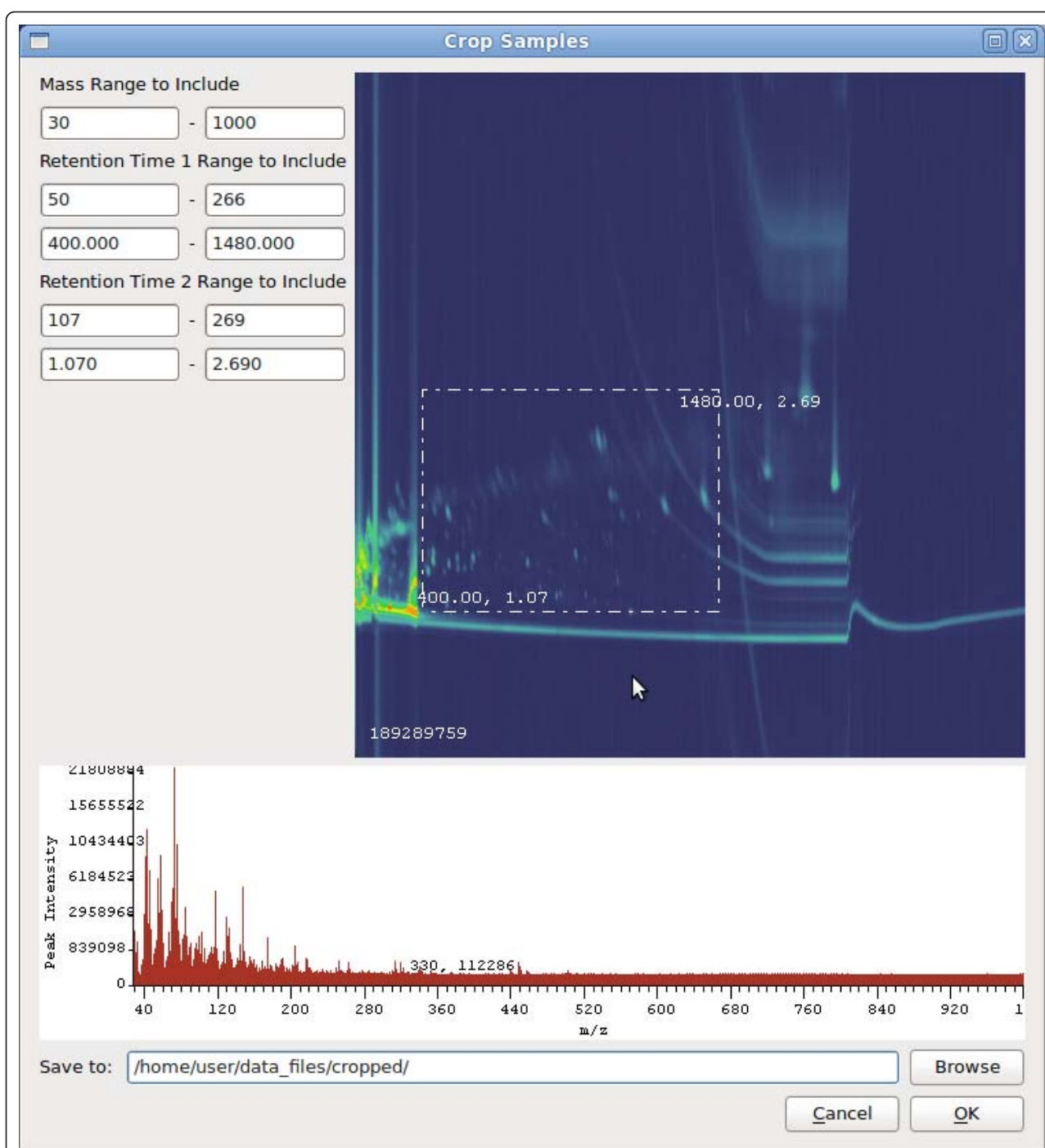
Figure 4 The OmicsVis tool. OmicsVis provides users with the ability to crop data sets, enabling them to remove solvent-saturated areas and reduce file size for ease of use.

negative differences falling downward from the baseline as demonstrated in Figure 5.

Cropping

Datasets created by GCxGC-MS can quickly become extremely large. A typical experiment setup for datasets used in this thesis involve collecting data over a time duration of 40 minutes or more, taking one reading every 10 milliseconds, with each reading containing an intensity value for 900 or more m/z values. Such a setup creates over 850MB of raw data per sample, with experiments containing from 10 to 40 samples. Cropping the datasets not only makes them faster and easier to store and work
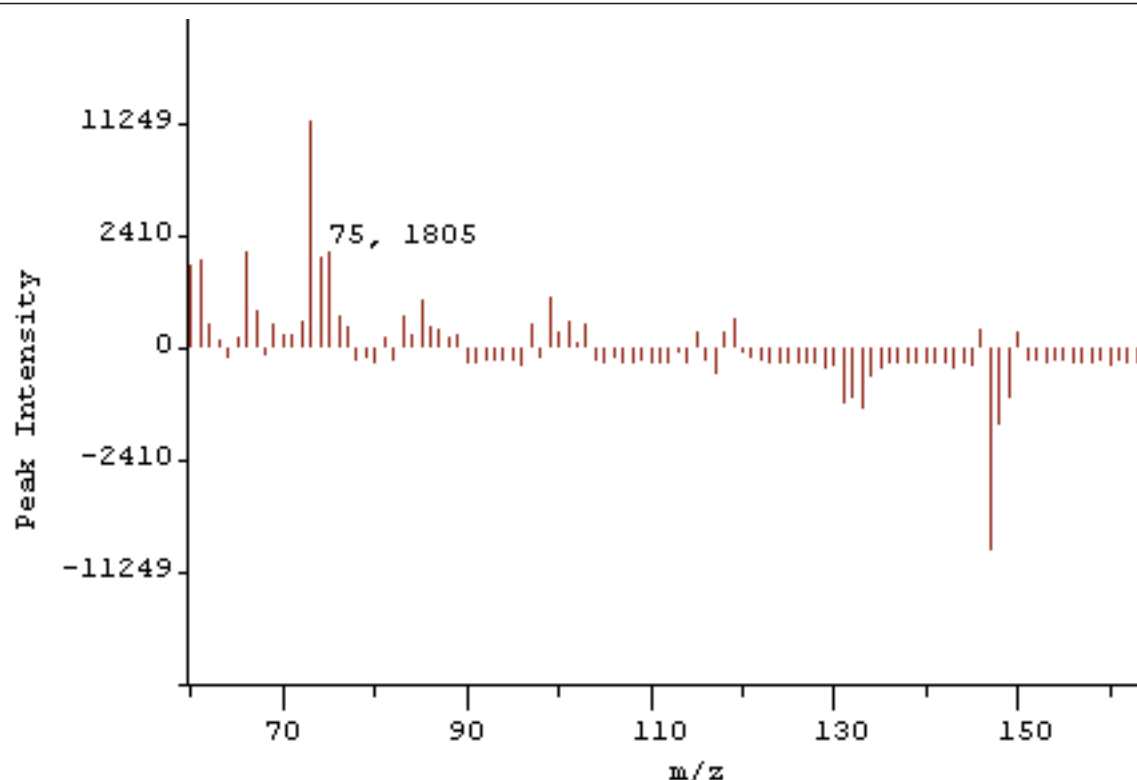
Figure 5 The Difference Spectrum. Visualizing the mass spectral difference between two samples.

with, but also allows uninteresting or problematic areas to be removed. Typically, a large amount of solvent will elute near the beginning of the data collection. The high intensities and large variation in this area can skew the scaling and data mapping schemes employed in analysis algorithms. This area can be graphically selected, and cropping can be performed across a group of samples based on those parameters. Figure 4 depicts the use of the cropping tool.

Mass filter

When a user is looking for bio-markers and meaningful differences between samples, the user will often identify a few mass values that make up some interesting compound. The mass filter tool allows users to select a unique mass or set of masses from the mass spectrum display. Total ion count data for active TIC visualizations is then re-acquired using only the selected masses. The color mapping is then rescaled to match the new range of values across all of the TIC displays, and the TIC displays are refreshed. This allows a user to quickly determine differences among samples for a selected set of mass values. While other applications provide mass filtering via dialog options, none currently support real-time updating of the mass filter in an interactive manner. Mass values can quickly be added to or removed from the filter as the data sets are being

interactively explored for differences in a unique set of mass values.

Volume visualization

Volume rendering is a well-established method for depicting the embedded structures in a three-dimensional scalar field. The most interesting feature of volume rendering is that important regions can be enhanced while distracting details can be hidden by adjusting their transparency. By incorporating non-photorealistic rendering into volume visualization, volume illustration [13] has proven to be very effective in feature-oriented visualization. The exploration of the mass spectral data can benefit from applying volume visualization in three ways. First, visualizing one dataset with the assistance of a multi-dimensional transfer function gives the user a global picture on the intensity distribution and possible peak patterns. Second, the user can freely select subregions to investigate the local mass spectra. The determination is either interactively specified, or automatically found by comparing the mass spectra in a small-sized window with the mass spectra of known metabolites stored in a database. It also facilitates comparing the spectra of different potential bio-markers. Figure 6 compares a selected region of four
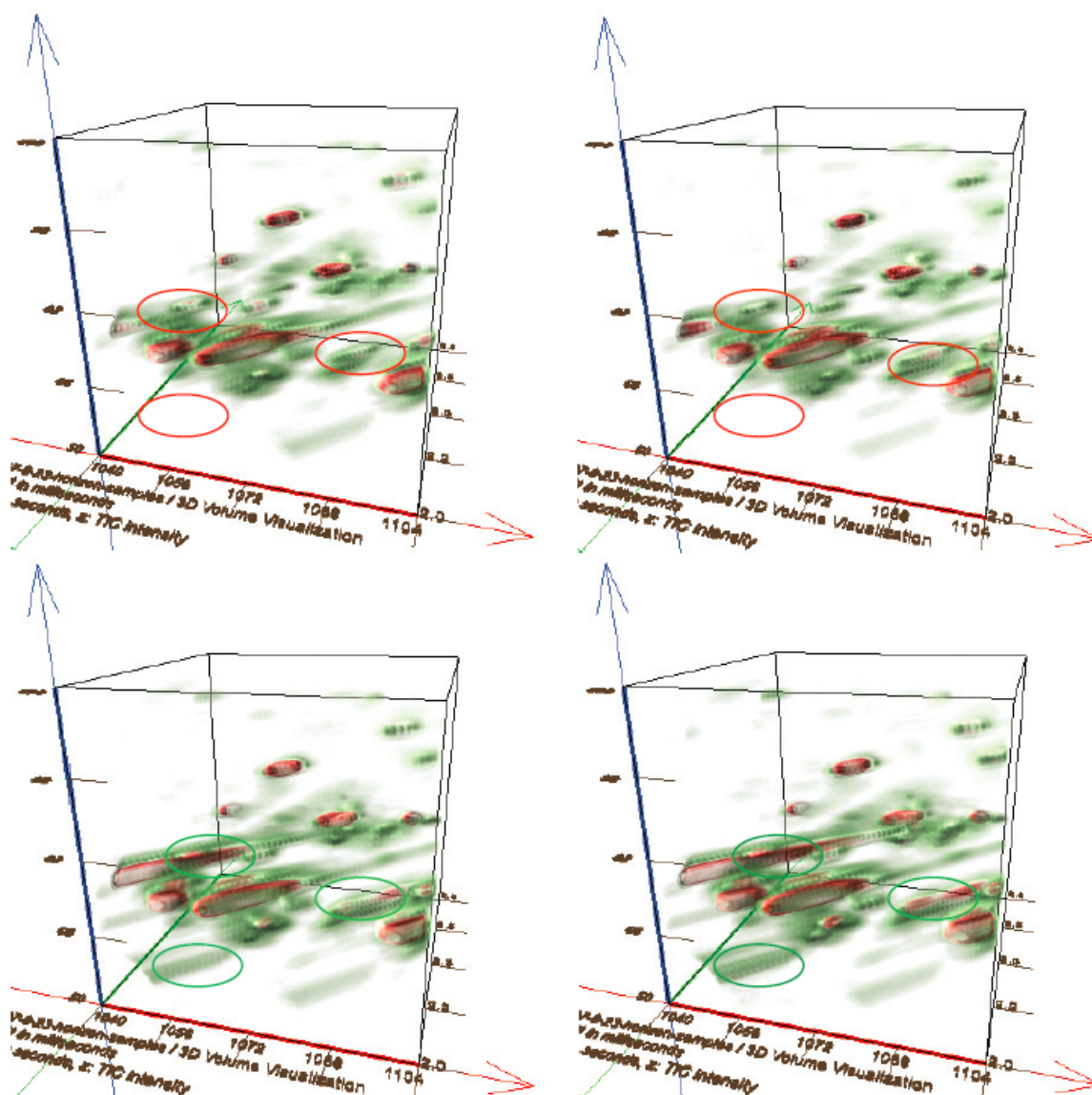
Figure 6 Volume visualization. Volume visualization of selected regions (the first retention time: [1104]; the second retention time: [2.0,2.64]; the m/z values: [50,114]). Top: two healthy samples; bottom: two cancer samples. Visible differences favor locating interesting points for further study. Three circles (red for healthy samples and green for cancer samples) indicate three regions with significant differences, whose locations can be easily.

samples. Third, by globally or locally comparing the mass spectra of a set of sample data with volume visualization techniques, the user's attention can be quickly directed towards the most interesting features, easing the task of finding or verifying the differences among samples

For datasets whose classifications are known, we can use volume visualization to depict the difference among two classes. For the healthy sample set and the cancer sample set, we compute the mean and normalized

variance of the values in each voxel, yielding a template data for each set, respectively. We then visualize the mean values of each template. The variance is used to modulate the opacity after the volume illumination is performed:

$$\alpha_{out} = \alpha_{in} \times (1.0 - k \times variance) \qquad (3)$$

where k is an adjustable constant. In our system, k is set to be 0.5. The variance-dependent opacity modulation enables an emphasis of the common features in

each class. Thus, color mapping schemes as described in the TIC Color Map section may be applied.

Mass spectrum based visual exploration

To accurately study the behaviour of a bio-marker or to differentiate different samples, cancer researchers need to check the mass spectrum at given locations, or of a specific potential bio-marker. However, the cycle of statistical bio-marker detection typically produces more than 6,000 candidates. Numerically checking their mass spectra would be a tedious and error-prone procedure. Our novel exploration scheme allows the users to study the mass spectrum of a set of peaks visually with both the two-dimensional and three-dimensional visual representations of the mass spectra. We also provide a comparative visualization approach that is capable of rapidly locating significant bio-marker candidates from a large set of metabolites, thus greatly reducing the user's time, and improving the exploration accuracy.

Two-dimensional mass spectrum exploration of mouse cancer

The results shown in this section were obtained using raw data exported from LECO's ChromaTOF software. Pre-processing could be applied to remove background levels, normalize samples, and align peaks. In these examples, however, it was interesting to see how the system performs without applying any of these transformations.

This set of samples consists of twenty-nine samples obtained from mice. Ten each were extracted from the body and yolk, while nine were extracted from the head. In each case, half of the samples were exposed to alcohol and half were not. Because alcohol is a depressant and slows down metabolic processes, it is expected that differences between the two samples should show up in the metabolomics data.

The first step taken is to convert the cdf files into raw data files. Since these particular files have very high background levels, the file size drops from about 1.3GB down to 677MB. Compressing the files further reduced the size down to about 245MB. Finally, the samples were cropped to remove regions that were empty or oversaturated with solvent. It was also shown that background level alone occupies mass values above 600, so this area was cropped. The resulting files averaged about 60MB, which can be read from networked file servers in a reasonable amount of time.

The first step after converting the data sets is to view the total ion count image from each one. Since each is displayed using a global color scale, quick visual comparison can identify unusual data samples. At this point, one is basically trying to obtain an overview of the data, and look for obvious outliers. For each pair of alcohol

and control means, a difference volume was calculated. Mean/standard deviation color mappings were created to help visualize the samples. There is a single red peak that indicates an area of higher concentration in the alcohol samples, but several green peaks can be noticed, standing out against the green hued background that is caused by the higher background level present in the control samples. These areas are also indicated by the mean/standard deviation color mapping in Figure 7. A researcher searching for bio-markers could then target these areas specifically for further analysis.

The linked views included in the system allow a user to investigate the mass spectra in real time. This form of focus+context can be used to form hypothesis about what masses or compounds are contributing the observed differences. Figure 8 demonstrates this feature. Without spending too much time working with these samples, a quick effort can be made using this system to look for meaningful differences between the groups. Using visual inspection, samples with missing or blurry peaks can be eliminated. Averages for the remaining samples in each group can then be computed. From the healthy group, multiple samples can be chosen and used for computation of the mean. While the variations present in these samples make conclusions hard to draw, researchers found such visual exploration techniques to be valuable. Particularly, researchers were able to visually inspect data and rule it out quickly, rather than spending days of analysis time trying to detect features and differences that may be completely obscured.

Three-dimensional mass spectrum exploration

We can also explore data in a three-dimensional visual representation to depict its properties such as the location, the uncertainty and the weight. This representation uses the (RT1, RT2, MS) triple coordinate space, and is modeled with a cylinder whose height and radius are proportional to the weight and uncertainty respectively, and whose appearance is encoded as a color-mapped texture representing its mass spectrum, as seen in Figure 9. The visual pattern of each peak is fixed during the exploration. The user can freely navigate in this three-dimensional space, visually check the mass spectrum of specific bio-marker candidates, and locate, explore and compare the properties of the selected metabolites of interest.

Our visualization technique compactly displays the mass spectra in a three-dimensional space, thereby easing the understanding and comparison. Although our mode may introduce occlusion problems, a user can annotate significantly different peaks, and employs other provided operations including two-dimensional and three-dimensional spectrum comparisons and statistics-based filtering to quickly reduce the peak number and remove possible visual clutter. The user can
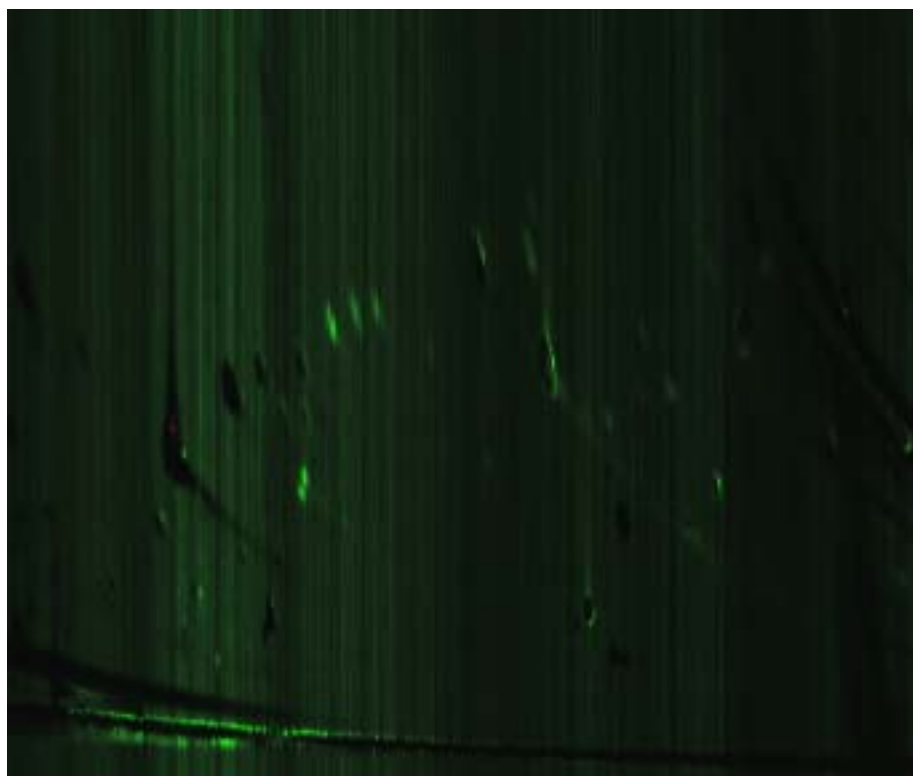
Figure 7 Alcohol mean colorized by comparing with the mean and standard deviation of the control samples. The red area indicates an unusual concentration in the alcohol samples.

progressively filter the bio-marker candidates and the comparison of multiple mass spectra is visually recognizable, and dramatically eases their differentiation. In addition, the user can further search the most significant ones with the assistance of the two-dimensional comparative exploration described below. Figure 9 depicts a simple procedure that locates an interesting potential biomarker (the second metabolite from the left as indicated with an arrow in (c) and (d)).

Discussion

Initial feedback from researchers working with GCxGC-MS data has been very enthusiastic. This system has provided them with their first opportunity to visualize multiple samples simultaneously. Enthusiasm has also been expressed about mean, standard deviation, and difference calculation for a set of samples. By visualizing these calculated data, the human eye can quickly identify differences. This system can be used to identify a particular peak or region of difference, and then the mass spectra can be explored to provide validation of differences, and hypotheses about the compounds involved. With this information, their existing tools can be used to obtain information about compound identification and intensity details much more quickly than was previously possible.

The researchers also frequently mentioned how pleased they were with the speed of the software. Other commercial systems will often take tens of seconds or minutes to even display a TIC image. Additionally, these software systems allow masses to be filtered, and individual spectra to be visualized, however, it is a slow and cumbersome process to change parameters and redisplay a new spectra or filter out different mass values. No other system currently used by this group was able to provide the fast, interactive filtering and mass spectra exploration of our system. As an example, the researchers can now quickly change the mass filter for a specific value, and slowly move through the entire range of mass values. Multiple samples can be visualized, and as the mass filter is updated the researchers can very quickly visually identify cases in which a unique mass has an unusual abundance in some samples. Currently, a version of this system is deployed for use on the Cancer Care Engineering Hub at Purdue University, http://ccehub.org.

The volume visualization tools have also proven to be effective for comparing the selected region of different samples. For instance, Figure 6 reveals some differences between the samples with and without breast cancer in the bottom left portion. Based on the visualization results, the researchers were able to locate the position
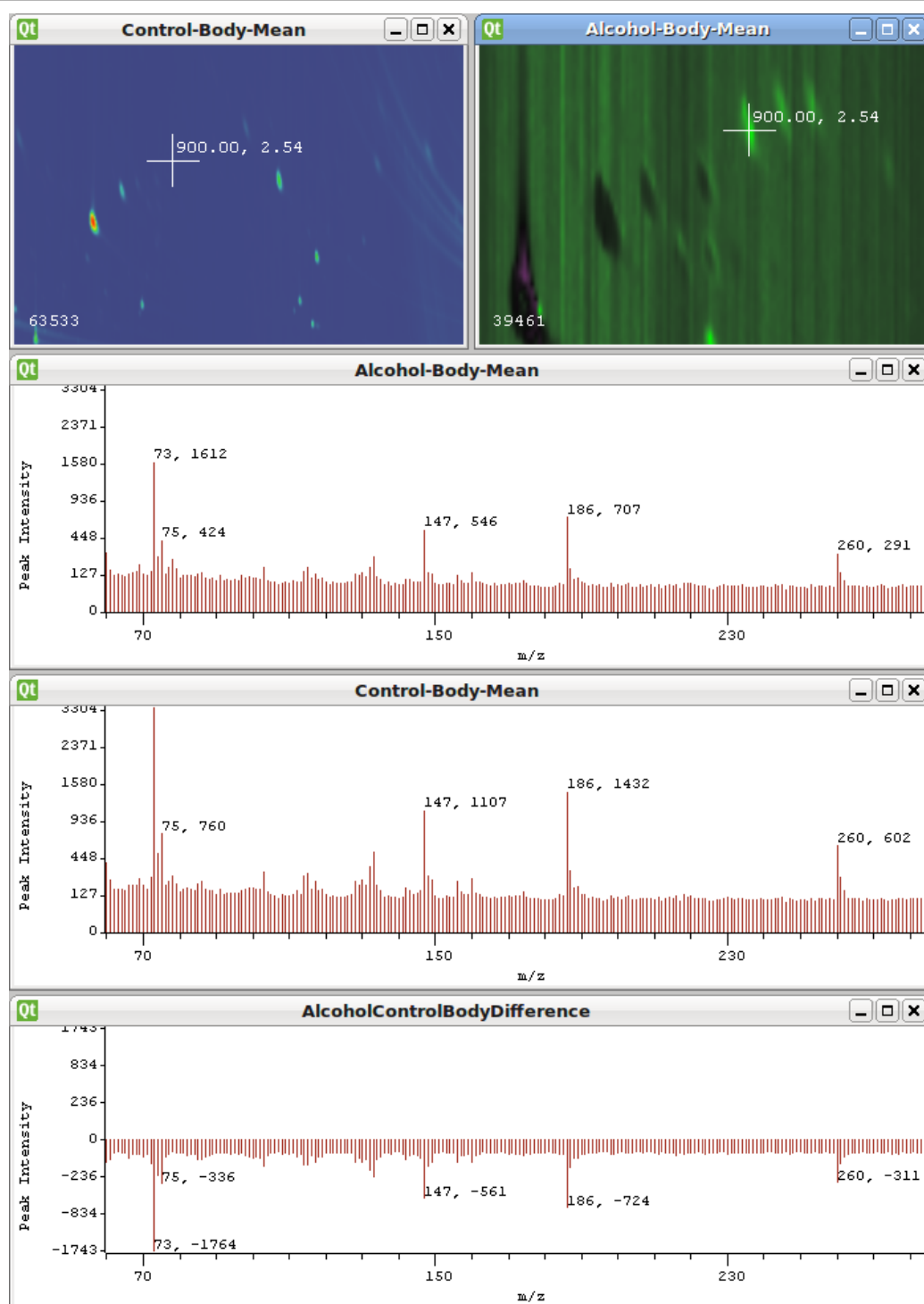
Figure 8 Linked views allow the mass spectra to be explored in real-time as the user interacts with the TIC image.

and check the detailed mass spectrum with the three-dimensional representation. However, we have found that visualizing a large-sized dataset is likely too cluttered to find significant differences easily. In the future, we expect to explore more powerful data visualization toolkits such as spectra-dependent transfer functions to address this problem. These approaches can easily be incorporated into the present visualization framework.

Figure 9 Exploring differences in canine samples. Interactive exploration of the 3500 detected peaks in two samples. The user interactively filters the bio-markers by adjusting parameters, (a) is the result (1300 peaks) by changing the thresholds from (0.01, 0.001) to (0.008,0.004) respectively (see the left top snapshot). (b) is the result (198 peaks) by interactively specifying a region in the region operating characteristics plane (see the left top snapshot). (c-d) Comparative exploration of 198 peaks in two samples. For instance, the bio-markers indicated with red arrows exhibit different mass spectra, which can be inspected by comparing their numerical values, color map views, and plot views provided in our system (see the views in the dashed red rectangle).

Conclusions and future work

Our system provides several benefits to researchers. We have found that the ability to visualize multiple samples and explore the data interactively greatly adds to one's understanding, particularly for users who were previously unfamiliar with the data. Sample visualization allows researchers to quickly validate some level of quality and consistency in their data. Using visual analysis and exploration can help researchers identify differences and bio-markers more quickly than the traditional,

purely analytic approaches. Finally, we have found that being able to visually analyze the data increases the level of confidence in the results obtained.

As our system was evaluated, we also received several suggestion for future work. Plans for future work involve producing output that can be used by other tools. For example, we could allow the user to select a single peak (or what appears to be a single peak) from a TIC image. This could be used to reconstruct a one-dimensional chromatogram for that region, and input that data into other existing tools that would then perform peak deconvolution (if necessary) and identification. Finally, many of these features could benefit from incorporating gradient based value mapping into the display. This was applied to GCxGC datasets in [14], and could be very effective when used with the types of comparative visualization techniques provided by this system.

List of abbreviations used

GC × GC-MS: Gas chromatography × Gas Chromatography - Mass Spectrometry Data; RT1: Retention Time 1; RT2: Retention Time 2; TIC: Total Ion Count; LCMS: Liquid Chromatography Mass Spectrometry; OPLS: Orthogonal Partial Least Squares.

Author details

¹Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. ²School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA. ³State Key Lab of CAD & CG, Zhejiang University, China.

Authors' contributions

PL carried out the system design and implementation and helped draft the manuscript. RM carried out the system design, provided input on the implementation, coordinated the project and helped draft the manuscript. WC carried out the system design and implementation. DSE participated in the design of the project and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Published: 18 May 2012

References

1. Reichenbach SE, Ni M, Kottapalli V, Visvanathan A: **Information technologies for comprehensive two-dimensional gas chromatography.** *Chemometrics and intelligent laboratory systems* 2004, **71**(2):107-120.
2. Livengood P, Maciejewski R, Chen W, Ebert D: **A Visual Analysis System for Metabolomics Data.** *IEEE Symposium on Biological Data Visualization* 2011.
3. Dimandja J: **A new tool for the optimized analysis of complex volatile mixtures: Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry.** *American Laboratory* 2003, **35**(3):42-53.
4. Hollingsworth BV, Reichenbach SE, Tao Q, Visvanathan A: **Comparative visualization for comprehensive two-dimensional gas chromatography.** *Journal of Chromatography A* 2006, **1105**(1-2):51-58.
5. Yao W, Yin X, Hu Y: **A new algorithm of piecewise automated beam search for peak alignment of chromatographic fingerprints.** *Journal of Chromatography* 2007, **1160**:254-262.
6. Reichenbach S, Ni M, Zhang D, Ledford E: **Image background removal in comprehensive two-dimensional gas chromatography.** *Journal of Chromatography A* 2003, **985**(1-2):47-56.
7. Sinha AE, Hope JL, Prazen BJ, Nilsson EJ, Jack RM, Synovec RE: **Algorithm for locating analytes of interest based on mass spectral similarity in GCxGC-TOF-MS data: analysis of metabolites in human infant urine.** *Journal of Chromatography* 2004, **1058**:209-215.
8. Shi J, Reichenbach S: **Restoration for comprehensive two-dimensional gas chromatography.** *Electro Information Technology, 2005 IEEE International Conference on* 2005, **6**:6.
9. de Corral J, Pfister H: **Hardware-accelerated 3D visualization of mass spectrometry data.** *Proceedings IEEE Conf Visualization* 2005, 439-446.
10. Linsen L, Locherbach J, Berth M, Bernhardt J, Becher D: **Differential protein expression analysis via liquid-chromatography/mass-spectrometry data visualization.** *Proceedings of the IEEE Conference on Visualization* 2005, 447-454.
11. Wiklund S, Johansson E, Sjostrom L, Mellerowicz E, Edlund U, Shockcor J, Gottfries J, Moritz T, Trygg J: **Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models.** *Anal Chem* 2008, **80**:115-122.
12. Linsen L, Löcherbach J, Berth M, Becher D, Bernhardt J: **Visual analysis of gel-free proteome data.** *IEEE Transactions on Visualization and Computer Graphics* 2006, 497-508.
13. Ebert D, Rheingans P: **Vollume Illustration: Non-photorealistic rendering of volume models.** *IEEE of IEEE Visualization* 2002, 253-264.
14. Visvanathan A, Reichenbach S, Tao Q: **Gradient-based value mapping for pseudocolor images.** *Journal of Electronic Imaging* 2007, **16**:033004.

SemanticPrism: a Multi-Aspect View of Large High-Dimensional Data
VAST 2012 Mini Challenge 1 Award: Outstanding Integrated Analysis and Visualization

Victor Yingjie Chen [1], Ahmad M Razip [2], Sungahn Ko [2], Cheryl Zhenyu Qian [3], David S.Ebert [2]

[1]Computer Graphics Technology [2]Electrical and Computer Engineering [3]Interaction Design
Purdue University

ABSTRACT

We present a visual analytics system SemanticPrism, which aims to analyze large-scale high-dimensional datasets containing logs of a million computers. SemanticPrism visualizes the data from three different perspectives: geo-temporal, time series curve, and pixel visualization. With each perspective, we use semantic zooming to present more detailed information.

Index Terms: H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces; H.1.2 [User/Machine Systems]: Visual Analytics;

1 INTRODUCTION

The provided data for VAST 2012 Mini-Challenge 1 has many dimensions: in addition to geographic location and time, it also has fields for activities, policies, and machine types. To make sense of the data, it is important to let an analyst see and compare all these different dimensions. Also, analyzing data for such a big and complex global organization, the analyst should not only be able to analyze the world as a whole, but also narrow down and investigate specific offices and the computers within those offices.

To meet these requirements, we developed the system SemanticPrism to visualize the given data from three aspects: geo-temporal visualizations of office health status, time series curves, and pixel visualizations of IP blocks. All these components are interlinked and provide two to four levels of semantic zooming to allow the user to drill down for more information.

2 THE SYSTEM

The system is developed using Adobe Flash, PHP, and MySQL. It is a web application that analysts can run using most modern web browsers. Also, the client-server structure of the web application is naturally suitable for such a problem by keeping the large-scale data in a central location.

2.1 Data transformation and aggregation

The first challenge is to transform the large amounts of data to make it effective for interactive analysis. Directly querying such a large dataset in its raw form is inefficient and can take hours to get a result. We thus created additional indices and tables to speed up the data query and enable responsive system performance. These new tables categorize branches, offices and computers, and aggregate computers with a combination of criteria (e.g. policy, office, and time). In a real-life implementation, such an aggregation and preprocessing could be performed while collecting data on the fly.

2.2 Geo-temporal visualizations

The default view of SemanticPrism is a geographic visualization

{chen489, mohammea, ko, qianz, ebertd}@purdue.edu

with a time slider (Figure 1). Offices are marked as dots of different shapes to encode the office types. To show the computers' health level, the office dots are colored according to the maximum policy violated by its computers at the chosen time.
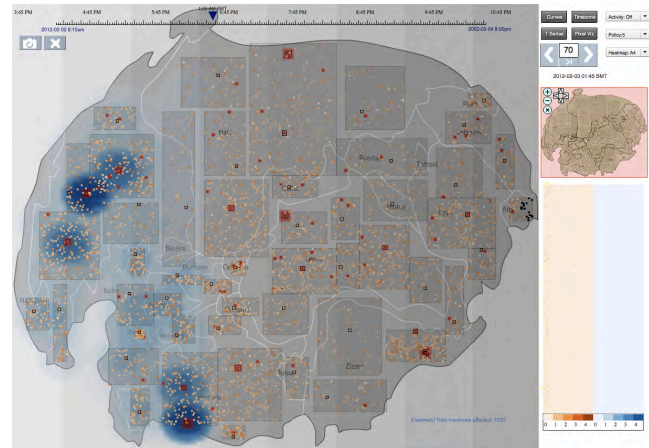
Figure 1: SemanticPrism map view – status at 2012-02-03 1:45am.

A dot with a darker shade of red represents higher policy violation of any computer in the office. With this visualization, if there is even one computer affected by a virus in any office (or other policy violation problems), the user is able to see it immediately. Dragging the time slider automatically updates the status of all offices to the newly selected time.
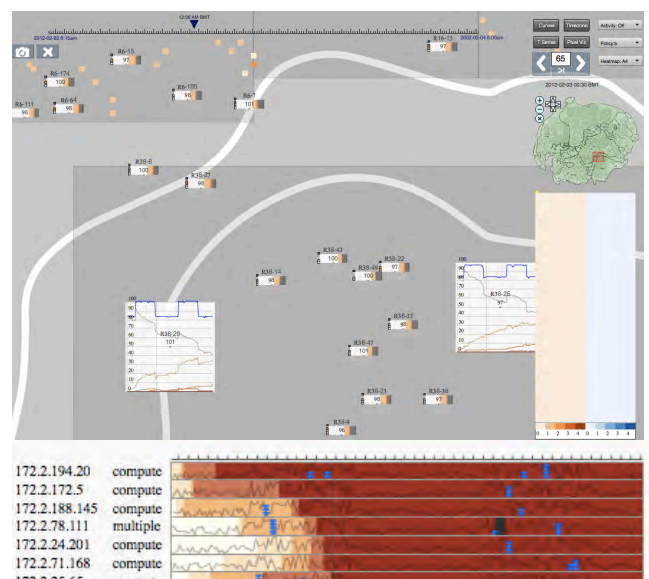
Figure 2: Four levels of semantic zooming to visualize an office

The SemanticPrism map (Figure 1) uses several layers to stack different information together. The heat map layer provides a

visualization of the geospatial distribution of computers with a certain policy or activity. The time zone layer shows the time zones and their local time. A layer of blinking dots highlights offices with a certain policy or activity flag for easy identification among the densely plotted office dots, with the size of the blinking dots reflecting the number of affected computers within that office. With this function, an analyst can easily identify abnormal activities and their growth extent, such as plugging in external devices to computers during nighttime or virus infections.

Semantic zoom allows the user to dynamically drill down and investigate the data in different levels of details (Figure 2). The user may use the navigator (Figure 1, right) to zoom and navigate the map. When zooming in, the space among office dots increase, thereby effectively providing more space to display more detailed information. Depending on this space, an office is visualized in one of four levels: Level 1 visualizes an office as a dot, colored to represent the highest level of policy violation of any computer in the office. Level 2 uses a horizontal color bar to show the percentage of computers in different policies, including being offline. With this visualization, the user may miss policies with small percentages of computers and deduce them as non-existent. To overcome this, icons are used to represent if computers with certain policies exist. Level 3 shows the growth curves of all policies in the office. The curves show the number of computers affected by each policy over time as well as their total. Level 4 shows the history status of each individual computer in the office (Figure 2, bottom). Each computer's policy status over time is visualized as different shades of red in a color bar. The curve in the middle of the color bar shows the number of connections. The computer's activity is visualized as stacked horizontal blue bars with the number of bars representing the activity flag number. The user uses this visualization to drill down to the lowest detail of a specific computer in a specific office.

The user can alternatively interact with the map to show more information. Clicking on a region box will show all offices in that region with level 2 details. The user can then zoom in the region view to show all offices in level 3 and 4 details. Clicking on an office dot will also pop up both the level 3 and 4 detail views.

2.3 Time series curves
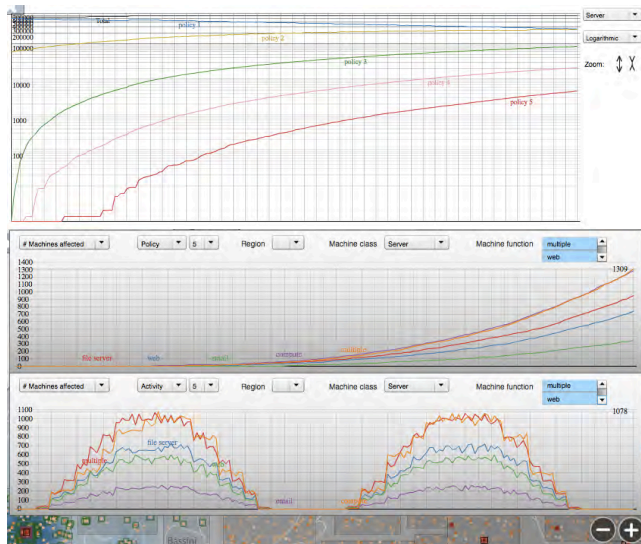
Figure 3: SemanticPrism Time Series Curves.

The SemanticPrism time series curves (Figure 3) provide an overview of the growth trends of policies, activities, and number of connections over the given time period. The default curve view

lets the user see the growth of policies and activities of one class of computers (Figure 3, top). The user can choose to use either linear or logarithmic scale to draw the curve. The logarithmic scale addresses the skewness of the curves towards large numbers and lets the user see the first moment a computer violates a policy, while the linear scale lets the user see the overall growth trend. The time series curve visualization also lets the user dynamically generate new curves and apply a combination of filters for computer class, computer functions, activities, and policies to visualize the number of computers affected or the number of connections. Having multiple panels allows the analyst to perform comparison among the generated curves of different filters.

2.4 Pixel visualizations
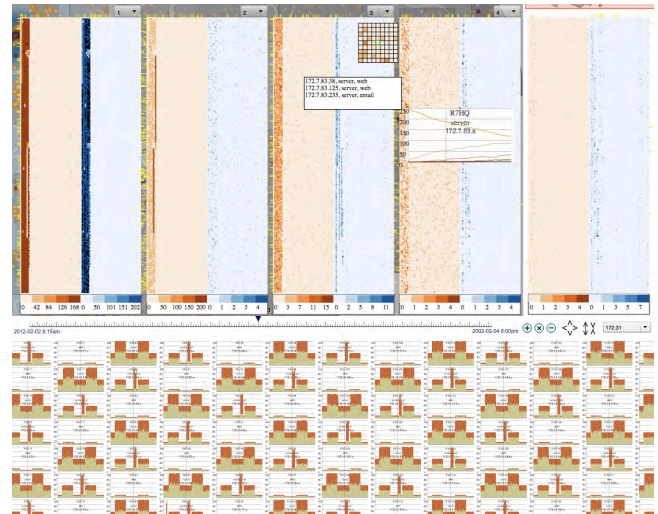
Figure 4: Pixel visualizations of IP bocks.

We incorporate a pixel visualization of IP blocks (Figure 4) to visualize the IP address space since it can provide hints to the network structure. Each panel shows the number of computers within an IP block that is affected by the selected activity and policy (Figure 4, top). In each of these five panels, the red (left) side is for policy and blue (right) side is for activity. Each pixel represents a group of computers in a particular IP address D-block. The X-axis encodes the IP's B-block (172.1 to 172.56), and the Y-axis encodes the C-block (0 to 255). The color of the pixel encodes the number of computers that carries the selected policy or activity flags in the D block.

The IP block pixel visualization has three levels of semantic zooming. The user can zoom in to see the time series curve of all C-blocks within one B-block (Figure 4, bottom). Zooming in further will show all individual computers grouped in the IP C-block using a similar format as shown in bottom of Figure 2.

3 Conclusion

With SemanticPrism, the interconnected three main components: geo-temporal visualization, time series curves, and pixel visualizations, help the analyst explore the data from different aspects. With semantic zooming, the analyst can explore the full spectrum of the data from getting an overview of the world as a whole, to locating problematic areas, to drilling down further to investigate individual computers.

4 Acknowledgement

Spatial Text Visualization Using Automatic Typographic Maps

Shehzad Afzal, Ross Maciejewski, *Member, IEEE*, Yun Jang, *Member, IEEE*,
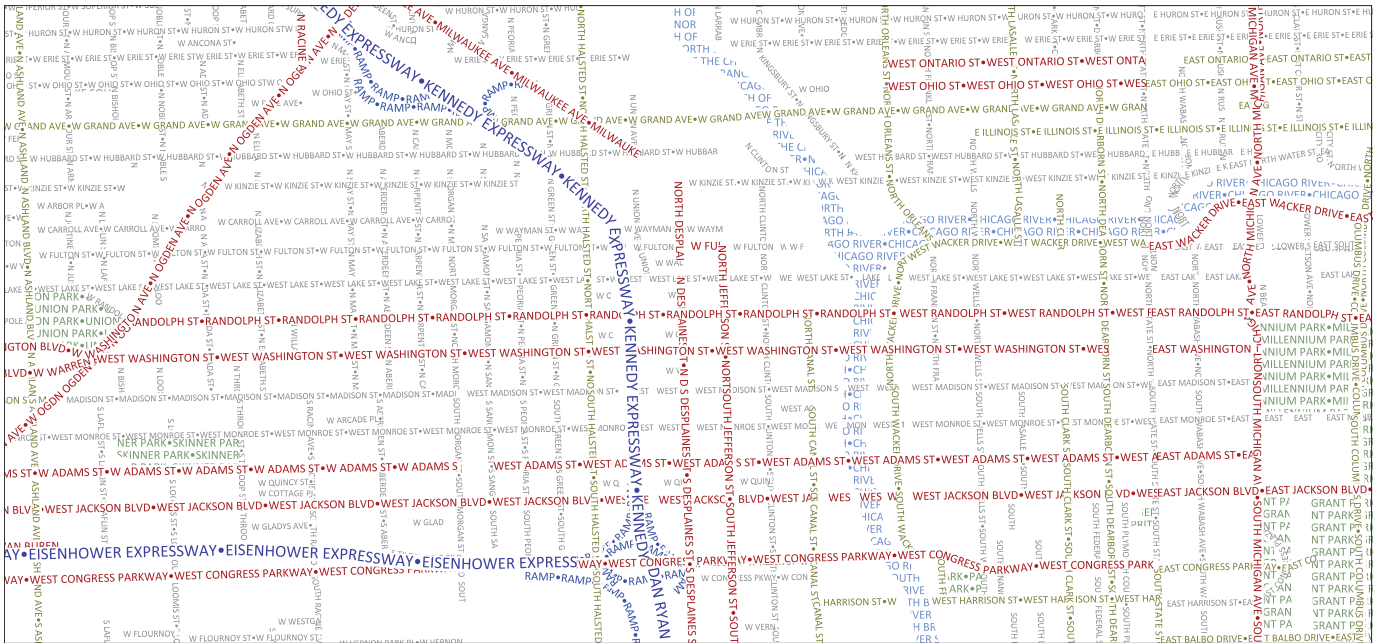Niklas Elmqvist, *Member, IEEE*, and David S. Ebert, *Fellow, IEEE*

Fig. 1. Typographic map for Chicago, IL built using our automatic visualization technique and geographical data from OpenStreetMap. Colors are used to signify different entity types based on a palette used by Axis Maps in commercially sold Typographic Maps.

Abstract—We present a method for automatically building typographic maps that merge text and spatial data into a visual representation where text alone forms the graphical features. We further show how to use this approach to visualize spatial data such as traffic density, crime rate, or demographic data. The technique accepts a vector representation of a geographic map and spatializes the textual labels in the space onto polylines and polygons based on user-defined visual attributes and constraints. Our sample implementation runs as a Web service, spatializing shape files from the OpenStreetMap project into typographic maps for any region.

Index Terms—Geovisualization, spatial data, text visualization, label placement.

◆

1 INTRODUCTION

Textual labels are an intrinsic part of most practical visualization techniques, but they typically play a supporting role in annotating visual entities such as lines, points, and shapes of different color, texture, and weight that make up the visual representation. However, a recent development in the design and infographics community has been visual representations consisting entirely of text [7] (also known as *calligrams* [27]); here, the labels themselves become the sole graphical features. One such interesting technique is Typographic Maps [1], introduced by the cartography company Axis Maps, where entire city maps are rendered only using the names of streets, highways, parks,

waterways, and monuments that together make up the spatial features of the city. The result is a detailed and highly aesthetic geographic map made entirely of the geographical labels (i.e., type) themselves.

However, Typographic Maps, while aesthetically pleasing, are manually created by human mapmakers in a lengthy and tedious process [14]. As a result, Typographic Maps exist only for a handful of North American cities, and most cities in the world will likely never have Typographic Maps made for them. Furthermore, the current Typographic Map approach is purely aesthetic, and does not exploit the full potential of visualizing data using text that has been the hallmark of the information visualization field since its inception [34, 45].

Accordingly, we present a technique for automatically generating a typographic map of any geographic region within seconds (Figure 1). Mimicking the practices of the human mapmakers, our technique wraps text along paths to create lines, and fills polygons with text to create shapes. Furthermore, an automatic technique opens the door to conveying additional data using the spatialized text. We show how to combine our typographic maps technique with spatial datasets to create thema-typographic maps where the individual characters in the map are scaled, colored, or highlighted according to the underlying spatial data. Examples of where this could be useful include conveying crime rate in a city, traffic information on a road network, or demographic data in a geographic region.

- *Shehzad Afzal, Niklas Elmqvist, and David S. Ebert are with Purdue University in West Lafayette, IN, USA. E-mail: {safzal, elm, ebertd}@purdue.edu.*
- *Ross Maciejewski is with Arizona State University in Tempe, AZ, USA. E-mail: rmacieje@asu.edu.*
- *Yun Jang is with Sejong University in Seoul, South Korea. E-mail: jangy@sejong.edu.*

The motivation for our work first and foremost comes from the high visual aesthetics of Typographic Maps—Axis Maps sells poster-sized versions of their maps that are in high demand—which is balanced by the high cost of creating them. An automatic typographic map technique will alleviate the production cost and would make such maps available for any geographic region in the world. The thema-typographic map concept makes the technique potentially useful for data visualization. However, our interest goes beyond this mere automatization process and into the domain of *visual asceticism*: there is something profoundly compelling about a visual representation made up entirely of labels, where the form of the data is also its semantics. The discipline of information visualization is often concerned with providing visual representations that allow people to interpret symbols as a quantity. In the case of cartography, both road names and road lines are symbols that need to be interpreted with respect to their geographic location. By combining the road line directly with the name, we are able to facilitate this interpretation, achieving a combination of symbolic and graphical aspects into a single hybrid representation.

We have implemented our automatic typographic maps technique as a web service that accepts requests for a particular position and region of the world and returns an SVG [43] file representing the typographic map for that region. Our implementation uses OpenStreetMap [29] and spatializes text onto the graphical features of the map. It can also generate thema-typographic maps if the user provides a spatial dataset with the map request. We compare the output of this implementation with a San Francisco map created by Axis Maps. We also show an example of using our thema-typographic map for West Lafayette, IN where high amounts of crime in an area scales up the characters.

The remainder of this article is structured as follows: we begin by reviewing the literature on text visualization, spatializing text, cartography, and label placement. We present our basic technique for automatically generating typographic maps from a geospatial dataset, and then show how these maps can be turned into thema-typographic maps. We describe our implementation and show some examples and close with our conclusions and plans for future work.

2 BACKGROUND

A *map* is a visual representation of a physical space depicting the relationship between the spatial elements of that space. Cartography, the study and practice of making maps, has been around since the cradle of civilization and has long concerned itself with how to best design these visual representations to present the most important and relevant features in clear, understandable, and actionable ways [15, 30, 36].

Below we review the relationship between cartography and more recent efforts from the geographic information science and visualization domains. We then present exciting innovations in infographics and graphic design on the creative use of typography—an intrinsically cartographic consideration—as visual form. We draw parallels to the field of text visualization and then show how these components can all be combined into typographic maps that convey shape, scale, and data in a single visual representation.

2.1 Cartography and Geovisualization

Scale has been long recognized as one of the most important visual variables to the human perceptual system [3, 4], and humans possess a high degree of spatiocognitive abilities that make it possible, even easy, for us to navigate in geographic space [35]. These facts all form the basic cognitive platform upon which maps and mapmaking are based. However, where traditional maps are static, the new fields of geographic information science (GIScience) and geovisualization deal with intrinsically interactive visual representations [36], commonly using *geographic information systems* (GIS) [23].

Outside of the GIScience domain, spatial and geographic datasets are a common data source for our own visualization field. Shneiderman includes 2D geographic maps as one of seven core data types for information visualization [34], and much recent work has explored the intersection of cartography and visualization (e.g., [10, 19, 47]).

Labels are a key feature of maps [36], static and interactive alike, and thus the fields of cartography and geovisualization have long studied the combination of textual labels on graphical features for making geographical maps. Seminal work [3, 4, 15] and handbooks on cartography [26, 30, 36] tend to include long lists of guidelines on appropriate label placement in different situations, as well as the use of typographic conventions to convey spatial information. These guidelines are also being transferred into automatic labeling algorithms in the digital and GIS domains, with work such as that by Kakoulis and Tollis [17] studying general feature labeling, and that by Kameda and Imai [18] looking at optimal label placement on points and curves.

Many of the label placement strategies outlined in the cited work above focus on conveying information about the spatial features on the map using the label as well. For example, labels may sometimes be curved or meandering [15], and typographic conventions also play a role, such as using italics for names of water features [36]. However, beyond these uses, labels are seldom used in cartography to convey abstract information beyond the spatial features themselves.

2.2 Combining Text and Shape

Labels generally play a supporting role in visual representations, but a radical new idea that is gaining traction in both academic and design communities is to generate graphics where the textual labels alone form the visual features: in other words, the labels become the image. This is actually a form of calligraphy and is known as a *calligram*.

From the academic side, perhaps the most well-known example is the extended graph labels technique proposed by Wong et al. [46]. Designed for node-link diagrams, this technique dispenses with traditional lines connecting nodes in the graph, and instead uses the textual label, curved and repeated as necessary, as the de-facto visual link. The WordBridge technique [20] builds on this idea, but takes it a step further by transforming both links and nodes into full-fledged word clouds (edge and node clouds, respectively). However, both of these use standard graph layout algorithms for the spatial position of nodes.

Chevalier and Diamond [7] review the design and infographic side of the text-as-shape paradigm in a recent miniature survey on the synergy between text analysis and fine art. One of their examples in particular, word cloud portraits, is especially relevant to our work because it also utilizes language to convey a visual form. The specific pieces discussed in the paper are Roscover's word cloud portraits for President Obama [32] and Steve Jobs [31]. However, the spatial layout in these portraits are designed primarily with aesthetics in mind.

Recent work by Maharik et al. [27] propose a method for creating *digital micrograms*, which are calligrams using small-scale and readable text. The smooth vector field method used in their work is potentially useful for our ideas, but their paper focuses on aesthetics, as opposed to the representational maps we study here. The key difference between our technique and that of Maharik et al. is that our input consists of a series of network nodes defining the road topology and theirs consists of a segmented image. Furthermore, our technique is faster (on the order of seconds instead of minutes) and is directly suited to geographical visualization where segments of lines and regions can be scaled based on thematic variables.

2.3 Text Visualization

Text visualization uses interactive visual representations to show information about documents beyond their actual text, and has been a prominent focus in information visualization since the field was established [34, 45]. Text is ubiquitous in our everyday life, and text visualization provides a lightweight and low-barrier approach to seeing a different perspective on this type of data.

Perhaps the most common text visualization technique today is the *word cloud* (or tag cloud) [2, 41], where terms are scaled proportionally to their relative frequency and placed on a visual space in some specific order. Popularized by the social media website Flickr in 2002 [41] (although there exist several examples of their use prior to this), word clouds are now in use by thousands of Internet websites.

However, despite their popularity, word clouds are plagued by a number of problems, such as awarding undue attention to long words, difficulties in comparing term size, and layouts that do not promote

(a) Aligning a character to a path.　　　　　　(b) Aligning (and repeating) a label to a path.
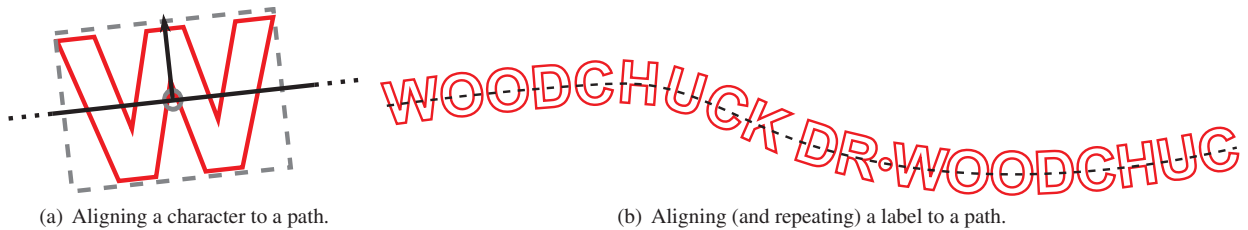
Fig. 2. Fitting text to a path. We use the path normal at each insertion point for a character to derive its orientation.

visual search [41]. The Wordle technique [42] proposes a more space-efficient randomized layout algorithm that results in highly aesthetic clouds. The improved ManiWordle technique [21] allows the user to directly control the cloud layout. A competing approach uses a circular rather than a square layout [25], and yet another allows for layouts optimized around a point or a line [20]. SparkClouds [24] add sparklines to individual terms in a word cloud to show their trend over time, but do not change the term layout. In general, none of these examples use the spatial position of terms for anything other than aesthetics.

Some more recent cloud-based techniques are starting to better utilize the position variable. Parallel tag clouds [9] use one spatial axis to display the temporal attributes of terms in a document corpus. TIARA [33, 44] combines a tag cloud with a trend graph to show changes over time. Clustered word clouds [8, 13] attempt to place commonly co-occurring terms close to each other. TagMaps [48] draw word clouds on top of geographical features, such as for conveying not just the content but also the locality of the keywords. In all of these techniques and systems, not only the size of the word but also its position is important for understanding the visualization.

2.4 Spatializing Text for Geovisualization

Finally, although our work arose from a need to spatialize text in several geospatial applications, we were heavily inspired by the recent Typographic Maps [1] published online by the cartography design firm Axis Maps. These maps, just like our typographic maps, use the names of streets, highways, parks, and city blocks to form a geographical map of the city itself. The maps have garnered much attention and sales on the Internet, and many visitors to the Axis Maps website are requesting maps be created for their cities. However, creating a Typographic Map is a manual and painstaking process performed by a human map-maker, where creating a single map may take several weeks [14]. As a result, Typographic Maps currently exist only for a select few North American cities, and new maps appear only rarely.

The difference between the original Typographic Maps and our visualization technique for generating typographic maps is that our approach is fully automated using geographic information from Open-StreetMap. While we have made efforts to replicate many of the design practices used by the mapmaker, our automated technique naturally lacks the truly creative and aesthetic touch afforded by a human designer making the map. However, with access to an automatic algorithm, we take the next step by using Typographics Maps to visualize data using a technique inspired by proportional symbol maps, where symbols on the map are used to convey a thematic variable.

3 Typographic Maps

Typographic maps are spatial visualizations where the graphical features making up the visual representation consist **only** of text of different size, rotation, and graphical properties. Each text object is arranged so that it conveys not only the *semantics* of the spatial data (i.e., the label), but also its *shape*. Thus, the visualization utilizes spatial position effectively by placing the labels in the area they belong.

Our focus in this paper is on geographic maps, but here we describe the automatic typographic maps technique in general terms. More specifically, in the text below, we will discuss the abstract input data expected by the technique as well as methods for spatializing text onto

polylines and onto regions of space. This discussion sets the stage for our implementation, presented later.

3.1 Data Model

The input data for typographic maps is an abstract 2D shape representation consisting of graphical *layers* and *objects*. Each layer has a name and represents a particular class of graphical objects in the overall representation. For example, for a geographical map, layers are entity types such as highway, ramp, street, block, park, etc. All graphical objects in the shape representation belong to exactly one layer.

As part of the text spatialization process, the user is asked to assign visual attributes, such as font size, color, and weight, to each layer to guide the output. For example, a highway could be assigned a larger font with a black color, while a smaller city street would have a smaller size and use a light gray so that it is not as visually prominent.

The graphical objects in the shape representation can typically be regarded either as 1D paths (e.g., polylines) or 2D regions (e.g., filled polygons). Each graphical object also has a label; for a geographic map, the label of a street would be its name. If a graphical object lacks a label, we can use the name of the layer (often the object type) it belongs to as a label. This label will form the text that will be drawn repeatedly in lieu of the graphical rendering of the path or region itself.

At this point in time, the user may want to define simple text transformation rules to optimize the visual output. For example, because labels will need to be repeated for larger graphical objects, the user may want to specify a separator that will be interleaved with multiple instances of the label. Some layer types may want to use a particular separating character; for example, for an interstate layer (a divided highway) on a geographic map, the separating character may be a road shield. In addition, the whitespace character may be problematic because it causes a visual discontinuity in the output, so the user may want to replace it with something else like an asterisk or dash. Finally, for certain domains, it may make sense to abbreviate or shorten the labels, such as writing "St" instead of "Street", "Ave" instead of "Avenue", and "Blvd" instead of "Boulevard" for a city map.

Finally, all shape representations use a layer ordering that governs which layers should be prioritized (i.e., on top of other layers) for spatial interference. In the example of geographical maps, an elevated highway in a city should clearly be prioritized over the city streets crossing under it, or a street winding through a park should be rendered on top of the graphical region representing the park.

3.2 Paths as Text

Rendering a path using text amounts to fitting the text to the path and repeating it for the duration of the path's length (Figure 2). Fitting a textual label, in turn, equates to iteratively placing each graphical character of the label on the next position centered on the path and rotating the character to align with the path normal at that position (Figure 2(a)). We choose not to warp the graphical character, but instead merely use a 2D rigid-body rotation to maximize readability of the label. This avoids graphical artifacts arising from paths with high curvature at the potential expense of suboptimal curve fit.

There are a number of additional points to consider in order to achieve the look of a continuous path using fitted text. First, the width of the line will clearly now be controlled by the font size—the larger the font, the thicker the line—as well as type face—different font types

(a) Clutter from label overlap.

(b) Character mask (invisible).
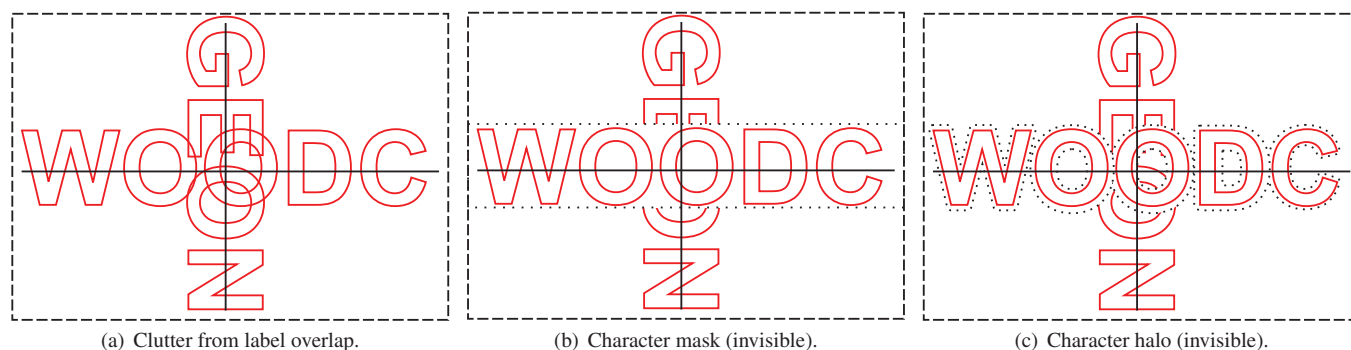
(c) Character halo (invisible).

Fig. 3. Reducing clutter from label overlap by adding (b) white masks and (c) halos to the background. Note that the dotted lines in (b) and (c) are added for illustrative purposes, and are invisible in a real implementation.
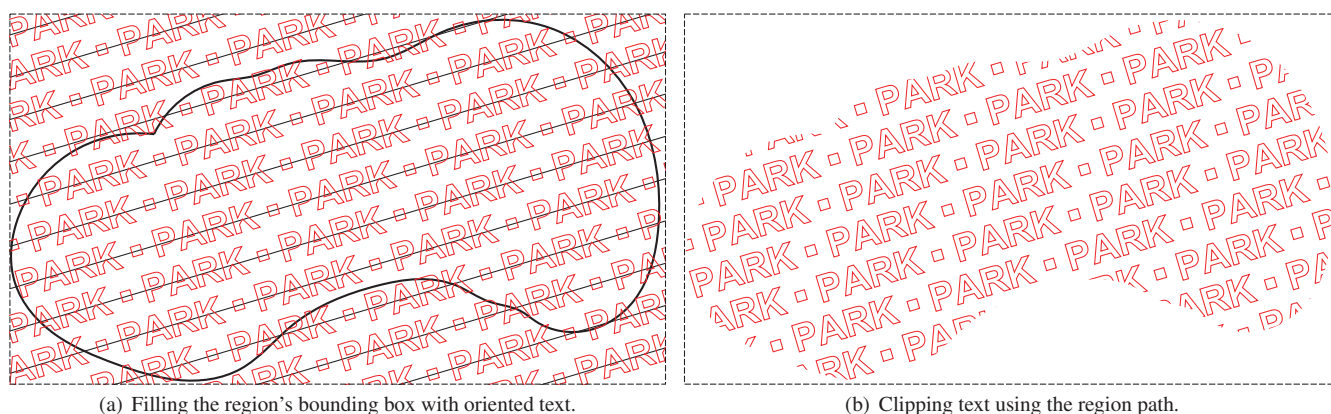

(a) Filling the region's bounding box with oriented text.

(b) Clipping text using the region path.

Fig. 4. Filling and clipping a region with text. Note that we use the same orientation for the whole region, and we also introduce a single-character offset for each line to mitigate character repetition effects across lines.

have different graphical appearances. This naturally places a lower bound on how small a path can be drawn so that the label forming it is still legible. Second, different font types can be used to achieve particular graphical and cartographic effects; for example, in current cartographic practice, a reverse oblique type face is often used to communicate water, such as rivers, lakes, or ponds [15, 36]. Furthermore, to achieve a uniform width, it is often best to use uppercase versions of each word to avoid the lowercase parts of a label taking less vertical space. Finally, as noted above, we introduce a separating character for interleaving repeated instances of a label, and sometimes we may even want to utilize a special whitespace character to avoid the visual discontinuities caused by an empty space in a label.

As mentioned above, our typographic maps technique tracks layer priorities to manage the correct order for overlapping layers, but we may need to do additional work to avoid visual clutter in these situations. Such clutter arises, for example, when two paths intersect, causing the situation shown in Figure 3(a). A simple solution is to add a white rectangle as a background (known in cartography as a *mask* [36]) behind each character as illustrated in Figure 3(b). A more advanced solution would be to add a so-called *halo* [36] (or *null halo*) behind the label; this is illustrated in Figure 3(c) (the use of halos for enhancing depth perception is also prominent in illustrative visualizations in 3D [11, 16, 37]). Another approach may be to use an outer stroke (i.e., an outline) in the background color (typically white). All solutions have strengths and weaknesses—for example, halos are common in some cartographic designs, but may introduce more clutter than the more regular appearance of masks. The Axis Map designs tend to use masks, presumably for this reason.

Finally, another important issue for maximizing the legibility of our labels is to consider its orientation. Cartography has many guidelines for text orientation that we may consider [15, 36]. For example, text

is obviously easiest to read when it runs from left to right and right side up. For paths with mostly vertical components, Byrne [6] demonstrated that marquee placement of text was outperformed by rotated horizontal text and that reading the rotated text from bottom to top or top to bottom had no impact. We employ rotated horizontal text reading from bottom top. Our typographic maps technique tries to enforce these orientation rules. In some cases, paths radically change orientation throughout their existence; consider, for example, a beltway circling a city. This makes it difficult to find an optimal orientation. For these situations, we split the path into segments depending on the predominant direction and orient the text on a per-segment basis.

The original Typographic Maps are all manually designed, and it is interesting to see that not all roads are labeled with consistent text orientation (compare "San Bruno Ave" and "James Lick Freeway" on the right side Figure 7(a)). Switching orientaton, particularly for adjacent roads, may be a way to better distinguish between roads and increase label readability. Adding such design guidelines to the layout algorithm is left for future work, however.

3.3 Regions as Text

We render a region as text simply by filling the interior of the region with the text, repeated as necessary (Figure 4). Because regions are two-dimensional areas, fitting the text to the area needs a radically different approach than for paths. Our solution takes the bounding box of the closed path representing the region and fills it with straight lines of the textual label, repeated as necessary (Figure 4(a)). The lines all have the same orientation, and an increasing character offset is used between adjacent lines to avoid unseemly visual artifacts arising from the tiling of the label. Finally, as Figure 4(b) shows, we use the closed path of the region itself as a clip path. However, in sticking with the text-as-shape paradigm, we do not draw the border itself.

Issues may arise when rendering two regions adjacent to each other, in particular if they share the same label. The best option in this case may be to use different visual attributes—such as font weight, size, or color—between the adjacent regions, but this may not always be an option (in particular if the regions belong to the same class, so they have the same color). In these situations, we vary the orientation of the lines in the adjacent regions, causing a visual discontinuity between the regions that will be perceived as a border (Figure 5). Additionally, regions could be segmented and transformed to text using the algorithm presented by Maharik et al. [27].

Fig. 5. Varying text orientation to distinguish between adjacent regions; useful if the regions have no other distinguishing graphical attributes.

An advanced rendering option that we have not yet explored is to use paths other than straight lines inside each region. For example, in Typographic Maps [1], water areas typically use a wavy pattern, presumably to better communicate the region type. In the future, we foresee extending this to other layer types or for additional spatial data, such as conveying contour lines on mountainous or hilly areas, prevailing currents in an ocean, or wind direction on a meteorological map.

4   THEMA-TYPOGRAPHIC MAPS

Thematic maps are geographic maps where a geospatial variable is visually encoded on the map [36], and have been used to visualize various demographical, political, and economical data for various regions in the world. For example, John Snow used a thematic map to identify a contaminated pump during a cholera outbreak in London in 1854 [40]. Examples of such maps include chloropleth maps [36] (coloring spatial aggregations), cartograms [12, 39] (distorting space and distance), and proportional symbol maps [36] (scaling map symbols).

Fig. 6. Scaling individual characters on a path to convey a thematic variable (such as traffic) in the underlying spatial dataset.

We introduce the concept of *thema-typographic maps* based on our automatic typographic maps technique. The basic idea is simple: instead of merely repeating the same textual label along a path or inside a region as described above, we use the font attributes—typically size, but color or intensity is possible—on a per-character level to convey the value of a statistical variable at each character's spatial location. Because the mapping operates merely on font attributes and does not affect the characters themselves, the semantics (i.e., the labels) of the typographic map is preserved. Figure 6 illustrates this simple concept for a path rendered using text; the idea is similar for regions. Note how the varying character size creates the impression of a band that is wide at the edges and that shrinks in thickness in the middle.

Because our thema-typographic map scale (or color) the elements of the map on a per-element level, we see many interesting applications for this technique. For example, Figure 8 shows a thema-typographic map generated using our approach where roads are scaled

proportional to crime in the area. This yields a visual mapping where areas with a high degree of crime are made visually larger, and those with less crime are smaller in size. Other uses include demographics, political, and traffic data being overlaid on a geographical map.

5   IMPLEMENTATION

We have implemented the automatic typographic maps technique as a web service called TYPOMAP that uses OpenStreetMap [29] as a geographic database. Given a map query with the bounding box in longitude and latitude, our web service returns an SVG [43] file representing the typographic map of that region. This file can then be viewed in a modern web browser or vector editor, or printed and viewed off-line.

5.1   Overview

Our web service implementation consists of three simple steps: (1) retrieving map data in XML format for a requested area on the globe from the OpenStreetMap XAPI service; (2) cleaning, filtering, and merging paths and regions in the data and storing it into an internal 2D vector database; and (3) rendering the vector regions and paths in the database as text using the SVG format, that is returned to the caller.

5.2   Data Source: OpenStreetMap

TypoMap uses the XAPI interface to perform read-only map queries to the OpenStreetMap (henceforth, OSM) web service. The return value for such map queries is an XML file in the OSM data format that we proceed to parse and use to populate an internal vector database.

OSM data is optimized for both rendering and routing, so the data is generally of high quality and well-suited for rendering as a typographic map. However, the path data is often redundant and somewhat irregular. Our XML parser performs low-level cleaning and filtering to ensure good results: bike trails and footpaths are omitted, and segments that are smaller than can reasonably be represented on the typographic map (small water features in particular) are also filtered. We then use a higher-level filter component where the resulting data is optimized for rendering as a typographic map based mostly on aesthetics; this involves steps such as (a) connecting separate paths, which occurs in OSM when a road changes name; (b) combining the divided lanes of a highway into the same path; and (c) managing path priority, such as weaving paths or giving horizontal paths priority over vertical.

OSM by default uses a slightly modified version of the spherical Mercator projection (also known as the Google Mercator projection since it was first introduced by Google Maps in 2005), and we adopt the same projection for mapping the longitude and latitude of map features onto the flat canvas of our SVG file output.

Furthermore, all map elements in the OSM data format contain freeform but standardized tags on map features. We use a subset of these standardized tags as the *layers* in our model with a predefined priority order as well as graphical attributes for font, color, and size.

5.3   Typographic Maps in SVG

Rendering typographic maps proceeds one layer at a time. Each layer in the data extracted from the data source has a specific font, color, and size (configurable upon invoking the web service). We render the map simply by rendering each layer at a time, with the lowest priority layer first (meaning that it will end up behind all other layers).

Both paths (polylines) and regions (i.e., the polyline defining the region border) in the TypoMap vector database are represented using the SVG `<path>` construct, defined once in the SVG file but given an identifier so that it can be used multiple times. However, depending on whether we are rendering a path or a region, we proceed differently:

- **Paths:** We first render the white background mask using a simple polyline with the appropriate stroke thickness (depending on the size of the characters in the layer). We motivate our choice of a mask by the fact that this is the solution the Axis Maps designs use. We then render the text using the `<textPath>` element, which wraps the specified text along the path. Because the SVG default is to use the path as a baseline (i.e., the text is placed on

(a) Original Typographic Map of San Francisco, CA.   (b) Automatic typographic map of San Francisco, CA.
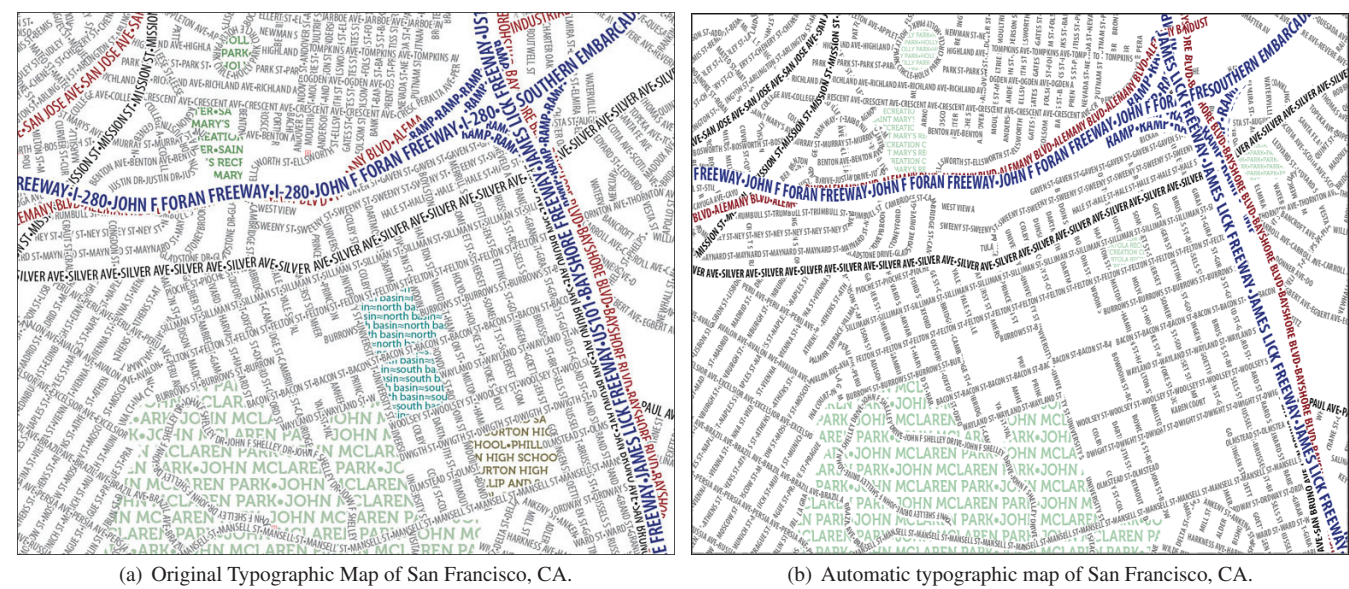
Fig. 7. Comparison between a manually created Typographic Map (left) and an automatic typographic map (right) of the same area of downtown San Francisco. The automatic one was generated as SVG in less than one second and then rendered to a same-resolution bitmap for faithful comparison. The original Typographic Map on the left is used with permission from Axis Maps.
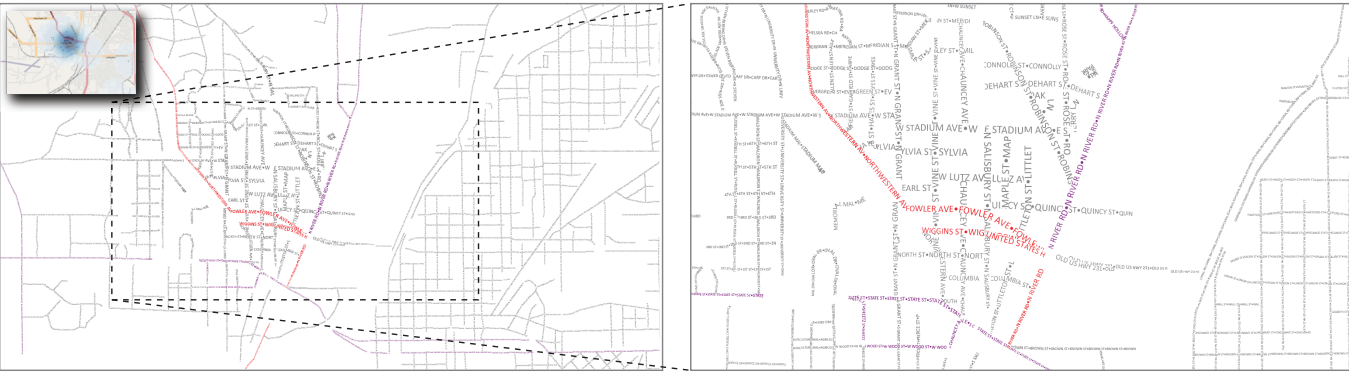

Fig. 8. Thema-typographic for West Lafayette, IN where the statistical variable being visualized is crime rate. larger text means more crime. The right picture shows a zoomed-in area around the Purdue campus, and the inset (upper left) shows the KDE map for the entire region.

the line), we must offset the text 50% of its height so that it is centered on the line instead.

Our implementation tries to orient text according to the cartographic rules outlined earlier; however, due to the arbitrary ordering of vertices and the merging of paths in the OSM database, our layout does not always succeed in enforcing them.

- **Regions:** We first fill the region outline with white to mask out any background features. We then set up clipping using the <clipPath> element with the region outline as the argument. This causes any subsequent rendering to be masked against the region outline. After that, we use the bounding box of the region to emit text on evenly spaces lined at the same (Figure 4), randomly chosen, angle. A more advanced implementation may find the primary axis of the region to use as an orientation, or may choose to use map coloring to ensure that adjacent regions get assigned one of four fixed orientations instead.

Our implementation uses Cascading Style Sheets (CSS) classes for the different map layers, which allows the viewer to change the graphical appearance of the map, if not the geometry itself. This is useful when conforming to the strict typography guidelines that cartography stipulates on font family, weight, case, color, and intensity [30, 36].

5.4 Thema-Typographic Maps in SVG

Rendering our thematic extension means slightly more complex SVG output. Since we are resizing individual characters, we must internally keep track of where different characters in a text fall in the 2D space of the map. This means that we must render the text on the path ourselves since the SVG <textPath> command gives no way for the designer to know where on the path individual characters will appear.

To achieve this, we use an internal low-fidelity renderer that, given a specific font and size, steps through characters in the text being rendered and uses the letter width and spacing to calculate the exact 2D location on the path where the character will be placed. Using this location, we look up the mapping variable in the spatial dataset and use its value to scale the size of that character. Individual characters in a text object can be modified in this way using the <tspan> element.

Our previous strategy of drawing a single white line on the path with the appropriate stroke width as a mask will now fail since the mask will have to be of varying width to accommodate the varying character size. Instead, our current implementation simply sets the stroke width to be the average of the minimum and maximum font size for a path. Unfortunately, this does result in some situations where there may be visual overlap. A better solution that we plan to implement in the future is to generate a white closed polygon as a mask using the envelope of the character bounding boxes along the path.

5.5 Implementation and Performance

TypoMap is implemented in C# and uses only the standard Microsoft .NET libraries. The core component is a collection of 2D vector shapes that is used as the internal geometric representation of the map to render. This also allows us to fully modularize the input and output components: The OSM XML input parser creates the vector representation, and the SVG renderer uses the vector representation for output. We are able to accommodate different input or output formats simply by exchanging these modules. For example, we could use replace the OSM parser with an image processing component that reads an image and generates vector paths and text to achieve output similar to Roscover's word portraits of Barack Obama [32] and Steve Jobs [31].

Rendering performance is on the order of 2-3 seconds (not counting network transfer time) even for large bounding boxes (the OSM XAPI allows for retrieving regions up to 100 square degrees). Generating a thema-typographic map takes on the order of 10 seconds because of the need to internally render the characters on the paths, but performance is still easily within acceptable parameters for a web service.

6 RESULTS

Figure 7 shows a side-by-side comparison between a Typographic Map for San Francisco provided by Axis Maps, and our automatically generated typographic map for the same region. Since we only have access to a rasterized version of the original, we have rendered our SVG file as a raster image for comparison. We have also chosen font faces, colors, and sizes to be similar to those in the original.

Comparing the two maps side-by-side like this, it is clear that there are differences in their appearance. In particular, the mapmakers who created the original Typographic Map have generalized the map in some places to better convey the spatial features—for example, small roads running parallel to a highway have often been manually removed, whereas our automatic typographic map makes no such distinction. Nevertheless, it should be noted that the original Typographic Map on the left was created by a mapmaker working full-time for two weeks, whereas our automatically generated version on the right required only 2-3 seconds to generate. Since our web service generates standardized SVG that could be imported directly into the Adobe Illustrator tool the Axis Maps designers employ, perhaps one use of our technique is as an initial rough rendition of a Typographic Map that a mapmaker can use as a starting point for further refinement.

The thema-typographic map in Figure 8 shows the spatial layout of West Lafayette, IN where crime rate has been visualized onto the map. We define character size to be proportional to the amount of crime in that region, which causes crime-stricken (and presumably dangerous) areas to become large, and safer areas to become small. In this way, the map intuitively conveys a feeling for the rate of crime in the city.

Finally, Figure 9 shows a large-scale typographic map of the downtown area of Seattle, WA where the surrounding Puget Sound and Lake Washington have been polygon-filled with their names. Figure 10 shows an inset around the conference hotel for VisWeek 2012. This map required on the order of 5 seconds to generate.

In creating these images, a number of the generation issues are due to topology issues associated with the OpenStreetMap data. Roads may start and stop and can be renamed or reclassified from node to node. These issues can result in poor textual overlays for streets. Our future work will look at ways of intelligently merging such problematic typology regions in order to produce more stylistic results. Also, depending on the length of the line segment and the length of the label, some map features may not be able to be labeled with the full name. Along with the labeling issue, such small streets can often cause clutter within the image. Future work will explore means of removing small line segments for clutter reduction as well as prioritizing smaller streets so that if overlap occurs, the algorithm will first ensure that at least one iteration of the street name is visible.

We sent the example images of our automatic typographic map technique in this paper to Axis Maps asking for their feedback. The representatives from Axis Maps thought the results were "fantastic" and "very impressive", especially for the depth sorting of the streets and the type casing our technique does. The person did point out two

weaknesses: (1) that additional and more aggressive road generalization is needed, and (2) that the polygon fills should be oriented at 45° to set them off from the street grid. We have since addressed both of these comments: our path generalizer can be fine-tuned to control the amount of generalization, and text in the polygon fills can be rotated in any orientation. He also asked whether they could get access to the program themselves at Axis Maps. We are currently preparing a demonstration release for their use.

7 DESIGN IMPLICATIONS: DRAWING WITH TEXT

While we have focused on geographic maps in this paper, the general concept of *drawing with text* can be applied to virtually any spatial representation. As evidenced by our literature review, combining text and shape has already been explored both within the academic community—such as in the extended graph labels by Wong et al. [46]—as well as within the art, infographic, and design communities—such as in Roscover's word cloud portraits [31, 32]. Our emphasis here is on how to automatically replicate the methodology and design aesthetic of human mapmakers, but it is worthwhile to consider the utility of drawing with text in a broader context.

Visual representations consisting entirely of text are visually interesting and may contain more information than the corresponding representation consisting only of paths and filled regions. As we have seen in our work, this allows us to convey not just the spatial extents of a visual feature, but also something about the semantics of that feature. There is something profoundly compelling about such a visual representation, and it approaches what we would like to call *visual asceticism* or *visual minimalism*.

Of course, by the same token, as data density goes up, so does the visual complexity. A complex visual representation can be difficult to understand and may impose a high cognitive load on the viewer. Therefore, a graphic consisting entirely of text may be counterproductive for situations when the user needs to make quick and accurate decisions. Consider a police officer trying to find the location of a call on an electronic map—the officer's performance may be degraded with a typographic map due to the extra distractions of all streets being drawn as text. On the other hand, here is an opportunity for our typographic maps technique: because our approach is automated, we can seamlessly switch between typographic and standard 2D vector graphics. Continuing our example with the police officer, we may choose to—for lack of a better word—*textualize* the street around the location of the call, but draw all other streets as vector graphics. The resulting map is preattentively distinct and may even be construed as being **less** visually complex than drawing **both** street and street label. Clearly the lesson here is that drawing with text is a powerful yet potentially confounding method, and should be used with caution.

The above example also touches upon another issue that we have not yet discussed in this paper: how do you interact with a spatialized text visualization? Because our typographic map technique produces SVG files as output, standard navigation operations such as panning and zooming are easy to perform. Furthermore, since the wrapped labels remain text in the SVG document, it is even possible to select and copy portions of the streets or regions on the map. However, additional and more complex interactions are certainly possible, even if we have not explicitly studied them in this paper. Examples include searching for entities on the map, highlighting spatial features (perhaps as a result of searching), or switching between drawing with standard vector graphics and with text inside a magic lens [5]. And what about embedding driving directions to a particular destination in the graphical path to the destination on top of the map itself?

The design space of drawing with text goes beyond interaction. Color is one dimension that we merely use to distinguish between different categories of features (similar to the Typographic Maps of Axis Maps), but there exist many other possibilities as well. For example, we discuss mapping the thematic variable for our thema-typographic maps onto a color scale for characters instead of their size. Another design dimension is supporting different languages and even character sets when drawing with text in general, and for our typographic maps
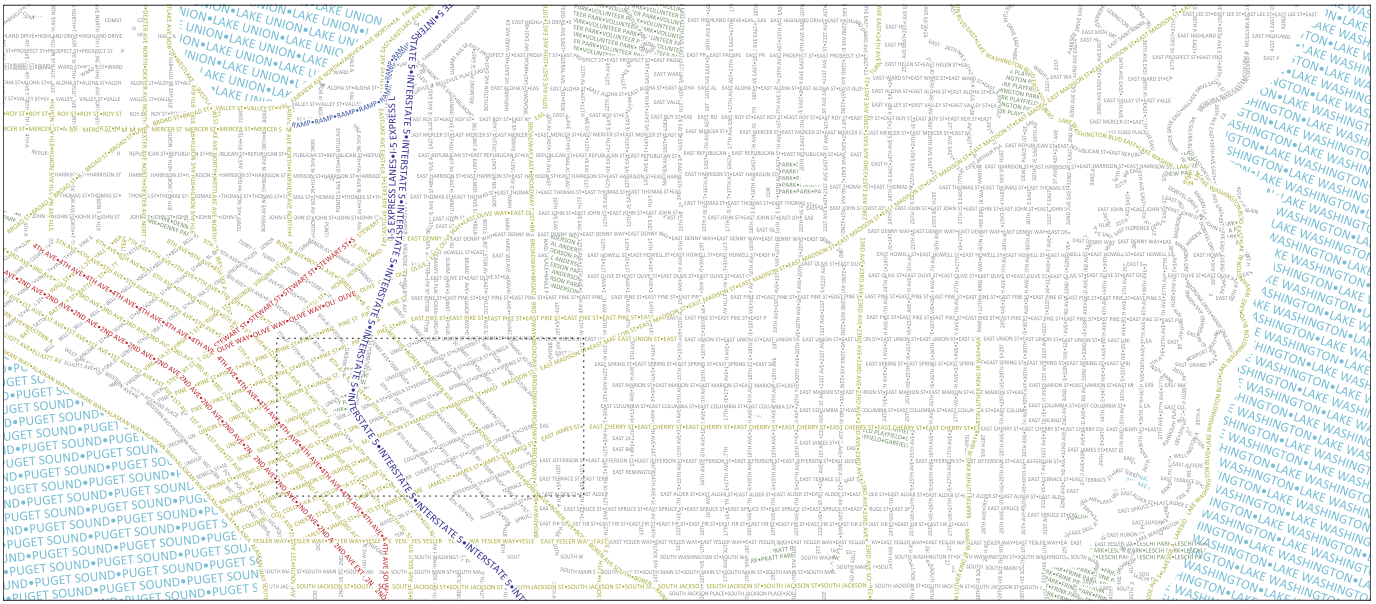
Fig. 9. Automatic typographic map of the downtown area around the VisWeek 2012 hotel in Seattle, WA. This image was generated using our TypoMap web service in less than 1 second based on OpenStreetMap data. The SVG can be viewed (and printed) using a modern web browser.

technique in particular. Maps are clearly relevant for a global audience, but we have yet to study how to support this for our technique.

This is also the place to discuss an oft-overlooked aspect of visualization that the concept of drawing with text exhibits: aesthetics with regards to beautiful imagery [22, 28]. As evidenced by the mainstream success of text visualizations such as tag clouds [2, 41], Wordle [42], and TagMaps [48], the notion of visualizing textual data seems to be inherently compelling to a general audience. Taking the step to use the text itself to represent graphical features has garnered even more attention [7]: Roscover's word cloud portrait of President Obama made the cover of *Time* magazine, and Axis Maps' Typographic Maps [1] are in high demand and are sold commercially on the company's website. In a recent paper [28], Vande Moere and Purchase argue for adopting commercial and artistic design practice to information visualization, and we think that the concept of drawing with text is an excellent compromise between these principles of aesthetics and the core need to accurately convey information using graphics.

Of course, aesthetically appealing imagery is just one side of the coin; for a text-based visualization such as ours to be useful, it has to be accurate as well as readable. While our technique at its core is designed for accuracy (unlike techniques such as the digital micrograms of Maharik et al. [27], which first and foremost prioritize a visually appealing appearance), there are still several aspects in which it could be improved. For example, we think there is still work to be done for our algorithm in improving readability by delimiting labels, as well as reducing visual clutter in geographically dense areas. Furthermore, when rendering small regions and paths, it is clearly important that at least one complete label is visible in the textualized representation. Alternatively, if the region or path is too small to fit even one complete label, perhaps the geographical feature should be omitted entirely, or shown using an iconified representation—a little like semantic zooming for typographic maps. Adding such advanced importance-based rendering and filtering functionality is left for future work.

At the same time, cartographic labeling often deliberately uses imprecise labeling to indicate the fluid nature of geographic features; for example, labeling a mountain range does not necessarily commit to a specific extents or precise location for the range. This approach is very much in line with what has been called Tobler's *first law of geography* [38]—"everything is related to everything else, but near things are more related than distant things"—and we look forward to seeing how these ideas can be extended in the future given this new take on making the labels the actual map.
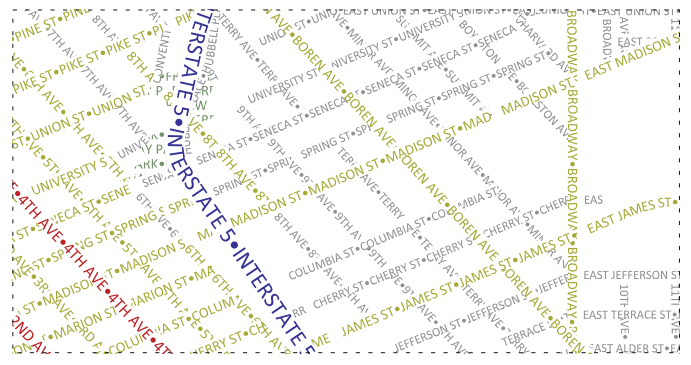
Fig. 10. Detailed area around the VisWeek 2012 hotel in Seattle, WA.

8 CONCLUSIONS AND FUTURE WORK

We have presented an automatic algorithm for generating typographic maps—geographic maps consisting entirely of text as the only graphical features—using the OpenStreetMap web service, allowing us to generate SVG renditions of any region on the globe within a matter of seconds. Using this framework, we have shown how to use our algorithm for creating thema-typographic maps, where the size of the text on the map can also be scaled based on spatial data features such as crime, demographics, or traffic information. We have also presented several examples that showcase the utility of this technique.

We see many avenues for future research. We are very interested in continuing to explore the use of spatial data features as a means of visualizing data using the typographic map representation. The step is also not far to spatialized Wordles, or Wordle Maps, where the labels no longer convey the name of the geographic feature, but instead carries semantic meaning specific to the dataset. Finally, our maps are currently static SVG files and do not incorporate any interaction, and an obvious extension is to add operations for navigating, drilling down, and changing the map layout.

ACKNOWLEDGEMENTS

REFERENCES

[1] Axis Maps. Typographic maps. http://www.axismaps.com/typographic.php.

[2] S. Bateman, C. Gutwin, and M. A. Nacenta. Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proceedings of the ACM Conference on Hypertext and Hypermedia*, pages 193–202, 2008.

[3] J. Bertin. *Sémiologie graphique: Les diagrammes - Les réseaux - Les cartes*. Editions de l'Ecole des Hautes Etudes en Sciences, Paris, France, les réimpressions edition, 1967.

[4] J. Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983.

[5] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. DeRose. Toolglass and Magic Lenses: The see-through interface. In *Computer Graphics (ACM SIGGRAPH '93 Proceedings)*, volume 27, pages 73–80, Aug. 1993.

[6] M. D. Byrne. Reading vertical text: Rotated vs marquee. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, volume 10, pages 1633–1635, 2002.

[7] F. Chevalier and S. Diamond. The use of real data in fine arts for insight and discovery: Case studies in text analysis. In *IEEE VisWeek Discovery Exhibition*, 2010.

[8] J. Clark. Clustered word clouds. http://neoformix.com/2008/ClusteredWordClouds.html, Oct. 2008.

[9] C. Collins, F. B. Viégas, and M. Wattenberg. Parallel tag clouds to explore faceted text corpora. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 91–98, 2009.

[10] J. Dykes, J. Wood, and A. Slingsby. Rethinking map legends with visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):890–899, 2010.

[11] D. S. Ebert and P. Rheingans. Volume illustration: Non-photorealistic rendering of volume models. In *Proceedings of the IEEE Conference on Visualization*, pages 195–202, 2000.

[12] H. Edelsbrunner and R. Waupotitsch. A combinatorial approach to cartograms. In *Proceedings of the Annual Symposium on Computational Geometry*, pages 98–108, 1995.

[13] Y. Hassan-Montero and V. Herrero-Solana. Improving tag-clouds as visual information retrieval interfaces. In *Proceedings of the International Conference on Multidisciplinary Information Sciences and Technologies*, 2006.

[14] D. Heyman. Axis Maps, personal communication, Mar. 2011.

[15] E. Imhof. Positioning names on maps. *The American Cartographer*, 2(2):128–144, 1975.

[16] V. Interrante and C. Grosch. Strategies for effectively visualizing 3D flow with volume LIC. In *Proceedings of the IEEE Conference on Visualization*, pages 421–424, 1997.

[17] K. G. Kakoulis and Ioannis G. Tollis. A unified approach to labeling graphical features. In *Proceedings of the ACM Symposium on Computational Geometry*, pages 347–356, 1998.

[18] T. Kameda and K. Imai. Map label placement for points and curves. *IEICE Transactions on Fundamentals of Electronics, Communications & Computer Sciences*, E86–A(4):835–840, Apr. 2003.

[19] D. A. Keim, S. C. North, C. Panse, and J. Schneidewind. Efficient cartogram generation: A comparison. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 33–36, 2002.

[20] K. T. Kim, S. Ko, N. Elmqvist, and D. S. Ebert. WordBridge: using composite tag clouds in node-link diagrams for visualizing content and relations in text corpora. In *Proceedings of the Hawaiian International Conference on System Sciences*, 2011.

[21] K. Koh, B. Lee, B. H. Kim, and J. Seo. ManiWordle: Providing flexible control over Wordle. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1190–1197, 2010.

[22] A. Lau and A. V. Moere. Towards a model of information aesthetics in information visualization. In *Proceedings of the International Conference on Information Visualization*, pages 87–92, 2007.

[23] R. Laurini and D. Thompson. *Fundamentals of Spatial Information Systems*. Academic Press, New York, 1992.

[24] B. Lee, N. H. Riche, A. K. Karlson, and M. S. T. Carpendale. SparkClouds: Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1182–1189, 2010.

[25] S. Lohmann, J. Ziegler, and L. Tetzlaff. Comparison of tag cloud layouts: Task-related performance and visual exploration. In *Proceedings of INTERACT*, volume 5726 of *Lecture Notes in Computer Science*, pages 392–404. Springer, 2009.

[26] A. M. MacEachren. *How Maps Work: Representation, Visualization and Design*. Guilford Press, New York, 1995.

[27] R. Maharik, M. Bessmeltsev, A. Sheffer, A. Shamir, and N. Carr. Digital micrography. *ACM Transactions on Graphics (Proc. SIGGRAPH 2011)*, 30(4):100:1–100:12, 2011.

[28] A. V. Moere and H. C. Purchase. On the role of design in information visualization. *Information Visualization*, 10(4):356–371, 2011.

[29] OSMF. OpenStreetMap. http://www.openstreetmap.org/. accessed March 2012.

[30] A. H. Robinson, J. L. Morrison, P. C. Muehrcke, A. J. Kimerling, and S. C. Guptill. *Elements of Cartography*. John Wiley & Sons, 1995.

[31] D. Roscover. Steven Paul Jobs. http://www.gorosco.com/#438803/Steven-Paul-Jobs.

[32] D. Roscover. Burdened. In *Time Magazine*, Feb. 2010.

[33] L. Shi, F. Wei, S. Liu, L. Tan, X. Lian, and M. Zhou. Understanding text corpora with multiple facets. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, pages 99–106, 2010.

[34] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[35] A. Skupin. From metaphor to method: Cartographic perspectives on information visualization. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 91–98, 2000.

[36] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard. *Thematic Cartography and Geovisualization*. Prentice Hall, third edition, 2009.

[37] N. A. Svakhine and D. S. Ebert. Interactive volume illustration and feature halos. In *Proceedings of the Pacific Conference on Computer Graphics and Applications*, pages 347–354, 2003.

[38] W. Tobler. A computer model simulating urban growth in the Detroit region. *Economic Geography*, 46(2):234–240, 1970.

[39] W. Tobler. Cartograms and cartosplines. In *Proceedings of the Workshop on Automated Cartography and Epidemiology*, pages 53–58, 1976.

[40] E. R. Tufte. *Visual Explanations: images and quantities, evidence and narrative*. Graphics Press, Cheshire, Connecticut, 1997.

[41] F. B. Viégas and M. Wattenberg. Tag clouds and the case for vernacular visualization. *interactions*, 15(4):49–52, 2008.

[42] F. B. Viégas, M. Wattenberg, and J. Feinberg. Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1137–1144, 2009.

[43] W3C. Scalable Vector Graphics (SVG). http://www.w3.org/SVG/. accessed March 2012.

[44] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. TIARA: a visual exploratory text analytic system. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 153–162, 2010.

[45] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 51–58, 1995.

[46] P. C. Wong, P. Mackey, K. Perrine, J. Eagan, H. Foote, and J. Thomas. Dynamic visualization of graphs with extended labels. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 73–80, 2005.

[47] J. Wood and J. Dykes. Spatially ordered treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1348–1355, 2008.

[48] Yahoo. TagMaps. http://tagmaps.research.yahoo.com/. accessed March 2012.

Spatiotemporal Social Media Analytics for Abnormal Event Detection and Examination using Seasonal-Trend Decomposition

Junghoon Chae*
Purdue University

Dennis Thom†
University of Stuttgart

Harald Bosch†
University of Stuttgart

Yun Jang‡
Sejong University

Ross Maciejewski§
Arizona State University

David S. Ebert*
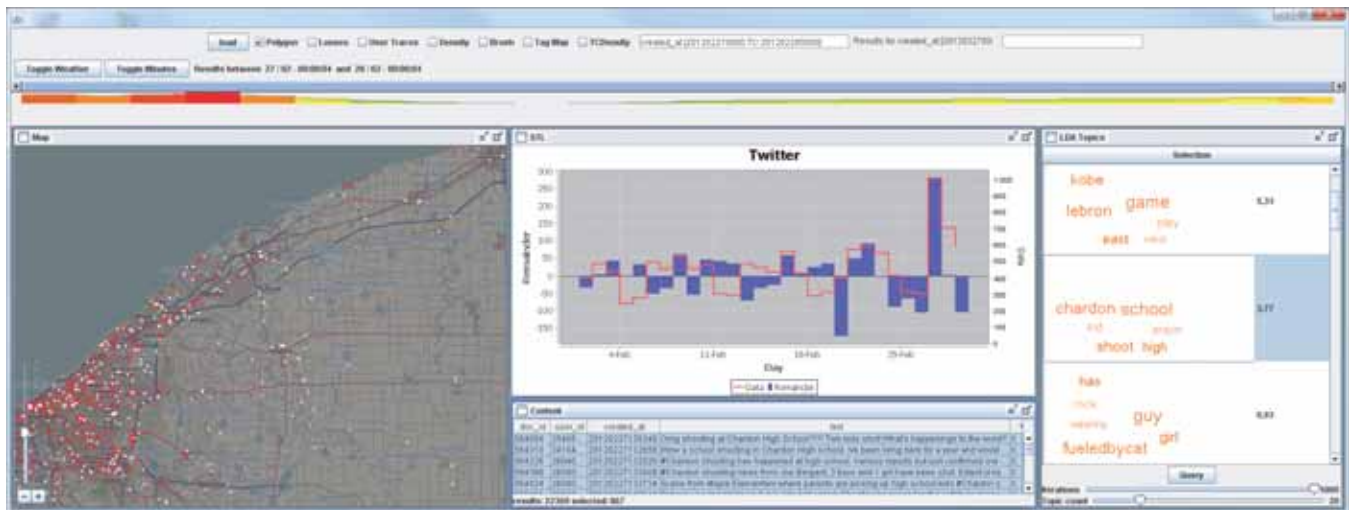Purdue University

Thomas Ertl†
University of Stuttgart

Figure 1: Social media analysis system including message plots on a map, abnormality estimation charts and tables for message content and topic exploration. It can be seen, how the Ohio High School Shooing on February 27, 2012 is examined using the system. The selected messages, marked as white dots on the map, show retrieved Tweets that are related to the event.

Abstract

Recent advances in technology have enabled social media services to support space-time indexed data, and internet users from all over the world have created a large volume of time-stamped, geo-located data. Such spatiotemporal data has immense value for increasing situational awareness of local events, providing insights for investigations and understanding the extent of incidents, their severity, and consequences, as well as their time-evolving nature. In analyzing social media data, researchers have mainly focused on finding temporal trends according to volume-based importance. Hence, a relatively small volume of relevant messages may easily be obscured by a huge data set indicating normal situations. In this paper, we present a visual analytics approach that provides users with scalable and interactive social media data analysis and visualization including the exploration and examination of abnormal topics and events within various social media data sources, such as Twitter, Flickr and YouTube. In order to find and understand abnormal events, the analyst can first extract major topics from a set of se-

lected messages and rank them probabilistically using Latent Dirichlet Allocation. He can then apply seasonal trend decomposition together with traditional control chart methods to find unusual peaks and outliers within topic time series. Our case studies show that situational awareness can be improved by incorporating the anomaly and trend examination techniques into a highly interactive visual analysis process.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—GUI; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Information filtering, relevance feedback

1 Introduction

Social media services, e.g, Twitter, Youtube, Flickr, provide a rich and freely accessible database of user-generated situation reports. As advances in technology have enabled the widespread adoption of GPS enabled mobile communication devices, these reports are able to capture important local events observed by an active and ubiquitous community. The different forms of social media content provided by the users, such as microposts, images or video footage, can have immense value for increasing the situational awareness of ongoing events.

However, as data volumes have increased beyond the capabilities of manual evaluation, there is a need for advanced tools to aid understanding of the extent, severity and consequences of incidents, as well as their time-evolving nature, and to aid in gleaning investigative insights. Due to the

*e-mail: {jchae|ebertd}@purdue.edu
†e-mail: {forename.surname}@vis.uni-stuttgart.de
‡e-mail: jangy@sejong.edu
§e-mail: rmacieje@asu.edu

143

large number of individual social media messages it is not straightforward to analyze and extract meaningful information. For example, in Twitter, more than 200 million Tweets are posted each day [31]. Thus, in a developing event, the relevant messages for situational awareness are usually buried by a majority of irrelevant data. Finding and examining these messages without smart aggregation, automated text analysis and advanced filtering strategies is almost impossible and extracting meaningful information is even more challenging.

To address these challenges, we present an interactive spatiotemporal social media analytics approach for abnormal topic detection and event examination. In order to find relevant information within a user defined spatiotemporal frame we utilize Latent Dirichlet Allocation (LDA) [1], which extracts and probabilistically ranks major topics contained in textual parts of the social media data. The ranks of the categorized topics generally provide a volume-based importance, but this importance does not reflect the abnormality or criticality of the topic. In order to obtain a ranking suitable for situational awareness tasks, we discard daily chatter by employing a Seasonal-Trend Decomposition procedure based on Loess smoothing (STL) [5]. In our work, globally and seasonally trending portions of the data are considered less important, whereas major non-seasonal elements are considered anomalous and, therefore, relevant.

However, due to the large volumes of data, the very specific syntax and semantics of microposts and the complex needs of situational analysis, it would not be feasible to apply these techniques in the form of a fully automated system. Therefore, our whole analysis process, including the application of automated tools, is guided and informed by an analyst using a highly interactive visual analytics environment. It provides tight integration of semi-automated text-analysis and probabilistic event detection tools together with traditional zooming, filtering and exploration following the Information-Seeking Mantra [24].

The remainder of this document is structured as follows: Section 2 is a review of related work. The automated methods to find and examine unusual topics and events are described in Section 3. In Section 4 we briefly introduce our visual analytics system Scatterblogs, which was already featured in previous works, and explain how the automated methods are integrated within a sophisticated iterative analysis loop. Finally we demonstrate the performance of our system based on selected case studies in Section 5 and discuss the approach in Section 6.

2 Related Work

In recent research, social media services have become a popular and influential data source for many domains. Researchers in the fields of data mining and visual analytics have found through studies among users and domain experts, that the analysis of such data can be essential for spatiotemporal situational awareness [15, 23]. Thus, as the size of social media data increases, scalable computational tools for the effective analysis and discovery of critical information within the data are a vital research topic. This section presents previous work that has focused on spatiotemporal and event related social media analysis.

2.1 Spatiotemporal Social Media Data Analysis

As social media platforms move towards location-based social networks (LBSNs) researchers have proposed various approaches to analyze spatiotemporal document collections, in general, and spatiotemporal social media data, in particular.

VisGets [7] provides linked visual filters for the space, time and tag dimensions to allow the exploration of datasets in a faceted way. The user is guided by weighted brushing and linking, which denotes the co-occurrences of attributes. Further works demonstrate the value of visualizing and analyzing the spatial context information of microblogs for social network users [9] or third parties like crime investigators [22] and urban planners [33]. With Senseplace2, MacEachren et al. [15] demonstrate a visualization system that denotes the message density of actual or textually inferred Twitter message locations. The messages are derived from a textual query and can then be filtered and sorted by space and time. Their work also has shown that social media can be a potential source for crisis management. With ScatterBlogs [2], our own group developed a scalable system enabling analysts to work on quantitative findings within a large set of geolocated microblog messages. In contrast to Senseplace2, where the analysts still have to find and manage the appropriate keywords and filters to gather relevant messages in the high volume of insignificant messages, we propose a semi-automatic approach that finds possibly relevant keywords and ranks them according to their 'abnormality'.

Special LBSN for certain domains, like Bikely[1] and Every-Trail[2] have an even stronger focus on the sharing and tracing of user locations. Ying et al. [37] present various location based metrics using spatial information of these LBSNs to observe popular people who receive more attention and relationships within the network. Similarly, there are many related works for non-spatial temporal document collections, for example IN-SPIRE [36], which is a general purpose document analysis system that depicts document clusters on a visual landscape of topics.

2.2 Social Media Event Detection and Topic Extraction

One of the major challenges in analyzing social media data is the discovery of critical information obscured by large volumes of random and unrelated daily chatter. Due to the nature of microblogging, message streams like Twitter are very noisy compared to other digital document collections. Recently, many researchers have tried to solve this challenge by means of automated and semi-automated detection and indication of relevant data.

Sakaki et al. [23] propose a natural disaster alert system using Twitter users as virtual sensors. In their work, they were able to calculate the epicenter of an earthquake by analyzing the delays of the first messages reporting the shock. Weng and Lee [34] address the challenge by constructing a signal for each word occurring in Twitter messages using wavelet analysis, thereby making it easy to detect bursts of word usage. Frequently recurring bursts can then be filtered by evaluating their auto-correlation. The remaining signals are cross correlated pairwise and clustered using a modularity-based graph partitioning of the resulting matrix. Due to the quadratic complexity of pairwise correlation, they rely on heavy preprocessing and filtering to reduce their test set to approx 8k words. As a result, they detected mainly, large sporting events, such as soccer world cup games, and elections. Our approach, in contrast, provides a set of topics through a probabilistic topic extraction algorithm which can be iteratively applied to subsets and subtopics within user selected message sets.

Lee and Sumiya [14] as well as Pozdnoukhov and Kaiser [19] present methods to detect unusual geo-social events by measuring the spatial and temporal regularity of

---

[1] http://www.bikely.com/
[2] http://www.everytrail.com/

Twitter streams. Lee and Sumiya propose a concept to detect unusual behavior by normalizing the Twitter usage in regions of interests which are defined by a clustering-based space partitioning. However, their results are mainly a measurements of unusual crowd behavior and do not provide further means for analyzing the situation. Pozdnoukhov and Kaiser observe abnormal patterns of topics using spatial information embedded in Twitter messages. Similar to our approach, they apply a probabilistic topic model (Online Latent Dirichlet Allocation) as a means of analyzing the document collection. A Gaussian RBF kernel density estimation examines the geo-spatial footprint of the resulting topics for regularities. The usual message count of identified areas is then learned by a Markov-modulated non-homogeneous Poisson process. The spatial patterns are shown as a static heat map. The resulting system does not provide interactive analytics capabilities.

Recently, researchers have applied LDA topic modeling to social media data to summarize and categorize Tweets [39] and find influential users [35]. Zhao et al. [39] demonstrate characteristics of Twitter by comparing the content of Tweets with a traditional news medium, such as the New York Times. They discuss and adapt a Twitter-LDA model and evaluate this model against the standard topic model and the so-called author-topic model [25], where a document is generated by aggregating multiple tweets from a single user, in terms of meaningfulness and coherence of topics and Twitter messages. In this work, we do not use the author-topic model, since a users Tweet timeline is usually a heterogeneous mixture of unrelated comments and messages and not a homogenous framework of interrelated topics like a traditional document. Furthermore, the evaluation of Zhao et al. [39] shows that the standard model has quite reasonable topic modeling results on Tweets, although the Twitter-LDA model outperforms the standard model. Works from Ramage et al. [20] also show promising results in LDA based Twitter topic modeling by evaluating another type of LDA model (Labeled LDA) [21]. ParallelTopics [8] also extracts meaningful topics using LDA from a collection of documents. The visual analytics system allows users to interactively analyze temporal patterns of the multi-topic documents. The system, however, does not not deal with spatial information, but takes an abnormality estimation into account.

In our previous work [26], we proposed a spatiotemporal anomaly overview based on a streaming enabled clustering approach that is applied for each term in the dataset individually. The resulting clusters can be used to generate a spatially and temporally explorable term map of large amounts of microblog messages as an entry point for closer examination. Even though the scalable event detection and our current approach share the same workbench, they can be used independently as well as complementary. The combination of LDA and STL allows for an ad-hoc analysis of a user selected set of messages regarding the topical distribution of messages and the abnormal presence of topics. Due to this characteristic, it provides an iterative analysis loop for qualitative analysis and drill down operations.

3 Spatiotemporal Social Media Analytics for Event Examination

Since several social media sources recently provide space-time indexed data, traditional techniques for spatiotemporal zooming, filtering and selection can now be applied to explore and examine the data. However, as message volumes exceed the boundaries of human evaluation capabilities, it is almost impossible to perform a straightforward qualitative

analysis of the data. In order to cope with the data volumes, traditional interaction and visualization techniques have to be enhanced with automated tools for language processing and signal analysis, helping an analyst to find, isolate and examine unusual outliers and important message subsets.

To address this issue, we present an interactive analysis process that integrates advanced techniques for automated topic modeling and time series decomposition with a sophisticated analysis environment enabling large scale social media exploration. In part 3.1 of this Section we first explain how the Latent Dirichlet Allocation, a well established topic modeling technique in the information retrieval domain, can be used to extract the inherent topic structure from a set of social media messages. The output of this technique is a list of topics each given by a topic proportion and a set of keywords prominent within the topics messages. In a subsequent step, our system then re-ranks the retrieved topic list by identifying unusual and unexpected topics. This is done by employing a seasonal-trend decomposition algorithm to the historic time series data for each topic, retrieving its seasonal, trending and remainder components. Using a z-score evaluation, we locate peaks and outliers in the remainder component in order to find an indicator of unusual events. While the LDA topic extraction is done primarily for Twitter data, the abnormality estimation is also applied to different social media data sources, such as Flickr and YouTube, for each topic. This is achieved by searching matching entries for each term of a topic and applying the same STL analysis on the resulting time series. The results are available to the analyst for cross validation. The details of this step are described in Subsection 3.2 and the complete detection model is formally described in Subsection 3.3. In Section 4, we describe how powerful tools based on these techniques are used within our analysis environment, Scatterblogs, in order to iteratively find, isolate and examine relevant message sets.

3.1 Topic Extraction

Our monitoring component collects space-time indexed Twitter messages using the Twitter-API. The received messages are preprocessed and then stored in our local database. When users of these services witness or participate in unusual situations they often inform their friends, relatives or the public about their observations. If enough users participate, the communication about the situation constitutes a topic that makes up a certain proportion of all messages within the database, or some messages within a predefined area and timespan. In most cases, however, the proportion will be smaller than that of other prevalent topics, such as discussions about movies, music, sports or politics. In order to extract each of the individual topics exhibited within a collection of social media data, we employ Latent Dirichlet

Rank	Proportion	Topics
1	0.10004	day back school today
2	0.09717	lls bout dat wit
3	0.09443	people make hate wanna
4	0.08226	**earthquake thought house shaking**
5	0.05869	**earthquake felt quake washington**

Table 1: An example of extracted topics and their proportions. We extracted topics from Tweets written on August 23, 2011 around Virginia, where an earthquake occurred on this day. One can see that topics consisting of ordinary and unspecific words can have high proportion values, while the earthquake related topics have a relatively low proportion value.

Number of Iteration Steps in the LDA process		
50	300	1000
foursquare pic hall brooklyn	time back night day	time night nyc day
time night day back	york ave brooklyn btw	york ave brooklyn park
newyork nyc tweetmyjobs finance	pic bar food nyc	**foursquare occupywallstreet mayor ousted**
york brooklyn ave street	**foursquare occupywallstreet park mayor**	newyork tweetmyjobs finance citigroup
york ave park btw	newyork tweetmyjobs finance citigroup	**san gennaro street italy**

Table 2: An example of topic model results depending on the number of iteration steps in the LDA process. The topics are extracted from the Tweets posted in New York City on September 17 and 18, 2011 where the Occupy Wall Street protest movement began and a famous festival, *San Gennaro* occurred. A higher number of sampling iterations provides a better topic retrieval describing the two different events.

Allocation, a probabilistic topic model that can help organize, understand, and summarize vast amounts of information.

The LDA topic model approach, as presented by David Blei et al. [1], is a probabilistic and unsupervised machine learning model to identify latent topics and corresponding document clusters from a large document collection. Basically, it uses a "bag of words" approach and assumes that a document exhibits multiple topics distributed over words with a Dirichlet prior. In other words, the LDA assumes the following generative process for each document: First, choose a distribution over topics, choose a topic from the distribution for each word, and choose a word associated with the chosen topic. Based on this assumption one can now apply a Bayesian inference algorithm to retrieve the topic structure of the message set together with each topic's statistical proportion and a list of keywords prominent within the topic's messages. Table 1 shows an example set of extracted topics resulting from the application of LDA to Twitter data ordered by the proportion ranking. The example social media data was collected from Twitter for the Virginia area on August 23rd. On this day, the area was struck by an earthquake with a magnitude of 5.88. As seen in the table, this earthquake event was captured as a topic within the Twitter messages.

In our system, the MALLET toolkit [18] is used for the topic analysis. Prior to the topic extraction, the stemming algorithm KSTEM by Krovetz [13] is applied to every term in the messages. The results of KSTEM are more readable and introduce fewer ambiguities than the often used Porter stemmer.

For the unsupervised LDA classification and topic retrieval one has to define two parameters: the number of expected topics and the number of iterations for the Gibbs sampling process[10], which is used in MALLET for the topic inference. The number of topics that should be chosen depends on the size of the document collection and the required overview level. A small number of topics (e.g., 10) will provide a broad overview of the documents, whereas a large number (e.g., 100) provides fine-grained results. The number of sampling iterations is a trade-off between computation time and the quality of discovered topics. To illustrate this, Table 2 shows the experimental results of the topic model using a varying number of sampling iterations while the number of topics was set to four. The topics were extracted from Tweets posted in New York City on September 17 and 18, 2011, where a large group of protesters occupied Wall Street in New York City[3]. A topic indicating the Occupy Wall Street protests can be seen when using at least 300 iterations. At the time of these protests, there was also a famous annual festival, the *San Gennaro*[4], occurring in Little Italy.

[3] http://occupywallst.org/
[4] http://www.sangennaro.org/

This can only be seen when using at least 1000 iterations. As shown in Table 2, the topics with 50 iterations do not indicate any meaningful events. The topics with 300 iterations, on the other hand, consist of more distinguishable classes. Finally, the topics with 1000 iterations obviously point out individual events which happened in the city.

3.2 Abnormality Estimation using Seasonal-Trend Decomposition

Abnormal events are those that do not happen frequently and usually they cover only a small fraction of the social media data stream. As shown in Table 1, even during an earthquake episode, highly ranked topics consist of ordinary and unspecific words. The third and fourth ranked topics include words indicating the earthquake event of August 2011: *earthquake felt quake washington*. From this observation in the distributions of ordinary and unusual topics over the social media data, it is necessary to differentiate the unusual topics from the large number of rather mundane topics. In order to identify such abnormal topics, we utilize Seasonal-Trend Decomposition based on locally-weighted regression (Loess) known as STL [5]. For each extracted topic of the LDA topic modeling, our algorithm retrieves messages associated with the topic and then generates a time series consisting of daily message counts from their timestamps. The time series can be considered as the sum of three components: a trend component, a seasonal component, and a remainder:

$$Y = T + S + R \tag{1}$$

Here Y is the original time series of interest, T is the trend component, S is the seasonal component, and R is the remainder component. STL works as an iterative nonparametric regression procedure using a series of Loess smoothers [6]. The iterative algorithm progressively refines and improves the estimates of the trend and the seasonal components. The resulting estimates of both components are then used to compute the remainder: $R = Y - T - S$. Under normal conditions, the remainder will be identically distributed Gaussian white noise, while a large value of R indicates substantial variation in the time series. Thus, we can utilize the remainder values to implement control chart methods detecting anomalous outliers within the topic time series. We have chosen to utilize a seven day moving average of the remainder values to calculate the z-scores, $z = (R(d) - mean)/std$, where $R(d)$ is the remainder value of day d, $mean$ is the mean remainder value for the last seven days, and std is the standard deviation of the remainders, with respect to each topic. If the z-score is higher than 2, events can be considered as abnormal within a 95% confidence interval. The calculated z-scores are thus used as abnormality rating and the retrieved topics will be ranked in the analytics environment according to this estimate.
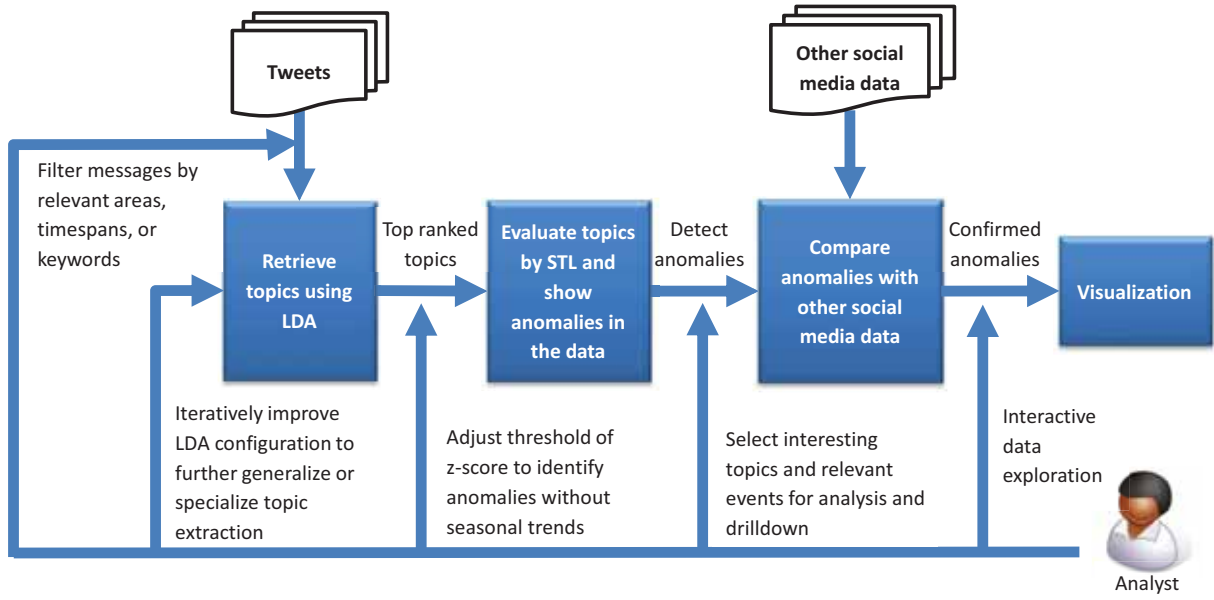
Figure 2: Overview of our iterative analysis scheme for event detection and examination.

3.3 Detection Model

To conclude this section, we formalize our abnormal event detection model based on the probabilistic topic extraction and time series decomposition.

An abnormal event is associated with a set of social media messages that provides its contents, location, and timestamp. To detect abnormal events for a given area and timespan, we define a set called *social spacetime* as follows:

$$S = (T, \Delta time, \Delta area, msgs) \qquad (2)$$

where T is a set of topics, $\Delta time$ is a time period (e.g., one day), $\Delta area$ is a bounded geographical region, and $msgs$ is a set of messages. The user selected parameters $\Delta area$ and $\Delta time$ define the analysis context for which all messages are loaded into the analysis system. In this context, the user selects a subset of messages ($msgs$) for which the LDA topic modeling procedure (described in Section 3.1) extracts the set of topics, $t_i \in T$. Each topic is defined as:

$$t_i = (M_i, W_i, z_i, Y_i, p_i) \qquad (3)$$

where W_i is a set of words describing the topic, M_i is a set of relevant messages, z_i is an abnormality score (z-score), Y_i is a time series, and p_i is a statistical proportion of the topic in $msgs$.

For each topic (t_i), our algorithm searches relevant messages (M_i) in the selected area ($\Delta area$) and time period ($\Delta time$) and a predefined time span of historic data preceding $\Delta time$ (e.g. one month). Messages are considered relevant if they contain at least one word in W_i. From M_i a daily message count time series (Y_i) is generated from the timestamps of the messages. The algorithm decomposes Y_i to obtain a remainder component series using the STL and calculates a z-score (z_i) from the remainder series. Lastly, it sorts the topics based on the z-scores.

For cross validation of each topic, we search for relevant entries in Flickr and YouTube by their meta-data that includes titles, descriptions, tags, and timestamps, using the respective APIs. We repeat the steps for generating a time series from the collected timestamps, applying STL to decompose the time series, and calculating the z-score from the remainder component series.

4 Interactive Analysis Process

The complete topic extraction, abnormality estimation, and event examination are tightly integrated into a highly interactive visual analysis workbench, that allows an analyst to observe, supervise, and configure the method in each individual step. The following sections introduce the details of this system and describe how the event detection is embedded within a sophisticated analysis process as shown in Figure 2.

4.1 Social Media Retrieval and Analysis System

Our modular analysis workbench ScatterBlogs was already featured in previous works [2, 26]. It proved itself very useful for fundamental tasks like collection, exploration and examination of individual, as well as aggregated, social media messages. The UI of the system is composed of several interconnected views and the main view houses a zoomable openstreetmaps implementation showing message geolocations on a world map. The system features a text search engine and visual content selection tools that can be used to retrieve messages, show spatial and temporal distributions and display textual message contents. Additional visualizations and map overlays provide the analyst with powerful inspection tools, such as a kernel-density heatmap similar to [16], to show aggregated and normalized message distributions and a movable lens-like exploration tool (called 'content lens') that aggregates keyterm frequencies in selected map areas [2]. To indicate spatiotemporal anomalies in the message set, the system features a mechanism to detect spatiotemporal clusters of similar term usage, and suspicious message clusters can be represented as Tag Clouds on the map [26]. For the real-time collection of messages using the Twitter Streaming API the system features a scalable extraction and preprocessing component. This component was used to collect Twitter messages since August 2011 and it currently

147

processes up to 20 Million messages per day[5], including the almost complete volume of up to 4 million messages that come with precise geolocation information.

4.2 Visual Topic Exploration and Event Evaluation

Results from the topic retrieval and event detection as described in Section 3 can be iteratively refined by means of visual result presentation and interactive parameter steering. Both, the final result of event detection as well as intermediary findings during data filtering and topic extraction can be used by the analyst to adjust the process in order to identify interesting topics and keyterms as well as relevant map areas and timespans for a given analysis task. New insights can be generated on each of four individual analysis layers which, in conclusion form an iterative analysis loop from data filtering to result visualization:

- **Spatiotemporal Data Filtering:** The analyst selects an initial spatiotemporal context of Twitter messages to be represented in the visualization and to serve as a basis for analysis. He can do so by using textual as well as spatiotemporal query and filter mechanisms that load the relevant base message set from a larger database into active memory. The analyst can further filter the base set and remove unimportant parts by using a time-slider, depicting temporal message densities, or polygon and brush selection tools. Using these tools the analyst can gain an initial impression of the spatial and temporal distribution and location of messages that could be relevant for his analysis task.

- **LDA Topic Examination:** In the subsequent step the analyst can choose to start the topic extraction either on the whole analysis context or on some subset of selected messages. At this stage he can utilize the configuration parameters of LDA extraction to interactively explore available topics by generalization and specialization. In this regard the most important parameter is the number of topics that have to be defined for the topic model inference. If the analyst decreases the number using the provided tools, the extracted topics will be more general. If he increases it, they will be more specific and thus candidates for small but possible important events. Once topics are generated from the data they will be presented to the analyst through a list of small tag clouds for each topic. He can now select the topics from the list to see their individual message distribution on the map and the temporal distribution in the time-slider.

- **STL Evaluation:** Depending on the analyst's choice, the topics can be evaluated and ordered based either on absolute topic frequency or based on abnormality estimates that have been computed using STL. As described in Section 3.2, a valid estimate of abnormality depends on the computation of z-scores from data seven days prior to the observed time frame. Therefore, the STL evaluation will extend the data examination to a range prior to the selected spatiotemporal context, if data is available. Once abnormality is computed for each topic, the topic list will be ordered according to the values and the topics with most outstanding abnormality are highlighted.

- **Crosscheck Validation:** Each selection of messages is accompanied by charts showing the total time series and the remainder components for the selected message set using STL. This is true for spatiotemporal selections as well as for selections using the LDA topic list. In addition to the geolocated Twitter messages this STL is at the same time performed for data that has been extracted from supplemental services like Flickr and YouTube. Based on the multiple charts the analyst can crosscheck the importance and abnormality of examined events and topics.

In our system, the analyst is supposed to iteratively use these means of semi-automated processing, visualization and interaction to refine the selection of messages up to a point where he can begin to examine individual message details. For this task, he can then utilize tools like the content lens for small scale aggregation or the table view to read the messages textual content. The application of these tools is shown in Figure 3. Usually the most valuable messages will be reports from local eyewitnesses of an important event or from insiders for a given topic. Thus, to retrieve large quantities of such messages helping to understand an ongoing event or situation will be the final goal of the iterative process. Unusual topics, suspicious keyword distributions and events with high STL abnormality discovered on the repeatedly traversed analysis layers can guide the analysis from a very broad and general overview to very specific topics and a relatively small message set suitable for detailed examination.

5 Case Study

In this section, we present three case studies for our system covering different types of events including the Chardon High School Shooting, the Occupy Wall Street protests in New York, and the 2011 Virginia Earthquake. The first case shows how analysts can use our system efficiently to find and
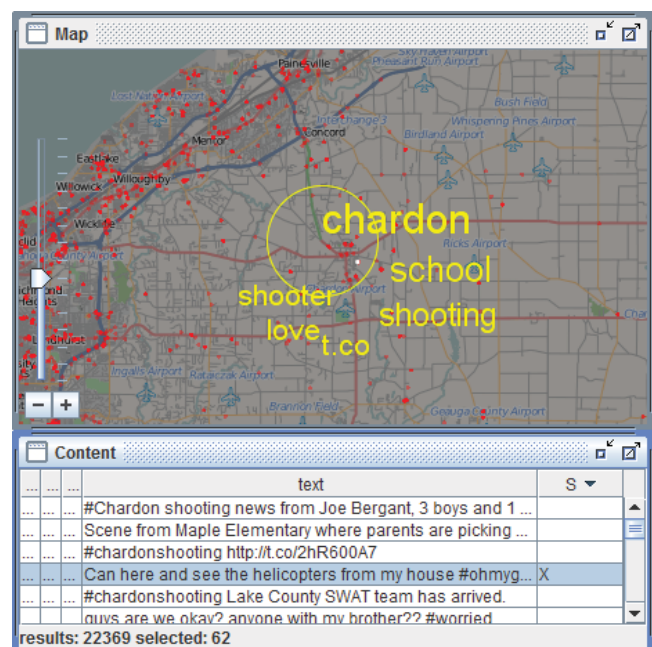
Figure 3: Examining the location of the Chardon high school shooting with a text aggregating content lens.

---

[5]This is just a 10% sample of the total 200 million Twitter messages due to API rate limitations.
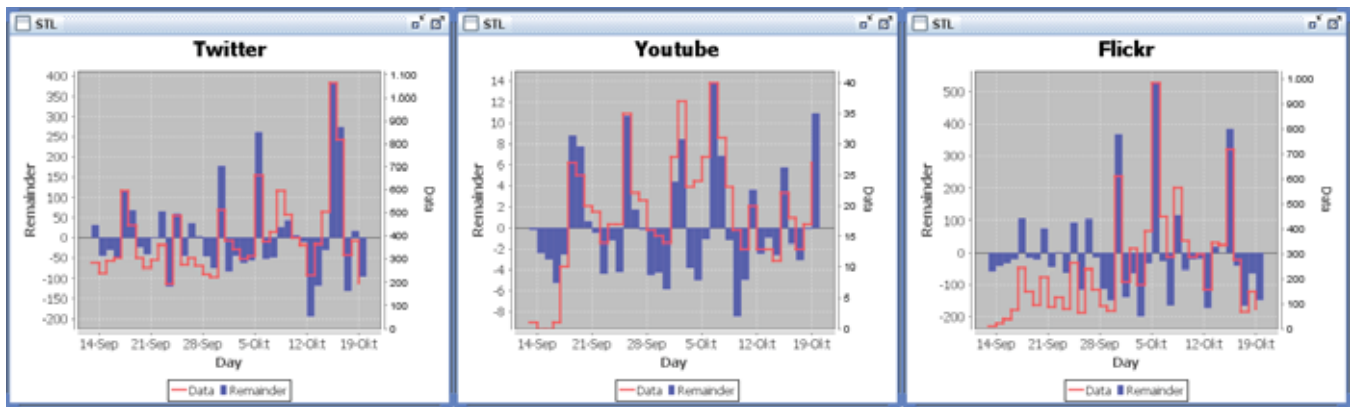
Figure 4: Cross validation of an event using Twitter, Flickr, and YouTube data for the Occupy Wall Street Protests. The protests occurred on Sep. 17 and 30, Oct. 5 and 15. The line charts show the remainder components R (blue) and the original data volumes Y (red) for the STL evaluation. The scales on the right and left side of each chart view are adapted to the maximum values.

explore an abnormal event. The second case highlights the differences between social media types by cross validation of a planned event. Finally, the last example showcases the effects of an abrupt, unexpected, natural disaster.

5.1 Ohio High School Shooting

On February 27, 2012, a student opened fire inside the Chardon High School cafeteria in the early morning. The gunman killed one student and injured four, from which two eventually died after the incident.

To examine this incident we first locate and select the broader Cleveland area on the map and select a time frame covering three days from February 26 to February 28. Using the text search engine and a wildcard query ('*') we can establish an exploration context showing all messages plotted on the map with their respective contents and meta data listed in a separate table view. First, we want to get a broad overview of the topics discussed in the region and thus we select all messages in the area and apply the LDA extraction tool to the current selection. In order to see the most general topics, we chose a low parameter value for the number of topics and a high iteration count to achieve good separation. At this level of semantic detail, the extracted topics indicate messages about the NBA all-star game (February 26 in Orlando) with keywords like *kobe, game, dunk and lebron* as well as the showing of the movie 'The Lion King'on TV with keywords *king, lion, tv*. If we look at the STL-Diagrams of these topics and the computed z-scores, we also see a peak for these events. By clicking on the retrieved topic representations the associated messages are highlighted in each view. By reading some of the message contents (e.g. *'Watching my fav. Movie on ABC family..... Lion King!!!!', 'Can't wait till the dunk contest starts!'*), the analyst can easily disqualify these from further analysis.

To get a higher semantic resolution we can now increase the number of topics and slightly decrease the iteration count in order to achieve a fast computation. By selecting 20 topics, the topic indicating the shooting event is extracted and indicated by keyterms like *shooting, chardon and school*, alongside the other topics. Although the proportion of the topic is not very high compared to the others, the topic receives a very high z-score (i.e., 3.77) and is ranked among the top five topics (highlighted in orange). Figure 1 demonstrates the system view of this observation. An analyst can now select the incident topic to see the spatial distribution

of associated messages on the maps as well as the temporal distribution in the timeslider histogram. By examining messages using the content lens to aggregate topics over map areas as well as the tools for reading individual message contents, we can easily distinguish between messages informed by media reaction and messages of actual observers in the Chardon High School area. In this case, after isolating the messages from local observers, we find messages like *'Omg shooting at Chardon High School?!?!'* and *'Helicopter overhead. We are on scene. Message from school says students moved to middle school'*.

5.2 Occupy Wall Sreet

Starting on September 17, 2011 in the Wall Street financial district in New York City, people have been gathering for the Occupy Wall Street protest movement. The movement against economic inequality has since spread to other major cities throughout the world. Various social media services including Twitter, Facebook, Flickr and Youtube have been utilized both by the participants and the global media for communication and reports about the movement in forms
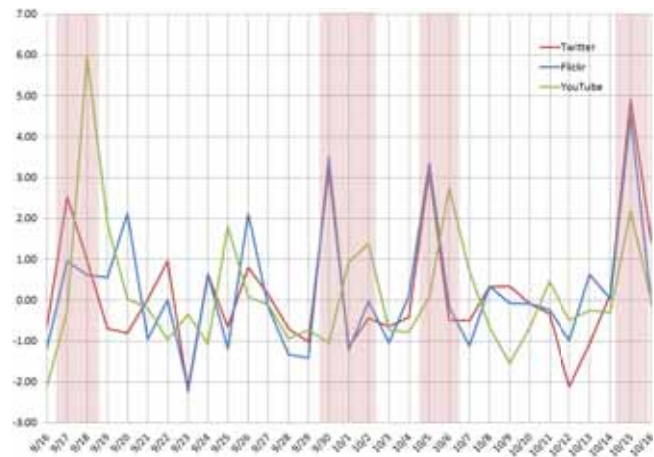
Figure 5: Abnormality and correlation on multiple social media sources. As a result of high z-scores around the same time periods, we found a strong correlation between the three social media sources. Marked regions correspond to periods where at least 2 providers received scores over 2.0.
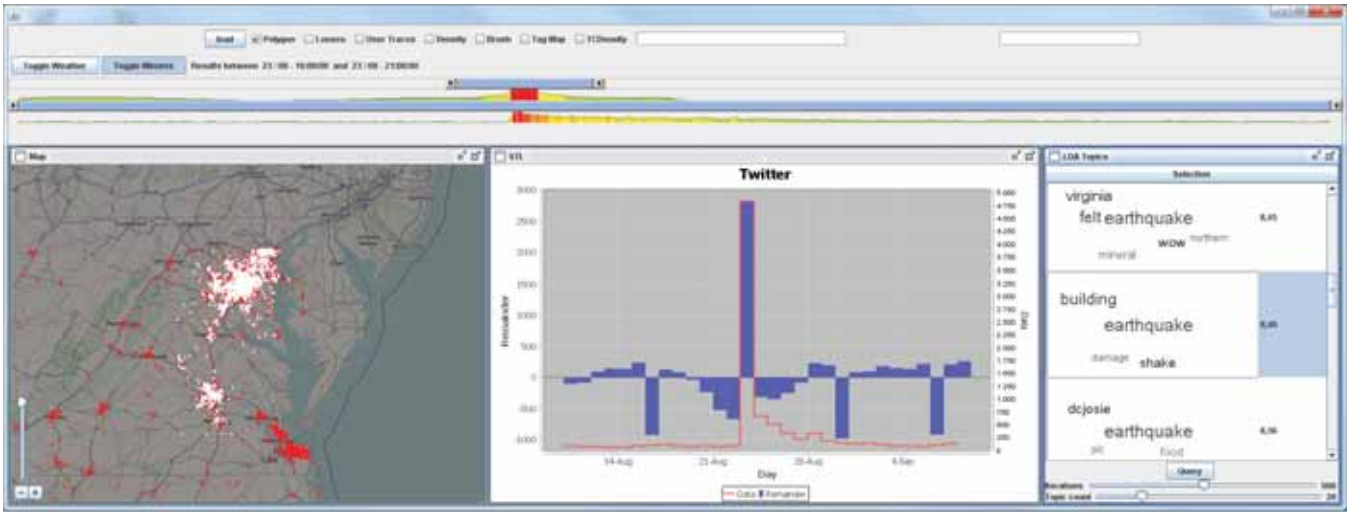
Figure 6: Virginia earthquake on August 23rd, 2011. Our abnormal event detection system detects the earthquake event using our STL based anomaly detection algorithm. The abnormality degree is extremely high on August 23rd, 2011 (times are given in UTC).

of text, images and videos. For the related extracted topic (*occupywallstreet, wall, takewallstreet, takewallst, park*), Figure 4 shows the results of our abnormality estimation for the three social media services Twitter, Flickr, and YouTube over the course of one month. As shown in Figure 5, in each of the marked regions, at least two of the services show z-scores over 2.0 and they correspond to actual events during the Occupy Wall Street protests. From this experimental result, one can derive a strong correlation between the three social media data sources. The related data volumes and remainder (R) are shown in Figure 4 for all three providers.

As shown in Figure 5, on September 17 (the first day of the protests with approximately 1,000 participants [30]), only the Twitter stream received an abnormal score while the Flickr and YouTube data artifacts are delayed by 1-3 days. We attribute this initial delay to the simple nature of Twitter usage compared to Flickr and YouTube where the data potentially has to be recorded, edited, and uploaded and is thus more labor intensive. Additionally, eighty protesters where arrested while marching uptown on September 24, but even though Flickr and YouTube reaction on this event created higher z-scores in the following days, they were not significant enough to register an event. The following spikes of high z-scores overlap with a march across the Brooklyn Bridge (Oct. 1 [29]), a large demonstration (Oct. 5 [27]), and globally coordinated protests (Oct. 15 [28]).

5.3 2011 Virginia Earthquake

For the last use case we examine a magnitude 5.8 earthquake that occurred on the afternoon of August 23rd 2011 in Mineral, Virginia [32]. Starting with the minute of the earthquakes occurrence, Twitter users posted more than 40.000 earthquake-related Tweets reporting tremors they felt along the East Coast [11]. Among these were messages like: *'EARTHQUAKE!!!!!!!'*; *'Whoa!!!! Just experienced an earthquake here in Virginia!!!!'*; and *'Omg I just felt an earthquake'*. Figure 6 gives an impression how our system is applied to examine this event.

For the analysis we begin with selecting the Virginia area from Baltimore to Virginia Beach and three days around the 23th. A topic extraction with 5 topics and just 100 iterations already retrieves two earthquake related topics showing

that this event is very prominent within the selection. By clicking these topics one can observe that the highest density of earthquake messages can be found in the Washington, Baltimore and Richmond areas.

To observe the areas in more detail we combine the topic selection with a spatial selection of the three cities and reapply the topic extraction. This time we use 20 topics with 500 iterations. Since we are now operating only on earthquake related messages, the retrieved topics all contain earthquake as a dominant keyword. On this level of detail we can see topics indicating that buildings have been evacuated due to the earthquake (*earthquake, people, evacuated, early, building*) and that damage has been caused (*earthquake, building, shake, damage*). The z-scores for all top ranked topics are now very high (often above 8.0) and thus indicate the high abnormality of this event.

Finally, when going into even higher detail with 100 topics and 1000 iterations we can see smaller events within the big earthquake event. For example, one topic indicates that damage was caused to the Washington Monument and by clicking on the topic we can see messages like *'damage to Washington Monument'*; *'Washington Monument is tilting?!? '*; and *'Helicopter just landed next to Washington Moniment, west side. #DCearthquake '*. There are also misleading messages, indicating that the damage to the Washington Monument was just false rumors: *'the Washington monument was not damaged in any way from the earthquake. #rumor'*. However, media crosschecks show that visible damages did in fact happen and will probably cost the city 15 million dollars to repair[6].

At this point, it is important to note, that while several earthquake topics produced significant z-values in Twitter, the event did not produce high z-scores in Flickr and YouTube. This is probably due to fact, that many people will write a quick message after a shock has been felt by themselves, but it takes quite some time until images or videos are uploaded from cameras to Flickr and YouTube. The event also demonstrates that large and unexpected events will produce immediate and significant reactions in

---

[6]http://www.huffingtonpost.com/2012/03/14/washington-monument-did-e_n_1344422.html

services like Twitter and they can thus easily be detected by using our system.

6 Discussion

In this section we want to discuss four important notes and observations relevant to the presented approach.

Event Types: As was demonstrated with the three case studies, events in social media can be categorized into two different types. The 2011 Virginia Earthquake and the Ohio High School Shooting can be categorized as abrupt or disaster events, while Occupy Wall Street can be considered a social and planned event. The two types of events have quite distinguishable features. For the abrupt events, there is a strong change in daily counts mainly in the text based Twitter messages. For the planned event, the Twitter signal may still be faster, but due to the gradual increase and decrease, it is less pronounced. In contrast, Flickr and YouTube have delayed, but very prominent changes, for planned events; however, we could not find significant signals for abrupt events. This reflects that video and photo recording happen rarely during abrupt events. Social events, e.g., Occupy Wall Street or election debates, however, have a high impact on such multimedia based social media; Relevant videos, photos, and even meta-data (e.g., descriptions, tags) allow analysts to find additional information about them. We, therefore, think that cross validating events among multiple social media types is important in order to establish situational awareness.

Base Data: Regarding the base data, it is important to note, that our approach depends on geo-located Twitter messages with precise coordinates, which are only a fraction of the whole Twitter stream. While this fraction still consists of several million messages per day, it is not a representative sample of the population, because it mainly covers mobile users equipped with GPS enabled devices. We think, however, that mobile users, who share their daily experiences freely, are the most relevant group for situational awareness scenarios. Some studies [4, 17] tried to overcome the problem of location information scarcity in Twitter messages, which adds another source of uncertainty. First, the user's self reported locations can be outdated. Second, the geo-coding of the location can be considerably wrong due to place name ambiguities. Furthermore, we have just shown the feasibility of the approach for Twitter, Flickr, and YouTube data, but it can easily be adapted to other social media providers like Facebook or Forsquare as well, in order to widen the sample of the population.

Probabilistic Models: In this work, we use STL to decompose time series of topic streams. There are many alternative statistical models for this task, such as DHR (Dynamic Harmonics Regression) [38] and SARIMA (Seasonal AutoRegressive Intergrated Moving Average) [3]. DHR and SARIMA models are particularly useful for forecasting and STL can also be used for prediction based on seasonal (periodic) time series [12]. Our main reasons for choosing STL was the fact that it is non-parametric, can be computed faster than SARIMA [12] and needs less training data for equally good results.

End User Feedback: We requested informal feedback from users within our institutes and received comments and suggestions. To compare the LDA topic modeling plus the seasonal-decomposition based abnormality analysis versus only the LDA topic modeling, we enabled our system to switch between these modes. The users were impressed by the fact that both results (two lists of topics) from two different modes were quite different. Highly ranked topics by LDA

topic modeling consisted of ordinary words, while the combined analysis was indicating unusual events. They noted that the tightly integrated visual analysis workbench was useful to apply the automated methods. Furthermore, they suggested a function allowing people to see a pattern of abnormality for a user-defined topic.

7 Conclusion

In this paper, we presented an interactive abnormal event detection and examination system for the analysis of multiple social media data sources. The system uses an abnormality estimation scheme based on probabilistic topic modeling and seasonal-trend decomposition to find and examine relevant message subsets. This scheme is tightly integrated into an highly interactive visual analytics system, which supplements tools based on automated message evaluation with sophisticated means for parameter steering, filtering and aggregated result set exploration. Three use cases demonstrated the visualization and user interaction within the system and its capabilities to detect and examine several different event types from social media data. The ability to crosscheck findings based on three distinct social media sources revealed the kinds of correlations that can be expected from various event types.

For future work, we will further investigate context-based analysis and improve the current detection algorithm to allow for a faster analysis. Due to the fast-paced and low quality nature of microblogging, we will also investigate the effects of additional preprocessing options like automated spell-checking or synonym recognition under the constraint of preventing ambiguities. Furthermore, we want to supplement the system with real-time monitoring features, demanding additional means for adaptive attention guiding as well as interaction techniques for use in high pressure environments. For the final system we are currently preparing a thorough evaluation to test it in cooperation with crisis management personnel and other domain experts.

References

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.

[2] H. Bosch, D. Thom, M. Worner, S. Koch, E. Puttmann, D. Jackle, and T. Ertl. Scatterblogs: Geo-spatial document analysis. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 309 –310, 2011.

[3] G. E. P. Box and G. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, 1990.

[4] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 759–768, New York, NY, USA, 2010. ACM.

[5] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess (with discussion). *Journal of Official Statistics*, 6:3–73, 1990.

[6] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.

[7] M. Dörk, S. Carpendale, C. Collins, and C. Williamson. VisGets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Information Visualization 2008)*, 14(6):1205–1212, 2008.

[8] W. Dou, X. Wang, R. Chang, and W. Ribarsky. Paralleltopics: A probabilistic approach to exploring document collections. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 231 –240, oct. 2011.

[9] K. Field and J. O'Brien. Cartoblography: Experiments in using and organising the spatial context of micro-blogging. *Transactions in GIS*, 14:5–23, 2010.

[10] T. Griffiths and M. Steyvers. Finding scientific topics. In *Proceedings of the National Academy of Sciences*, volume 101, pages 5228–5235, 2004.

[11] L. Indvik. East coasters turn to twitter during virginia earthquake. Retrieved March 30, 2012, `http://mashable.com/2011/08/23/virginia-earthquake/`, 2011.

[12] B. Jiang, S. Liang, J. Wang, and Z. Xiao. Modeling modis lai time series using three statistical methods. *Remote Sensing of Environment*, 114(7):1432–1444, 2010.

[13] R. Krovetz. Viewing morphology as an inference process. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '93, pages 191–202, New York, NY, USA, 1993. ACM.

[14] R. Lee and K. Sumiya. Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, LBSN '10, pages 1–10, New York, NY, USA, 2010. ACM.

[15] A. MacEachren, A. Jaiswal, A. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford. Senseplace2: Geotwitter analytics support for situational awareness. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 181 –190, oct. 2011.

[16] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert. A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):205–220, Mar. 2010.

[17] J. Mahmud, J. Nichols, and C. Drews. Where is this tweet from? Inferring home locations of twitter users. In *International AAAI Conference on Weblogs and Social Media*, 2012.

[18] A. K. McCallum. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

[19] A. Pozdnoukhov and C. Kaiser. Space-time dynamics of topics in streaming text. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '11, pages 1–8, New York, NY, USA, 2011. ACM.

[20] D. Ramage, S. Dumais, and D. Liebling. Characterizing microblogs with topic models. In *ICWSM*, 2010.

[21] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning. Labeled lda: a supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP '09, pages 248–256, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[22] E. Roth and J. White. Twitterhitter: Geovisual analytics for harvesting insight from volunteered geographic information. In *Proceedings of GIScience*, 2010.

[23] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.

[24] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336 –343, sep 1996.

[25] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths. Probabilistic author-topic models for information discovery. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 306–315, New York, NY, USA, 2004. ACM.

[26] D. Thom, H. Bosch, S. Koch, M. Woerner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *IEEE Pacific Visualization Symposium (PacificVis)*, 2012.

[27] N. Times. Major unions join occupy wall street protest. Retrieved June 25, 2012, `http://www.nytimes.com/2011/10/06/nyregion/major-unions-join-occupy-wall-street-protest.html`, 2011.

[28] N. Times. Occupy wall street protests worldwide. Retrieved June 25, 2012, `http://www.nytimes.com/2011/10/16/world/occupy-wall-street-protests-worldwide.html`, 2011.

[29] N. Times. Police arresting protesters on brooklyn bridge. Retrieved June 25, 2012, `http://cityroom.blogs.nytimes.com/2011/10/01/police-arresting-protesters-on-brooklyn-bridge`, 2011.

[30] N. Times. Wall street protest begins with demonstrators blocked. Retrieved June 25, 2012, `http://cityroom.blogs.nytimes.com/2011/09/17/wall-street-protest-begins-with-demonstrators-blocked`, 2011.

[31] Twitter. 200 million tweets per day. Retrieved March 1, 2012, `http://blog.twitter.com/2011/06/200-million-tweets-per-day.html`, 2011.

[32] United States Geological Survey (USGS). Magnitude 5.8 - virginia. Retrieved March 30, 2012, `http://earthquake.usgs.gov/earthquakes/recenteqsww/Quakes/se082311a.php`, 2011.

[33] S. Wakamiya, R. Lee, and K. Sumiya. Crowd-based urban characterization: extracting crowd behavioral patterns in urban areas from twitter. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '11, pages 77–84, New York, NY, USA, 2011. ACM.

[34] J. Weng and B.-S. Lee. Event detection in twitter. In *International AAAI Conference on Weblogs and Social Media*, 2011.

[35] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 261–270, New York, NY, USA, 2010. ACM.

[36] P. C. Wong, B. Hetzler, C. Posse, M. Whiting, S. Havre, N. Cramer, A. Shah, M. Singhal, A. Turner, and J. Thomas. IN-SPIRE Infovis 2004 contest entry. In *IEEE Symposium on Information Visualization*, Oct. 2004.

[37] J. J.-C. Ying, W.-C. Lee, M. Ye, C.-Y. Chen, and V. S. Tseng. User association analysis of locales on location based social networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '11, pages 69–76, New York, NY, USA, 2011. ACM.

[38] P. C. Young, D. J. Pedregal, and W. Tych. Dynamic harmonic regression. *Journal of Forecasting*, 18(6):369–394, 1999.

[39] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li. Comparing twitter and traditional media using topic models. In *Proceedings of the 33rd European conference on Advances in information retrieval*, ECIR'11, pages 338–349. Springer-Verlag, Berlin, Heidelberg, 2011.

Simon Fraser University, CA

SIMON FRASER UNIVERSITY

Evaluating Analytic Performance

Linda T. Kaastra

University of British Columbia
MAGIC, 3645 FSC Building, 2424
Main Mall, UBC, Vancouver BC
Canada V6T 1Z4
+1 (858) 652 2811
lkaastra@magic.ubc.ca

Richard Arias-Hernandez

Simon Fraser University
250-13450 102 Ave., Surrey BC
Canada V3T 0A3
+1 (778) 960-8974
ariasher@sfu.ca

Brian Fisher

Simon Fraser University
250-13450 102 Ave., Surrey BC
Canada V3T 0A3
+1 (858) 381 7001
bfisher@sfu.ca

ABSTRACT

In this position paper we propose a performance science approach to evaluation of visual analytics systems.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Measurement, Design, Experimentation, Human Factors.

1. INTRODUCTION

The visual analytics approach challenges the visualization community to understand how graphical visualizations can help to solve real-world problems. This requires us to conduct research that will help us to better understand the reasoning processes of expert human analysts as they take place in a given environment to accomplish a specific task. We must also understand algorithmic methods of solving problems and how human and computational analytics can be combined most effectively through highly interactive visualizations.

In this paper we would like to draw a distinction between an *analysis*, which is situated in a particular problem-solving environment, a given analyst, problem, dataset, and evaluation criteria, and an *analytics*, which we think of as the validation of a set of methods for solving a class of problems under different situations, by different analysts, and so on. From the perspective of formal logic, an analytical approach can be considered valid insofar as it is truth-preserving, i.e. an analysis that follows a valid analytic method will, if given true antecedents, generate true results. In cases where data are uncertain, these methods can be used as well to determine the consistency of data and in so doing to identify data that might be erroneous or misleading. While a formal logic (e.g. predicate calculus) holds over a broad class of problems and environments, a more focused analytic method might be effective only on a small class of applications or situations. Modifications of classical logic such as modal logic have advantages in that they can address a particular problem (e.g. deontics for reasoning about laws and regulations), may be more computationally tractable (e.g. in closed world versus open world systems) and may avoid assumptions that do not hold in a given situation (e.g. fuzzy-logic's rejection of the principal of the

excluded middle). For the purpose of visual analytics, one goal of our effort must be to find ways of validating visual analysis processes that are sufficient to provide a warrant (ala Toulmin [1]) for a decision-maker to have confidence that they safely may take action based on that particular analysis process.

In cases were performance is critical, e.g. air travel, health care, and economic policy, those who choose systems for actual use should demand strong evidence that the model is appropriate and internally consistent under the specific conditions (e.g. that variables in the model are reasonable approximations of the real conditions, and discretization of problems takes place at a reasonable grain).

The design of visual analytic support technologies imposes additional demands on validation. While good decisions are the final measure of success, we will require other, more sensitive metrics of cognitive performance to design these mixed-initiative visual analysis systems, for a variety of reasons. First, we cannot be certain that the types of problems, forms of data, or performance conditions of problem solving will be consistent. Any changes in those parameters for a given situation might well render our metric of success invalid for predicting performance in this new class of problems/situations.

For formal logics, such as the predicate calculus, an analytics can be proven. However, for more constrained, situated logics that are employed by experts in their area of expertise, this is more difficult. These situated reasoning processes cannot be validated without an understanding of the structure of the task and the environment in which it takes place. Thus visual analytics requires us to conduct a good deal of field research in order to understand professional expertise, including aspects of expertise that the experts themselves are unaware of: perceptual expertise and learning, pattern recognition, visuospatial attention, cognitive "chunking", as well as biases and reasoning errors that might impact accurate decision-making.

In previous work we have addressed challenges to the design and evaluation of visual analytics processes (and by implication, technologies that support them): unconscious processing at both the general cognitive and perceptual cognitive level, individual differences (both innate and acquired, e.g. perceptual expertise), and how human cognitive agents distribute cognitive processing among themselves for optimal coordination. These have implication for assessment of training and organizational communication structures as well as for design of technology that supports these processes. Together with our colleagues we have called for a reconceptualization of visual analytics as a translational cognitive science that has as its goal the coordination of multiple disciplinary perspectives (e.g. data mining, statistics, modeling, psychology and decision science) to design and evaluate not only technologies but also training and communication methods.

For this workshop we propose to build on previous work on multi-agent coordination of cognitive processes for visual analytics by viewing the coordination of action as a performance. Seen through the lens of Performance Science, user actions have both

communicative and pragmatic aspects. As in musical performance [2] actions taken by emergency managers in a multi-display situation room have at the same time communicative purposes-- they are seen and interpreted by others-- as well as their pragmatic function (e.g. timing individual activities in coordinated team workflow and using human gestures and/or material artifacts as team coordination devices [5, 6, 7, 8, 9]). We propose an approach to understanding cognition in analysis that focuses on the interaction of humans and information at multiple levels of activity, along multiple procedural tracks, and including multiple layers of meaning. The goal is to concentrate on coordination of action as well as the signaling aspects of individual actions to better understand the joint activity of analysis.

We take musical performance as a metaphor for performance in analysis, arguing that a highly skilled visual analyst would bring to bear some of the same perceptual, cognitive, interactive, and motor performance adaptations that expert musicians develop in the course of their training. Much of this expertise is tacit, and is learned through a long and complex process of training. Research methods developed in the Performance Science community for examining tacit knowledge in expert performance are adapted and used to examine performance of both the solo "instrumental" analyst and for an "analytic ensemble" of two or more analysts who coordinate their actions in time as they collaborate on understanding a situation, making decisions, and collaborating in the performance of analysis.

We propose to first apply performance science methods to our "pair analytics" studies on aircraft safety that we conduct with the Boeing Company, later integrating them with our more complex analysis of multi-agent coordination in emergency management with the Provincial and City Emergency Operations Centres in the Lower Mainland of British Columbia.

In the pair analytics method, a structured analysis session pairs a "visual analysis expert" or VAE (who is trained in the tradecraft of visual analysis as well as the use of the tools) with a domain-specific "subject matter expert" or SME (e.g. Boeing aircraft safety engineer) to address a visual analysis problem using a preselected set of visualization tools. Video and audio recording capture movements, voice, gesture, and posture of the participants, while screen capture gives a clear time-synched view of interaction with the visual analytics tool-- usually a commercial product such as Tableau, IN-SPIRE, or Starlight. Because the analysis is carried out collaboratively, individual contributions to the analysis are made transparent through the joint activities of analysis: conversation, gesture, and tool use. The individuals thus can be said to "perform analysis" more or less as a duet.

We use a combination of open and axial coding methods from Grounded Theory [4] structured by a general framework for understanding spoken communication called Joint Activity Theory (JAT) developed by Herbert H. Clark and presented in his 1996 book "Using Language"[3]. Clark proposed that the study of language must take into account the social units and activities that give rise to conversation. He gives the example of music performance as a situation in which the coordination of activity is the primary goal - while playing a duet requires each musician to know their part, the work of rehearsal is geared specifically on the coordination that enables the performers to act as an ensemble. Clark argues that conversation also involves coordinated effort – the speaker and listener coordinate their attention, their activity, and their signaling of intent and understanding to the other party. The strength of JAT is in the concepts that enable a field researcher to analyze the tacit processes of coordination in skilled collaboration.

We adapt Clark's framework to take into account the coordination of communicative and pragmatic action by participants in technology-rich environments, in particular for the collaborative use of visualization systems. For our purposes, the key aspects of Clark's framework are: levels of action, tracks, layers of meaning, common ground, coordination devices, and Event Boundaries.

Taken together, they provide a powerful lens for interpreting collaborative activity in pair analytics.

Levels of Action

Human interactions never take place in a single line [3]. For example, in music, if someone plays a C, that action can take place at 4 levels. The player executes the necessary actions to perform middle C. They present "C" as a signal. They signal, with the presentation of "C" that they are beginning the first movement *sotto voce*, at a slower tempo than last time. The "C" is a proposal for the group to join in and play, and/or for the people in the room to participate as expected in that situation – a lesson, a master class, a performance, an impromptu concert, and so on. Thus, there are four levels of musical activity in collectives:

> Executing behavior
> Presenting signal
> Signaling common ground to another actor
> Proposing joint project

We analyze action in visual analytic activity the same way. For example, if the VAE moves a screen capture to the history bar in TABLEAU, he is presenting a signal "capture screen." He signals to the SME: "by capturing this screen and placing it in the history of analysis I acknowledge the mutual understanding that we are moving the analysis to a new analytical project."[5] The screen capture action is a performance of the transition to a new phase of analysis.

Tracks

A conversation takes place in primary and secondary tracks. The primary track includes the content of the discussion. The secondary track is the process by which the content is being negotiated, and includes vertical and horizontal conversation markers. In the pair analytics data, we found that conversation markers consist of words, inflections, gestures (with head and hands), and onscreen gestures [5].

Layers of meaning

Meaning can be negotiated in more than one layer. There is a basic layer, where the actions are executed, but the meaning might be understood at a higher layer. This is the case in staged actions as well as personal interactions that include jokes, lying, sarcasm, and verbal irony. In the paired analytics data, the paired analyses are always taking place in the context of data collection. As mentioned earlier, this creates a performance situation where individual analytical activity must be communicated either directly through speech, or indirectly through gesture, positioning, inflection, and tool use. This offers a window on some of the tacit processes of collaborative analysis.

Event Boundaries and Common Ground

Each activity has a negotiated entrance, body, and exit. In order to analyze what is happening at any moment in the data, we start by identifying the events and their boundaries. Then we can proceed to label the tacit processes involved in the negotiation. For example, a handshake begins when one person signals, with posture and possibly speech, that a handshake is about to happen. This is followed by the handshake itself, which concludes when one party signals their intent to move on by releasing their grip. This same pattern of events can be seen in a range of coordination activities such as an initial approach to the counter of a store, placing items on the table, paying with a credit card, or bagging and leaving. However, joint activities can only take place if participants assume certain knowledge on the part of the other person. Participants have a set of expectations and proceed to advance their understanding by drawing on joint perceptual

experience or joint actions. We call this "the negotiation of common ground" [3].

Common ground is a concept with a rich history in the study of mind. Of the three possible representations for common ground, we use the one presented in Clark as the most basic (CG-Shared), therefore the one with the broadest application for analyzing joint activities.

CG – Shared

Common ground (shared basis)

p is common ground for members of community C if and only if:

> 1. every member of C has information that basis *b* holds
> 2. *b* indicates to every member of C that every member of C has information that *b* holds.
> 3. *b* indicates to members of C that *p*.

Common Ground is *not* a representation for what individuals know in a joint activity. It is the *basis* that each person uses as justification for participating in the activity. Clark says, in practice, people can justify the basis for their common ground, in part based on the *quality of evidence* (what we can infer from the evidence).

The principle of justification – In practice, people take a proposition to be common ground in a community only when they believe they have a proper, shared basis for the proposition in that community [3].

Common ground can be perceptual, procedural, and content specific; and it can be personal or communal. *Perceptual common ground* is the things that I am aware that you are also aware of (the more formal term is "joint salience") that form the basis for our interaction. *Procedural common ground* is the things that I am aware that you know (shared task representations) about how our interaction should proceed. *Content specific common ground* is the stuff that I am aware that you know about the topic we are discussing or the music we are playing. *Personal common ground* is the things that a person might share with only a select group or even just one person. This common ground can be built up over years of living with a person in the case of family, or through a personal exchange. *Communal common ground* includes all of the social knowledge we share with the many different groups to which we belong. We use this to categorize people and this process largely shapes how we interact with people we do not know. Until we cultivate personal common ground, we operate from these larger categories of belief, awareness, knowledge and assumptions [3].

Of course these things are constantly being negotiated. The VAE cannot know for sure that the SME is exactly aware of something in the way he believes. But there is a justified basis for believing so, and a justified basis for using that to move toward greater clarification, coming closer to whatever they want to achieve through the joint activity. This basis structures how both parties attend to aspects of the environment, and what they attend to, what they make decisions about, and what they do when they participate in the activity. These processes are often tacit.

Coordination Devices

Coordination of activities takes place around jointly salient aspects of the shared task. The jointly salient aspects include verbal speech, actions, shared representations, and shared tools for analysis. These aspects of the shared task environment also meet the requirements of the personal, shared, and public goals for the activity [5].

Conclusion

Taken together, the concepts of JAT allow researchers to investigate the multi-dimensional nature of collaboration around a visual analytic task. The SME and VAE work together, building a shared task representation and negotiating common ground as they employ complex visualization tools. Their performance with the data need not be seen as a tangled web of interactions. Rather, the contents of their interactions with each other and the visualization tools offer an intriguing glimpse of the deeply situated and emergent nature of analytical activity.

If our goal is to evaluate performance analytics in collaborative analysis we must begin by conceptualizing analysis as a joint activity. This broadens the scope to include both the products of analysis and the process through which it takes place. This is compatible with the distributed cognition perspective taken by Hutchins (1995). Metrics for the evaluation of coordination per se include fluidity of the process, timing, and coordination – how the activities of analysis progress. Our adaptation of Clark's Joint Activity Theory and pair analysis protocol is the first step toward a framework that can address coordination of action on a larger scale. We are beginning this work in our laboratory with a focus on collaborative decision-making in public health as well as emergency management.

2. ACKNOWLEDGMENTS

3. REFERENCES

1. Toulmin, S. (1958). The Uses of Argument. Cambridge University Press. Cambridge, UK.

2. Kaastra, Linda. "Annotation and the Coordination of Cognitive Processes in Western Art Music Performance." International Symposium on Performance Science. Toronto, Ontario, 2011.

3. Clark, H. H. (1996) Using Language. Cambridge University Press. Cambridge, UK.

4. Strauss, A. & Juliet Corbin. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (2nd Ed.). SAGE Publications. Thousand Oaks, CA.

5. Arias-Hernandez, R, Kaastra, L.T., and Fisher, B. (2011) Joint Action Theory and Pair Analytics: In-vivo Studies of Cognition and Social Interaction in Collaborative Visual Analytics. In L. Carlson, C. Hoelscher, and T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 3244-3249). Austin TX: Cognitive Science Society.

6. Goodwin, C. & Goodwin, M., Formulating Planes: Seeing as a Situated Activity, In: Y. Engestrom & Middleton, D. (Eds.), Cognition and Communication at Work, Cambridge University Press, Cabridge, MA, 1996

7. Harper, R. & Hughes, J., 'What a F-ing System! Send 'em all to the same place and then expect us to stop 'em hitting': Making technology work in air traffic control. In G. Button (Ed.), Technology in working order: Studies in work, interaction and technology, Routledge, London, 1993, pp 127-144

8. Heath, C. & Luff, P., "Collaboration and control: Crisis management and multimedia technology in London Underground line control rooms," Journal of Computer-Supported Cooperative Work (CSCW) 1(1), Springer Netherlands, 1992, pp. 69-94.

9. Hutchins, Edwin. (1995). Cognition in the Wild. MIT Press. Massachusetts, USA.

From Cognitive Amplifiers to Cognitive Prostheses: Understandings of the Material Basis of Cognition in Visual Analytics

RICHARD ARIAS-HERNANDEZ, TERA M GREEN and BRIAN FISHER

School of Interactive Arts and Technology, Simon Fraser University

The most salient ways in which data visualization and interactive techniques have been understood as the material basis of cognition in the emergent field of visual analytics are discussed. Three main dominant understandings have captured the imagination and theorizations of researchers and technicians in this field: data visualizations and interactive techniques as cognitive amplifiers, cognitive prostheses, and cognitive mediators. The analysis of this treatment of materiality in cognition provides an up-to-date report on whether remarks on the situated character of cognition and the active role of human agents have, in effect, been incorporated in this field or not. We argue that even though visual analytic researchers have incorporated some of the ideas of situated cognition and tempered traditional arguments of information processing from cognitive science, understandings of the role of materiality in cognition are still marked by universalisms and ascriptions of exacerbated agency to visual representations.

KEYWORDS cognitive studies of technology, distributed cognition, action theory, visual analytics

Introduction

Defined as 'the science of analytical reasoning supported by interactive visual interfaces' (Thomas and Cook 2005), visual analytics (VA) is a post-9/11 technoscientific endeavour. Its research agenda was drafted in 2004 by visualization researchers from academia, industry, and government laboratories on the request of the US Department of Homeland Security. The initial objective of VA was to address analytical challenges in intelligence analysis and public safety posed by the threat of terrorism and natural disasters. As dramatically illustrated by the events of 9/11, intelligence analysis failed to connect the dots among dispersed yet available information,

which, if appropriately considered, could have prevented those attacks in the US. In retrospect, this weakness in intelligence analysis was explained in terms of the human and technological inability to cope with the information overload produced by enormous amounts of constantly generated, intelligence-related data.

During the first years of development of VA, researchers concentrated their efforts in the development of computationally based, interactive, visual representations as well as on the development of several assumptions about analytical cognition during visual analysis (Pirolli and Card 2005). The foundational assumption in VA is that, in contrast to computation alone, visualization can harness the human mind's innate 'visual intelligence' to gain novel insights into situations characterized by complex data that may contain uncertainty in fact or relevance to the problem, or time and location of occurrence. Two basic types of computer-based artefacts have been the focus of design and development in VA to harness this visual intelligence: visual representations and interaction techniques.

By developing computer-based visual representations of data that are designed to trigger instinctive perceptual responses, VA is expected to create systems that allow the human analyst to offload cognitive processes, such as comparing values or detecting outliers, into the more basic pre-cognitive, sensorial system (Card *et al*. 1999). For example, an analyst may visually mark outliers with different colours, shapes or sizes. This kind of offloading of cognition emphasizes the use of visualizations as external representations, which are coupled with the analyst's internal knowledge representations.

By designing interactions with these visualizations, VA aims at extending these systems to allow the analyst to offload some other smaller cognitive processes, such as juggling the memory of multiple search results or annotating the strength of a relationship between two visualized concepts into what are essentially outputs of the motor system. There are, for example, the outputs of saving the state of an analytical process or creating annotations about the reasoning process. These artefacts of interaction can be both real-time, such as mousing behaviours and menu interaction, as well as asynchronous, such as when an analyst creates notes for future use by herself or by other analysts who will use the visualization afterwards.

In this study, we discuss how these two kinds of artefacts, visualizations and interaction techniques, have been understood as the material basis of cognition in VA. Our main argument is that these understandings still rely on traditional cognitive models that focus on universalisms and assumptions of humans as passive cognitive agents while downplaying recent models that emphasize the situatedness and active role of humans in tight couplings with external representations-processes. For example, mainstream cognitive models in VA assume a homogeneous, universal, and rather passive cognitive agent that couples with visual analytic systems to augment her cognitive skills (Liu and Stasko 2010). More recently, however, research in the cognition of visual and spatial representations has challenged this view showing that cognitive agents display a wide variety of cognitive behaviours when interacting with visual analytic systems (Keehner *et al*. 2008). This expression of heterogeneous, situated and active cognitive agents that negotiate the terms of their coupling to external representations problematizes the mainstream cognitive models used in VA. Our discussion centres around what we consider are the three more dominant understandings of the materiality of cognition in VA: cognitive amplifiers, cognitive prostheses, and cognitive

mediators. The first two understandings are related to theories of distributed and external cognition and the third understanding is related to Activity Theory. The following sections present these understandings in details, highlight similarities and differences, and discuss their implications for changing notions of cognition. We conclude this study with an analysis of potential implications for revitalizing these understandings with more assertive notions of human agency and situatedness.

Living in a material world: theories and understandings of the material basis of cognition in InfoVis and visual analytics

VA is an outgrowth of two closely related disciplines: information visualization (InfoVis) and scientific visualization (Wong and Thomas 2004), and most of its current understandings about the material basis of cognition have been taken from psychology and cognitive paradigms adopted in these fields. In this section, we review the uptake of two of these paradigms: distributed cognition and Activity Theory, and the way they have shaped understandings of the materiality of cognition in VA.

Cognitive amplifiers: InfoVis and visual analytics take on distributed cognition

The distributed cognition paradigm has been developed in cognitive sciences (Norman 1993; Zhang and Norman 1994; Larkin and Simon 1987) and in cognitive anthropology (Lave 1988; Hutchins 1995) since the mid 1980s. Its main tenet is that human cognition is a phenomenon that is not bounded to internal states located in the human brain. On the contrary, human cognitive processes emerge as a result of the tight coupling of internal (i.e. mental) representations-processes with external representations-processes (Scaife and Rogers 1996). The external representations are normally embodied in physical form in a material object, or artefact (e.g. maps, charts, compasses, and stars used in naval navigation). The external processes can be implemented either by artefacts (e.g. calculators, computers) or by other human agents. According to this paradigm, an understanding of cognition cannot be limited to internal, mental representations but it should also include interactions among humans, and interactions between humans and artefacts. In other words, cognition is not the exclusive result of processes occurring inside the human brain but the result of interactions across individuals of a social group and across individuals and their physical environment.

Within the theories and models of distributed cognition, two perspectives can be distinguished. First, consider cognitive anthropological perspectives of distributed cognition, such as Hutchin's and Lave's. These perspectives take as their unit of analysis a sociotechnical *system*, in which cognitive processes, such as memory or problem solving, are distributed in a network of collaborative agents (humans and nonhumans). These approaches emphasize as much the interaction between social agents as they emphasize the interaction between individuals and artefacts. In other words, these analyses do not privilege individuals over social groups, or individuals over artefacts. It is rather the system as a holistic unit that captures the attention of the researchers. Methodologically, these approaches are also characterized by their use of ethnography.

A second perspective, also known as 'external cognition' (Scaife and Rogers 1996), corresponds to cognitive science approaches to distributed cognition. Donald Norman, J. H. Laird, and Herbert Simon, among other cognitive scientists, have heralded this approach. Central to external cognition theories is the concept proposed by Norman of 'cognitive artefacts' (Norman 1993). Cognitive artefacts are external aids invented by humans for the purpose of overcoming limits imposed in memory, thought, and reasoning. One class of cognitive artefacts is graphical inventions of all sorts, such as maps or data charts. Before the 1990s, traditional cognitive modelling relied heavily on internal representations, such as mental models (Johnson-Laird 1986), to explain cognitive phenomena. However, external cognition approaches have been instrumental in shifting the attention of researchers towards the interaction, or 'coupling', between internal and external representations (i.e. cognitive artefacts) in the production of naturalistic cognitive behaviour.

Different from cognitive anthropological perspectives, external cognition perspectives do not emphasize social interaction or cultural influences in their models. The unit of analysis is not a sociotechnical system including social structures, culture, individuals and tools, but rather a more limited interaction between the internal representations of *one* individual and external representations (i.e. cognitive artefacts). External cognitive approaches also tend to assume universal properties of individuals and artefacts, instead of accounting for situated variability. In other words, the 'system' in external cognition approaches is a more reductionist interaction between a standard individual interacting with one more or less deterministic artefact. External cognitive perspectives tend to privilege individuals over social groups, and cognitive artefacts over individuals.

From these two approaches to distributed cognition, the external cognition approach developed by cognitive scientists has been the one adopted by InfoVis and VA communities, and the one approach that has mostly shaped their research and design agenda (Dykes *et al*. 2005, Campbell *et al*. 2008, Liu *et al*. 2008). This perspective has also been the strongest influence to sediment understandings of interactive, visual representations as 'cognitive amplifiers'. For example, the very definition of InfoVis states that its purpose is 'the use of computer-supported, interactive, visual representations of abstract data to *amplify cognition*' (our italics, Card *et al*. 1999, 7). This strong emphasis on interpreting interactive, visual representations, as amplifiers of cognition was grounded by Card *et al*. in their seminal book 'Readings in Information Visualisation: Using Vision to Think' (1999).

Card *et al*. presented external cognition as 'the way in which internal and external representations and processing weave together in thought' (Card *et al*. 1999, 1). They argued that external, visual representations are aids that extend human cognitive abilities (e.g. working memory and computation) and amplify cognitive performance. To illustrate this point, Card *et al*. compared how long it takes for individuals to multiply a pair of two-digit numbers in their heads versus doing the multiplication in longhand using pencil and paper. They argued that the reduction in time by a factor of five produced in the pencil–paper condition results from the aid of the visual representation. Card *et al*. argued that one of the challenges of mental multiplication is to hold partial results in working memory, until they can be used. By using the visual representation, individuals are not required to hold partial results in their memory. Instead, an individual would simply write those numbers down. Additionally, the visual structure produced by neatly aligning numbers

in columns speeds up the retrieval of information when partial results are required. In so doing, the visual representation 'extends a person's working memory' (Card *et al*. 1999, 2). Moreover, the authors claimed that some artefacts such as slide rules, calculators or computers, which outperform the unaided mind in computational processing, could easily externalize the necessary calculations or transformations required by complex multiplications, augmenting an individual's computational abilities.

Further, Card *et al*. argued that *visualizations* could amplify cognition in six major ways: '(1) by increasing the memory and processing resources available to users, (2) by reducing the search for information, (3) by using visual representations to enhance the detection of patterns, (4) by enabling perceptual inference operations, (5) by using perceptual attention mechanisms for monitoring, and (6) by encoding information in a manipulable medium' (Card *et al*. 1999, 16). This argument was consistent with research done by Larkin and Simon (1987), Norman (1993), Kirsh and Maglio (1994), Kirsh (1995), and Scaife and Rogers (1996). Indeed, Larkin and Simon's (1987) research on using diagrams to solve physics problems argued that diagrams help an analyst with searching, recognition, and inference in three ways: (1) by grouping together information that is used together, large amounts of search are avoided, (2) by grouping data about an object, visualizations can avoid symbolic labels, leading to reductions in search and working memory, and (3) by offloading cognitive inferences done symbolically into inferences done with simple perceptual operations.

In a similar vein, Norman's influential book 'Things that make us smart' (Norman 1993) advanced the concept of cognitive artefacts and their use in experiential and reflective thought. He also supported the argument of using visualizations as aids to expand the working memory available for solving a problem. According to Norman, visual aids facilitate the reflective process 'by acting as external memory storage, allowing deeper chains of reasoning over longer periods of time than it would not be possible without the aids' (Norman 1993, 25). Kirsh's research on the Tetris game (1995) also highlighted the potential of using interactive visual interfaces to offload mental processes such as rotations of spatial icons, which are computationally expensive, into perceptual and motor processes with are faster and more accurate. Kirsh (1995, 65) argued that using visual–spatial arrangements (1) reduce memory load of tasks; (2) reduce the number of steps involved in internal computation; and (3) simplify visual search and categorization. Scaife and Rogers (1996) followed on Norman's argument by supporting the idea that visual representations could be used as external repositories of information to free up working memory for other aspects of thinking. Internal representations when using visual aids, they argued, do not need to be complete but rather can allow working memory to keep a minimum amount of information about the physical location of relevant points on the visual display and pointers to other equally important locations. In other words, the representation is distributed in internal dynamic pointers *and* external and detailed, visual information (Scaife and Rogers 1996, Pylyshyn 2003).

Since VA is an outgrowth of InfoVis, understandings of the role of interactive, visual representations as 'cognitive amplifiers' were also adopted in VA, and these ideas have become mainstream within the discipline. The most representative example of this point is found in the research agenda for VA drafted in 2004 and synthesized in the book 'Illuminating the Path' (Thomas and Cook 2005). This research agenda starts by endorsing Card

et al.'s perspective of interactive, visual representations of data as 'amplifiers of cognition' (Thomas and Cook, 2005, 46–47), and throughout the agenda there are consistent interpretations of visual aids as 'augmenting cognition' or as 'amplifying cognition', such as:

> 'Information visualisation amplifies human cognitive capabilities'. (Thomas and Cook 2005, 46)

> 'Augmenting the cognitive reasoning process with perceptual reasoning through visual representations permits the analytical reasoning process to become faster and more focused'. (Thomas and Cook 2005, 69)

> 'Visual representations are the equivalent of power tools for analytical reasoning'. (Thomas and Cook 2005, 70)

This treatment of interactive, visual representations as cognitive amplifiers in InfoVis and VA has implications. Here we discuss two. First, even though theories of external cognition emphasize the system created by the coupling of internal and external representations as the unit of analysis, the overemphasis of external representations in VA and InfoVis has systematically underplayed internal representations and interaction/coupling in understandings of external cognition (Yi *et al*. 2007, Liu, Nersessian, and Stasko 2008). The treatment of other components of the distributed cognitive system, such as interaction techniques and mental models, has remained underdeveloped in these fields. More importantly, by neglecting internal cognitive processes, interaction and coupling with external visual representations has resulted in the taken-for-granted assumption of agency granted to visualizations in processes of amplification of cognition. In other words, assuming that visualizations per se are cognitive amplifiers automatically grants them agency and control of the cognitive process over other components of the system — internal representations, motor interaction, and coupling of internal/external representations. This philosophical view assumes that, with visualizations, humans automatically and unconsciously, offload internal mental processes into external perceptual and motor processes. However, this may not be so. More nuanced theories of external cognition have challenged this view by arguing that human agency drives motor processes that make changes to external representations in order to save computations in the head (Kirsh 1995); conventions of visual representation have to be internalized for their effective use (Hegarty 2011); and internal, cognitive control of interactive behaviour minimizes effort by using a least effort combination of the mechanisms available to it rather than automatically expending perceptual-motor efforts to conserve lesser amounts of cognitive efforts (Gray and Fu 2004; Gray *et al*. 2006). What these cognitive scientists have highlighted is that practitioners of InfoVis and VA have privileged external representations as amplifiers of cognition, when research has shown that there is little reason to think that (1) people automatically offload internal cognitive processes into visualizations or that (2) visual representations have a privileged status in relation to other components of external cognition (e.g. mental models, perception, motor interaction, coupling of internal/external representations, etc.).

A second aspect is that cognitive science researchers in external cognition generally acknowledge that cognitive artefacts in fact *do not* amplify cognition (Larkin and Simon 1987, Norman 1993, Hutchins 1995, Kirsh 1995, Gray and Fu 2004). For them, the amplification of cognition is a misleading notion.

Rather, what happens in distributed cognitive systems is that cognitive tasks *change* once external aids are incorporated (i.e. cognitive tasks change to perceptual and/or motor tasks). When we use cognitive artefacts such as pencil and paper, maps or other visual representations, our internal working memory does not increase, nor does the speed of our internal computational processes. Rather, it is the nature of the task at hand that changes. For example, in the case of the previous illustration of the multiplication of a pair of two-digit numbers, mental multiplication usually involves the following cognitive tasks: computation of intermediary numbers, holding partial results and carriers in working memory, and retrieval of partial results when needed for further calculation (e.g. addition). When a cognitive artefact such as pencil and paper is used, the same goal is achieved by performing *different tasks*: perception and motor systems (i.e. vision and use of the hand) are used to write down the numbers on the paper and align the numbers in columns; long term memory is used to recall results of multiplications of single digits; perception-motor systems are used to write down temporary results and carriers; visual perception is used to retrieve partial results from the paper; and mental computation is used for simple addition of single digits. As this example illustrates, it is not that working memory amplifies when pencil and paper are used, rather unaided cognitive tasks that required intense use of working memory get simplified or are no longer necessary, once external aids are available. Hutchins puts it this way: it is not that the cognitive properties of individual minds get amplified; it is rather the effective coupling of individuals and cognitive artefacts that produces cognitive properties of a system that are more effective than those produced by the individual minds alone (Hutchins 1995). Even though this more reserved view of augmentation of cognition seems to be widespread in the cognitive science community, practitioners of InfoVis and VA still seem to subscribe to the idea that interactive visualizations in fact amplify cognition (Liu *et al*. 2008).

Cognitive prosthetics in visual analytics: from universal cognitive abilities to individual differences

Another understanding of the role of interactive visualizations in cognition in InfoVis and VA is that of 'cognitive prostheses'. Understandings of interactive visualizations as cognitive prostheses are directly derived from those of cognitive amplifiers. An interactive visualization is a cognitive prosthesis when it is designed and used to supplement a cognitive limitation or to reconstitute a cognitive ability that is considered to be impaired. As is the case with prosthetic devices that are designed to replace or supplement body parts, a cognitive prosthesis requires a standard of 'normal' faculties, skills or abilities to compensate for. For example a standard of vision of 20/20, which could be measured to discriminate among individuals' state of vision, is used to determine if someone's vision is falling behind what should be expected to be the ideal condition. The purpose of glasses, a prosthetic device, is then to correct for this limitation so that artefact-aided vision gets closer to the ideal standard. Similarly, for cognitive prostheses to compensate for limitations in cognitive faculties or abilities, some standards have to be determined so that gaps in individual capabilities can be found and diagnosed.

As problematic as these standards or normalizations are in their work in society as regulatory devices (Foucault 1995, Bowker and Star 1999), they are central to the work of modern science. Cognitive science, for example, has advanced several metrics to determine standards for cognitive skills and

abilities that are 'hard wired' in humans and are assumed to be universally distributed. Some examples of these are: preattentive processing rates for form, colour, motion and spatial position (Ware 2004); Miller's famous 'magic' number: 7 ± 2 'chunks' that can be held in working memory (Miller 1956); subitizing ranges that capture a feeling of immediately knowing how many items lie within a visual scene (Kaufman *et al*., 1949); and the small number of visual elements or objects (4–6) that can be indexed (FINSTed) in early stages of visual perception (Pylyshyn 1989). Characteristic of these standards is the tendency to configure a prototypical human cognitive agent, very similar to the standard individual that it is assumed in understandings of cognitive amplifiers. Another strategy to determine differences in cognitive abilities is using statistical analyses to discriminate among groups that respond to cognitive task better (i.e. faster or more accurate) than others. For example, using tests, such as Guay and McDaniels' visualization of viewpoints (Keehner *et al*. 2008), for scoring spatial ability — the ability to mentally store and manipulate visual–spatial representations accurately. These sorts of tests allow researchers to score individuals and statistically group them as having high or low abilities.

Understandings of computer-generated visual representations of data being used as cognitive prostheses are less common in InfoVis than those of cognitive amplifiers, and are normally found under the label of 'assistive technologies' (Lamming *et al*. 1994). For example, Nugent *et al*. (2008) and Alm *et al*. (2007) have developed stationary and mobile devices that include visual and interactive interfaces as test beds to validate their use as memory aids for people suffering from mild dementia. Czerswinski *et al*. (2004) redesigned the Windows XP bar to support automatic generation of visual reminders to resume tasks after interruptions and reduce the difficulty of multitasking in some users. In addition to these examples, InfoVis and human–computer interaction (HCI) has had a long tradition of developing customizable visual interfaces for dyslexic people, shortsighted people, and the elderly.

In VA, a few applications have also been designed as cognitive prostheses. Arnott *et al*. (2006), for example, designed interactive 'galaxy' views in the user interface of an augmentative and alternative communication (AAC) system as a cognitive aid for non-speaking individuals. The visualization was designed to support individuals with severe communicative impairments in retrieving information from a biographical AAC database. Another line of development in VA that intersects with understanding of cognitive prostheses has clearly come from the domain of intelligence analysis. Researchers of intelligence analysis (Heuer 1999, Cooper 2005) have proposed a set of 'analytic pathologies' that could be overcome with the help of VA tools. Among these pathologies are: ignoring high-profit documents due to data overload (Patterson *et al*. 2001); confirmation bias (Johnston 2005); oversensitivity to consistency, persistence of impressions based on discredited evidence, illusory correlations (Heuer 1999); and not considering levels of trust and certainty in sources of data (Johnston 2005). Several of the visual analytic tools developed under the sponsorship of US Department of Homeland Security have been designed to provide computer-support to analysts in order to overcome these so-called analytic pathologies. For example, the Scalable Reasoning System (Pike *et al*. 2009) is a visual analytic tool that provides features to ascribe levels of certainty to pieces of evidence in order to discriminate them during processes of generation and falsification

of hypotheses. Some other tools have been tailored to specific methods that help overcome some other cognitive biases. For example, the ACH tool developed by the Palo Alto Research Center provides a quick visual aid to help on the analysis of competing hypotheses and in the reduction of confirmation bias.[1] More recently, Green *et al.* (2009) have proposed using artificial intelligence (AI) agents in mixed-initiative VA systems to help human analysts 'neutralize' confirmation bias and provide an automatic recommendation system for consideration of alternative hypotheses:

> 'Through observation of what interests the human collaborator, the computer can suggest information that is semantically related, but up to this point, has not been considered ... The human is free to explore or to reject suggestions. But by making the effort in ensuring that nothing important is overlooked, the computer works to counteract human cognitive biases that can interfere with complete mental modelling'. (Green *et al.* 2009, 3)

Most of these understandings of artefacts as cognitive prostheses in InfoVis and VA have relied on similar assumptions to those discussed in the section of cognitive amplifiers. The reason for this, as mentioned before, is that understandings of cognitive prostheses are corollary to theories of augmentation of cognition. Therefore, a similarly strong emphasis has also been placed in granting agency to the visual artefacts. As long as an adequate prosthesis is matched to an individual with the cognitive limitation it is supposed to overcome, then the artefact should do the trick. These assumptions again are not fully supported.

Contradictory evidence about using interactive visual aids as cognitive prostheses has also come from the cognitive sciences. For example, studying individual differences in spatial abilities, Keehner *et al.* (2008) tested the effect of the spatial ability of a group of individuals on the effects of external, interactive visualizations. They wanted to test if provision of an interactive visualization could work as a cognitive prosthesis for low-spatial individuals who had poor internal abilities to mentally rotate and imagine cross-sections of 3D objects. Their results showed that individuals with low-spatial ability got lower accuracy scores in cross-section tasks than individuals with high-spatial ability, but all of these scores were independent of participants' use or not of interactive visualizations as aids. In other words, there was no evidence that, at least in this particular case, the external visualization could act as a cognitive 'prosthetic' that could compensate for low internal mental abilities.

Cognitive mediators: activity theory in InfoVis and visual analytics

Activity Theory has its roots in Leontiev's (1978) psychological analyses of activity and Vygotsky's (1978) work on child development, from which he derived his insights on the 'zone of proximal development'. In the 1980s, Engeström reformulated the theory and it is in large part this strand of Activity Theory that has been drawn upon by human–computer interaction (Nardi, 1996), and more recently by information visualization and VA researchers (Zhao *et al.*, 2008; Gotz and Zhou, 2009). The aim of Activity Theory is to understand the 'unity of consciousness and activity' (Kaptelinin and Nardi, 2009). Activity Theory is as much a theory of consciousness as it is a theory of activity. Human experience is seen as mediated through signs and

artefacts within an activity system, yet the system is given motive, intentionality and meaning through consciousness, which is uniquely human. In Activity Theory, the main unit of analysis is the *activity* itself.

An activity is composed of a subject, object, actions, operations and mediators. A subject is an individual or social group engaged in an activity. An object is what motivates the activity; the motif or need that shapes action, and it is intrinsic to the subject. An object, in Activity Theory, is not to be confused with a physical artefact. An object is better understood as a 'goal' or, as Kaptelinin and Nardi (2009) suggest, the 'object of the game'. Actions are processes undertaken in pursuit of the object or goal within the activity. Actions are conscious formulations aimed at fulfilling the goal. For example, where the goal may be to get a coffee from your favourite café, an action might be to find your coat to go outside or ask if someone else is going for coffee and offer to buy them a coffee in exchange for going to the café for you. Operations are processes that have become routine and are unconscious. For example, for an experienced driver, changing gears is a routine process. Operations are related to actions, and in fact, the same process of changing gears is an action to a novice driver, i.e. the inexperienced driver is conscious of the goal and the process required to shift from one gear to another. Mediators are external elements that facilitate actions or operations. Mediators can be human or nonhuman. Both, people and tools can act as mediators of actions and operations. Activity Theory holds that the elements of an activity system are not fixed but can change as conditions change. The object is not immutable and can be transformed in the course of an activity, yet it is not the moment-by-moment dynamism of situated actions (Kaptelinin and Nardi 2009).

The notion of mediation through artefacts is a central concept in Activity Theory (Kaptelinin and Nardi 2009). From an InfoViz or VA perspective informed by this theory, interactive visualizations are understood as mediators of activity, and interactions of users — subjects — with interactive visualizations are understood as actions or operations in a 'computer-mediated activity' (Kaptelinin and Nardi 2009). Several InfoVis researchers have incorporated Activity Theory or some conceptual elements of it in their research and design practices. For example, Zhao *et al.* (2008) have mapped movement of population in geovisualizations using activity as the transformational dimension, along with time and space. Matthews (2006) also incorporated Activity Theory in her design of glanceable peripheral displays. She conceptualized peripheral displays as mediators of operations (i.e. unconscious processes). These displays had visual information about secondary tasks during multitasking and could remain at the unconscious level of peripheral vision monitoring (Matthews 2006). The mediation of activity by tools has also been used in VA to capture 'actions' (i.e. conscious processes) that are mediated by human–computer interactions with VA tools. These actions are captured in time sequences and feed visual representations to provide a history of analytical reasoning (Gotz and Zhou 2009; Shrinivasan and van Wijk, 2009). Green *et al.* (2011) have also proposed to understand VA tools as cognitive mediators of analytical activity to guide the interactive design and evaluation of VA tools.

There are several differences and similarities between Activity Theory and external cognition approaches in their treatment of physical artefacts, such as interactive visualizations. Both approaches include material artefacts as part of cognitive activity and ascribe a vital role to them. In other words, they

make technology visible in cognitive theory. However, their treatment of materiality in cognition is different. For Activity Theory, the locus of cognition is *human activity*, not the isolated individual mind or the material artefacts. On the other hand, for external cognition, the locus of cognition is the *systemic interactions* between internal and external representations. Systemic cognition (i.e. augmented cognition) is an emergent property that cannot be ascribed to only one of the components of the system, whether human or nonhuman. This conceptual difference impacts on the level of attention and importance granted to humans, activity, artefacts, and interactions. Activity Theory privileges human activity over artefacts as the source of cognition, treating external artefacts as tools that humans choose to use to support or *mediate* their activities. Activity Theory only treats humans as cognitive agents. External cognition, on the other hand, does not grant any special privileges to any of the components of the system, but rather to the system as a whole as the source of systemic cognition, privileging the system over human and nonhuman components. External cognition distinguishes between systemic cognition and cognition that occurs at the components level as two different levels of cognitive phenomena. By insisting in symmetric treatment of the components of the system, external cognition theories cannot avoid granting artefacts status as cognitive agents.

Another related difference that Activity Theory places on the treatment of artefacts is related to agency. As mentioned in the section of cognitive amplifiers, InfoViz and VA take on external cognition shifted ascriptions of agency from the system 'user-interacting-with-a-visualization' towards the computer-based visual artefact, and skewed ascriptions of agency from the system to the artefact. On the contrary, Activity Theory has a clear and strong position that only ascribes agency to human subjects. The theory's understanding of agency is grounded on the idea of a purposeful subject that orients her activity towards the satisfaction of her needs. The theory clarifies that only living things have needs, and excludes any non-living thing, or artefact, from any sort of claim over agency. From an action theory perspective, a computer-based, interactive visualization cannot be considered as having agency of any sort.

Activity Theory also re-introduces the element of social interaction that is present in anthropological perspectives of distributed cognition but that is absent from external cognition perspectives adopted in InfoVis and VA. Since activity is mediated not only by tools but also by other humans (e.g. Vygotsky's zone of proximal development), Activity Theory gets closer to anthropological versions of distributed cognition, such as Hutchin's, in their inclusion of social groups and culture as constitutive of consciousness and human activity. This is an element that has been underplayed in adaptations of Activity Theory in InfoVis and VA, which tend to focus on partial aspects of the theory, but one that could be relevant for information visualizations scenarios that emphasize collaboration, such as collaborative VA or uses of visualizations in CSCW. It also promises to be a rich and useful approach for what is being currently developed as 'knowledge visualization' (Wang and Mu, 2009, Zhang *et al*. 2010).

Can we bring agency and situatedness to visual analytics?

So far in this paper, we have discussed three understandings of the material basis of cognition in VA: cognitive amplifiers, cognitive prostheses and

cognitive mediators. We highlighted how these understandings usually have found their way first in information visualization and how they have found their way into VA, owing to the big overlap between InfoVis and VA communities. We also explored how some defining theories from cognitive science provided the theoretical underpinnings of these understandings. However, it is necessary to emphasize here again that the translation of these theories from cognitive science to information visualization and VA has not been without modification, adaptation, appropriations and opportunistic selection of partial aspects of these theories. One example that we offered was the treatment of 'augmentation of cognition' by InfoVis and VA, something that is not heralded by cognitive scientists who work on external cognitive theories (Gray *et al*. 2006, Hegarty 2011). This incomplete translation has had as a consequence an overemphasis in ascriptions of agency to external, visual representations and interactive techniques, and a consequent underplay of the active, human cognitive agent. This explains why most of the developments in this field have focused on the design of computer-based visualizations and interactive techniques rather than on studies of human cognition in its interaction with visual analytic tools.

Another aspect that we have found underplayed in VA and that is evident in anthropological perspectives of distributed cognition, action theory, and some cognitive science theories, is 'situatedness'. The concept of situatedness, when applied to cognition (Suchman 1987), refers to the opportunistic character of cognitive activity as produced by the resources of the immediate situation. Situated perspectives of cognition are characterized for their acknowledgement of the particular, their emphasis on culture as shaper of cognition, and their resistance for developing universal models of cognition or theories that abstract cognition from phenomenological experience. Two examples of these perspectives are situated cognition (Suchman 1987) and embodied cognition (Dourish 2001). Cognitive anthropological perspectives of distributed cognition (Hutchins 1995, Lave 1988), Activity Theory (Kaptelinin and Nardi 2009), and some contemporary versions of extended cognition (Gray and Fu 2004, Gray *et al*. 2006) also share the same interest for 'situatedness' of cognition. However, this is one of the aspects that have been lost in the translation of these theories to InfoVis and VA, which tend to overemphasize universal models of human cognition and standard generic conceptions of the human agent. Even adaptations of Activity Theory, such as those of Gotz and Zhou (2009), seem to ignore the aspects of situated activity and ascribe to a fixed set of universal actions to characterize the interactions of users with visual analytic tools. It would seem that information visualization and visual analytic communities cannot deal with the contingency that characterizes human activity and persist in pursuing the quest for standard models and universals to characterize the role of visualizations and interaction in cognition.

The abundant ecology of materiality in cognition in VA, as discussed here,[2] shows how central artefacts are in the mainstream cognitive narratives in this field. From this initial exploration and analysis, it is evident for us that some elements of these narratives of the material basis of cognition have departed from traditional cognitive theories that demarcated binaries between mind and brain and that excluded technology from playing an important role in cognition. However, it is also evident that these same narratives have also been subsumed by deeply grounded universalisms in order to fit into the dominant scientific orders of information visualization and VA. Some

consequences of this have been the intense search for a universal prototype or standards of human cognitive abilities and the quest for a holy grail of general principles for the design of information visualization and VA technology. Another consequence of this commitment to universalism is that other cognitive-related approaches that emphasize situatedness and human agency have received much less attention, development and visibility in the information visualization and the VA community. This may partially explain why some theories from sociology and anthropology, which could be relevant to the understanding of the relations between human cognition and the material world (e.g. symbolic interactionism, situated cognition, and actor-network theory) have not permeated in the theoretical frameworks currently used in VA. It may also explain why the whole field of VA has not been able to attract more social scientists to help shape its research and design agenda.

Notes

[1] 'Analysis of Competing Hypothesis', Palo Alto Research Center, accessed November 18, 2011, http://www2.parc.com/istl/projects/ach/ach.html.

[2] We need to clarify that there are more understandings that inhabit the imaginations of researchers and designers in VA, such as: cognitive affordances, cognitive scaffolds, cognitive extractors, and cognitive armatures. However, we decided to limit this paper only to amplifiers, prostheses, and mediators, and expand our analysis later to encompass other minor understandings.

Bibliography

Alm, Norman, Arlene Astell, Gary Gowans, Richard Dye, Maggie Ellis, Phillip Vaughan, and Alan F. Newell. 2007. An Interactive Entertainment System Usable by Elderly People with Dementia. *Proc. 4th Int. Conf. on Universal Access in Human–Computer Interaction*.

Arnott, John L., Leishi Zhang, David O'Mara, Norman A. Alm, and Andrew Taylor. 2006. Information Visualisation in the User Interface for Augmentative and Alternative Communication. *Technology & Disability* 18(3): 147–161.

Bowker, Geoffrey C., and Susan L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge: The MIT Press.

Campbell, Bruce D., Huseyin O. Mete, Tom Furness, Suzanne Weghosrt, and Zelda Zabinsky. 2008. Emergency Response Planning and Training through Interactive Simulation and Visualisation with Decision Support. *IEEE Proc. Conference on Technologies for Homeland Security* 176–180.

Card, Stuart K., Jock D. Mackinlay, and Ben Schneiderman, eds. 1999. *Readings in Information Visualisation: Using Vision to Think*. San Francisco: Morgan Kaufmann Publishers Inc.

Cooper, Jeffrey R. 2005. *Curing Analytic Pathologies: Pathways to Improved Intelligence Analysis*. Washington, DC: Center for the Study of Intelligence.

Czerwinski, Mary, Eric Horvitz, and Susan Wilhite. 2004. A Diary Study of Task Switching and Interruptions. *Proc. CHI 2004* 6(1): 175–182.

Dourish, Paul. 2001. *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge: The MIT Press.

Dykes, Jason, Alan M. MacEachren, and Menno-Jan Kraak. 2005. Advancing Geovisualisation. In *Exploring Geovisualisation*, edited by Jason Dykes, Alan M. MacEachren and Menno-Jan Kraak, 693–703. Kidlington: Elsevier Ltd.

Foucault, Michel. 1995. *Discipline and Punish: The Birth of the Prison*. New York: Vintage.

Gotz, David and Michelle X. Zhou. 2009. Characterizing Users' Visual Analytic Activity for Insight Provenance. *Information Visualisation* 8(1): 42–55.

Gray, Wayne D., and Wai-Tat Fu. 2004. Soft Constraints in Interactive Behavior: The Case of Ignoring Perfect Knowledge In-The-World for Imperfect Knowledge In-The-Head. *Cognitive Science* 28(3): 359–382.

Gray, Wayne D., Chris R. Sims, Wai-Tai Fu, Michael J. Schoelles. 2006. The Soft Constraints Hypothesis: A Rational Analysis Approach to Resource Allocation for Interactive Behavior. *Psychological Review* 113(3): 461–482.

Green, Tera M., William Ribarsky, and Brian Fisher. 2009. Building and Applying a Human Cognition Model for Visual Analytics. *Information Visualisation* 8(1), 1–13.

Green, Tera M., Ron Wakkary, and Richard Arias-Hernandez. 2011. Expanding the Scope: Interaction Design Perspectives for Visual Analytics. *Proc. of HICSS-44*. Computer Society Press.

Hegarty, Mary. 2011. The Cognitive Science of Visual–Spatial Displays: Implications for Design. *Topics in Cognitive Science* 3(3), 446–474.

Heuer, Richards J. 1999. *The Psychology of Intelligence Analysis*. Washington, DC: Center for the Study of Intelligence.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge: The MIT Press.

Janney, Richard W. 1997. The Prosthesis as Partner: Pragmatics and the Human–Computer Interface. *Proc. IEEE International Conference on Cognitive Technology*: 82–87.

Johnson-Laird, Philip N. 1986. *Mental Models*. Cambridge: Harvard University Press.

Johnston, Rob. 2005. *Analytic Culture in the US Intelligence Community: an Ethnographic Study*. Washington, DC: Center for the Study of Intelligence.

Kaufman, E.L., M.W. Lord, T.W., Reese, and J. Volkmann. 1949. The Discrimination of Visual Number. *The American Journal of Psychology* 62(4): 498–525.

Kaptelinin, Victor and Bonnie A. Nardi. 2009. *Acting with Technology: Activity Theory and Interaction Design*. Cambridge: The MIT Press.

Keehner, Madeleine, Mary Hegarty, Cheryl Cohen, Peter Khooshabeh, and Daniel R. Montello. 2008. Spatial Reasoning With External Visualisations: What Matters Is What You See, Not Whether You Interact. *Cognitive Science* 32(7): 1099–1132.

Kirsh, David and Paul Maglio. 1994. On Distinguishing Epistemic from Pragmatic Actions. *Cognitive Science* 18(4), 513–549.

Kirsh, David. 1995. The Intelligent Use of Space. *Artificial Intelligence* 73(1–2): 31–68.

Lamming, Mik, Peter Brown, Kathleen Carter, Margery Eldridge, Mike Flynn, Gifford Louie, Peter Robinson, and Abigail Sellen. 1994. The Design of a Human Memory Prosthesis. *The Computer Journal* 37(3): 153–163

Larkin, Jill H., and Herbert A. Simon. 1987. Why a Diagram Is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* 11(1): 65–100.

Lave, Jean. 1988. *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge: Cambridge University Press.

Leontiev, A. 1978. *Activity, Consciousness, and Personality*. Englewood Cliffs, N.J.: Prentice Hall.

Liu, Zhicheng, Nancy J. Nersessian, and John T. Stasko. 2008. Distributed Cognition as a Theoretical Framework for Information Visualisation. *IEEE Transactions on Visualisation and Computer Graphics* 14(6): 1173–1180.

Liu, Zhicheng and John T. Stasko. 2010. Mental Models, Visual Reasoning and Interaction in Information Visualisation: A Top-down Perspective. *IEEE Transactions on Visualisation and Computer Graphics* 16(6): 999–1008.

Matthews, Tara. 2006. Designing and Evaluating Glanceable Peripheral Displays. *Proc. DIS 2006*: 343–345.

Miller, George A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63(2): 81–97.

Nardi, Bonnie A. (ed.). 1996. *Context and Consciousness: Activity Theory and Human-computer Interaction*. Cambridge, MA: MIT Press.

Norman, Donald A. 1993. *Things That Makes Us Smart: Defending Human Attributes in the Age of the Machine*. New York: Addison–Wesley.

Nugent, Chris D., Richard J. Davies, Mark P. Donnelly, Josef Hallberg, Mossaab Hariz, David Craig, Franka Meiland, Ferial Moelaert, Johan E. Bengtsson, Stefan Savenstedt, Maurice Mulvenna, and Rose-Marie Dröes. 2008. The Development of Personalized Cognitive Prosthetics. *Proc. IEEE EMBS Conference*: 787–790.

Patterson, E. S., E. M. Roth, and D. D. Woods. 2001. Predicting Vulnerabilities in Computer-Supported Inferential Analysis under Data Overload. *Cognition, Technology & Work* 3(4): 224–237.

Pike, William, Joe Bruce, Bob Baddeley, Daniel Best, Lyndsey Franklin, Richard May, Douglas Rice, Rick Riensche, and Katarina Younkin. 2009. The Scalable Reasoning System: Lightweight Visualisation for Distributed Analytics. *Information Visualisation* 8(1): 71–84.

Pirolli, Peter and Stuart K. Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. *Proceedings of International Conference on Intelligence Analysis*. McLean, VA: Mitre.

Pylyshyn, Zenon W. 2003. Return of the mental image: are there really pictures in the brain?. *Trends in Cognitive Science* 7(3): 113–118.

Pylyshyn, Zenon W. 1989. The Role of Location Indexes in Spatial Perception: A Sketch of the FINST Spatial-Index Model. *Cognition* 32(1): 65–97.

Thomas, James J. and Kristin A. Cook, eds. 2005. *Illuminating the Path: the Research and Development Agenda for Visual Analytics*. Los Alamitos: IEEE.

Scaife, Mike and Yvonne Rogers. 1996. External Cognition: How Do Graphical Representations Work? *International Journal of Human-Computer Studies* 45(2): 185–213.

Shrinivasan, Yedendra B., and Jarke J. van Wijk. 2009. Supporting Exploration Awareness in Information Visualisation. *IEEE Computer Graphics and Applications* 29(5): 34–43.

Suchman, Lucy. 1987. *Plans and Situated Actions*. Cambridge, Mass.: Cambridge University Press.

Vygotsky, L. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, Mass.: Harvard University Press.

Wang, Xiao-yue, and Yan Mu. 2009. Visualisation Based on Concept Maps: An Efficient Way to Knowledge Sharing and Knowledge Discovery in E-science Environment. *Proc. IEEE FSKD 2009*: 144–147

Ware, Colin. 2004. *Information Visualisation: Perception for Design*. San Francisco: Morgan Kaufmann.

Wong, Pak Chung and Jim Thomas. 2004. Visual Analytics. *IEEE Computer Graphics and Applications* 24(5): 20–21.

Yi, Ji Soo, Youn ah Kang, John T. Stasko. 2007. Toward a Deeper Understanding of the Role of Interaction in Information Visualisation. *IEEE Transactions on Visualisation and Computer Graphics* 13(6): 1224–1231.

Zhang, Jianping, Da Zhong, and Jiahua Zhang. 2010. Knowledge Visualisation: An Effective Way of Improving Learning. *Proc. IEEE 2nd Int. Workshop on Education Technology and Computer Science* 1: 598–601.

Zhang, Jiaje and Norman, Donald A. 1994. Representations in Distributed Cognitive Tasks. *Cognitive Science* 18 1): 87–122.

Zhao, J., Forer, P. and Harvey, A.S. 2008. Activities, ringmaps and geovisualisation of large human movement fields. *Information Visualisation* 7: 198–209.

Notes on contributors

Richard Arias-Hernandez is a postdoctoral fellow at the School of Interactive Arts and Technology, Simon Fraser University. His current research focuses on bridging symbolic interactionism and applied psycholinguistics to solve coordination problems in collaborative visual analytics.

 Correspondence to: Richard Arias-Hernandez: ariasher@sfu.ca.

Tera Marie Green is a PhD student and graduate research assistant at the School of Interactive Arts and Technology, Simon Fraser University. Her research involves applied cognitive science for visual interfaces.

Brian Fisher is an associate professor of Interactive Arts and Technology and Cognitive Science at Simon Fraser University and associate director of the Media and Graphics Interdisciplinary Centre at University of British Columbia.

Report from Dagstuhl Seminar 12081

Information Visualization, Visual Data Mining and Machine Learning

Edited by

Daniel A. Keim[1], Fabrice Rossi[2], Thomas Seidl[3], Michel Verleysen[4], and Stefan Wrobel[5]

1   **Universität Konstanz, DE,** `keim@uni-konstanz.de`
2   **SAMM, Université Paris 1, FR,** `Fabrice.Rossi@univ-paris1.fr`
3   **RWTH Aachen, DE,** `seidl@informatik.rwth-aachen.de`
4   **Université Catholique de Louvain, BE,** `michel.verleysen@uclouvain.be`
5   **Fraunhofer IAIS – St. Augustin, DE and University of Bonn, DE,**
    `stefan.wrobel@iais.fraunhofer.de`

―――― **Abstract** ――――――――――――――――――――――――――――――――――――――

This report documents the program and the outcomes of Dagstuhl Seminar 12081 "Information Visualization, Visual Data Mining and Machine Learning". The aim of the seminar was to tighten the links between the information visualisation community and the machine learning community in order to explore how each field can benefit from the other and how to go beyond current hybridization successes.

1   Executive Summary

Daniel A. Keim
Fabrice Rossi
Thomas Seidl
Michel Verleysen
Stefan Wrobel

Information visualization and visual data mining leverage the human visual system to provide insight and understanding of unorganized data. Visualizing data in a way that is appropriate for the user's needs proves essential in a number of situations: getting insights about data before a further more quantitative analysis, presenting data to a user through well-chosen table, graph or other structured representations, relying on the cognitive skills of humans to show them extended information in a compact way, etc.

Machine learning enables computers to automatically discover complex patterns in data and, when examples of such patterns are available, to learn automatically from the examples how to recognize occurrences of those patterns in new data. Machine learning has proven

itself quite successful in day to day tasks such as SPAM filtering and optical character recognition.

Both research fields share a focus on data and information, and it might seem at first that the main difference between the two fields is the predominance of visual representations of the data in information visualization compared to its relatively low presence in machine learning. However, it should be noted that visual representations are used in a quite systematic way in machine learning, for instance to summarize predictive performances, i.e., whether a given system is performing well in detecting some pattern. This can be traced back to a long tradition of statistical graphics for instance. Dimensionality reduction is also a major topic in machine learning: one aims here at describing as accurately as possible some data with a small number of variables rather than with their original possibly numerous variables. Principal component analysis is the simplest and most well known example of such a method. In the extreme case where one uses only two or three variables, dimensionality reduction is a form of information visualization as the new variables can be used to directly display the original data.

The main difference between both fields is the role of the user in the data exploration and modeling. The ultimate goal of machine learning is somehow to get rid of the user: everything should be completely automated and done by a computer. While the user could still play a role by, e.g., choosing the data description or the type of algorithm to use, his/her influence should be limited to a strict minimum. In information visualization, a quite opposite point of view is put forward as visual representations are designed to be leveraged by a human to extract knowledge from the data. Patterns are discovered by the user, models are adjusted to the data under user steering, etc.

This major difference in philosophy probably explains why machine learning and information visualization communities have remained relatively disconnected. Both research fields are mature and well structured around major conferences and journals. There is also a strong tradition of Dagstuhl seminars about both topics. Yet, despite some well known success, collaboration has been scarce among researchers coming from the two fields. Some success stories are the use of state-of-the-art results from one field in the other. For instance, Kohonen's Self Organizing Map, a well known dimensionality reduction technique, has been successful partly because of its visualization capabilities which were inspired by information visualization results. In the opposite direction, information visualization techniques often use classical methods from machine learning, for instance, clustering or multidimensional scaling.

The seminar was organized in this context with the specific goal of bringing together researchers from both communities in order to tighten the loose links between them. To limit the risk of misunderstandings induced by the different backgrounds of researchers from the two communities, the seminar started with introductory talks about both domains. It was then mainly organized as a series of thematic talks with a significant portion of the time dedicated to questions and discussions. After the first two days of meeting, understanding between both communities reached a sufficient level to organize, in addition to the plenary talks, working group focusing on specific issues.

Several research topics emerged from the initial discussions and lead to the creation of the working groups. The subject that raised probably the largest number of questions and discussions is Evaluation. It is not very surprising as differences between the communities about evaluation (or quality assessment) might be considered as the concrete technical manifestation of cultural and philosophical differences between them. Indeed, in machine learning, automatic methods are mostly designed according to the following general principle: Given a quality measure for a possible solution of the problem under study, one devises an

algorithm that searches the solution space efficiently for the optimal solution with respect to this measure. For instance, in SPAM filtering a possible quality measure is the classification accuracy of the filter: it has to sort unsolicited bulk messages correctly into the SPAM class and all other emails in the HAM class. In a simple setting, the best filter could be considered as the one with the smallest number of errors. However, counting only the number of errors is usually too naive, and better quality measures have to be used, such as the area under the ROC curve: the Receiver Operating Characteristic curve shows the dependency between the true positive rate (the percentage of unsolicited bulk messages classified as SPAM) and the false positive rate (the percentage of correct emails classified as SPAM).

In information visualization, evaluation cannot rely only on mathematical quality measures as the user is always part of the story. A successful visualization is a solution, with which the user is able to perform better, in a general sense, compared to existing solutions. As in machine learning, a method is therefore evaluated according to some goal and with some quality metric, but the evaluation process and the quality metrics have to take the user into account. For instance, one display can be used to help the user assess the correlation between variables. Then, a quality metric might be the time needed to find a pair of highly correlated variables, or the time needed to decide that there is no such pair. Another metric might be the percentage of accurate decisions about the correlation of some pairs of variables. In general, a visualization system can be evaluated with respect to numerous tasks and according to various metrics. This should be done in a controlled environment and with different users, to limit the influence of interpersonal variations.

Among the discussions between members of the two communities about evaluation, questions were raised about the so-called unsupervised problems in machine learning. These problems, such as clustering or dimensionality reduction, are ill-posed in a machine learning sense: there is no unquestionable quality metric associated to e.g. clustering but rather a large number of such metrics. Some of those metrics lead to very difficult optimization problems (from a computational point of view) that are addressed via approximate heuristic solutions. In the end, machine learning has produced dozens of clustering methods and dimensionality reduction methods, and evaluations with respect to user needs remain an open problem. An important outcome of the seminar was to reposition this problem in the global picture of collaboration between information visualization and machine learning. For instance, if many quality measures are possible, one way to compare them would be to measure their link to user performances in different tasks. If several methods seem to perform equally well in a machine learning sense, then the user feedback could help to indentify the «best» method. It was also noted that many methods that are studied in machine learning and linked to information visualization, in particular dimensionality reduction and embedding techniques, would benefit from more interaction between the communities. At minimum, state-of-the-art methods from machine learning should be known by information visualization researchers and state-of-the-art visualization techniques should be deployed by machine learning researchers.

Another topic discussed thoroughly at the seminar was the visualization of specific types of objects. Relational data were discussed, for instance, as a general model for heterogeneous complex data as stored in a relational database. Graph visualization techniques provide a possible starting point, but it is clear that for large databases, summarization is needed, which brought back the discussion of the ill defined clustering problem mentioned above. Among complex objects, models obtained by a machine learning algorithms were also considered, in particular as good candidates for interactive visualizations. Decision trees give a good example of such objects: Given a proper visualization of the current tree, of some possible

simplified or more complex versions and of the effect of the tree(s) on some dataset, an expert user can adapt the tree to his/her specific goals that are not directly expressible in a quality criterion. The extreme case of visualizing the dynamic evolution of a self learning process was discussed as a prototype of complex objects representation: The system is evolving through time, it learns decision rules, and it evolves using complex (and evolving) decision tables.

Finally, it became clear that a large effort is still needed at the algorithmic and software levels. First, fast machine learning techniques are needed that can be embedded in interactive visualization systems. Second, there is the need for a standard software environment that can be used in both communities. The unavailability of such a system hurts research to some extent as some active system environments in one field do not include even basic facilities from the other. One typical example is the R statistical environment with which a large part of machine learning research is conducted and whose interactive visualization capabilities are limited, in particular in comparison to the state-of-the-art static visualization possibilities. One possible solution foreseen at the seminar was the development of some dynamic data sharing standard that can be implemented in several software environments, allowing fast communication between those environments and facilitating software reuse.

Judging by the liveliness of the discussions and the number of joint research projects proposed at the end of the seminar, this meeting between the machine learning and the information visualization communities was more than needed. The flexible format of the Dagstuhl seminars is perfectly adapted to this type of meeting and the only frustration perceivable at the end of the week was that it had indeed reached its end. It was clear that researchers from the two communities were starting to understand each other and were eager to share more thoughts and actually start working on joint projects. This calls for further seminars ...

2 Table of Contents

Working Groups

3 Overview of Talks

3.1 Graph visualization methods and data mining: results, evaluation, and future directions

Daniel Archambault (University College Dublin, IE)

> **License** (cc) (i) (=) Creative Commons BY-NC-ND 3.0 Unported license
> © Daniel Archambault

Graph visualization and data mining methods have many areas of common interest.

In the introductory talk for this session, I will cover some of my recent results on graph visualization applicable to this topic, outline methods of visualization research, and identify some possible areas of future collaboration.

3.2 Steerable Large Scale Data Analytics

Daniel Archambault (University College Dublin, IE)

> **License** (cc) (i) (=) Creative Commons BY-NC-ND 3.0 Unported license
> © Daniel Archambault

In this short talk, I cover some ideas on steerable data analytics. In this area, I think that we should strive to strengthen the coupling between data mining or clustering processes and visualization in order to enable real time analysis. I give potential ways to achieve this goal with possible applications to the area of social media analysis and community finding.

3.3 Multivariate data exploration with CheckViz and ProxiViz

Michael Aupetit (Commissariat a l'Energie Atomique – Gif-sur-Yvette, FR)

> **License** (cc) (i) (=) Creative Commons BY-NC-ND 3.0 Unported license
> © Michael Aupetit
> **Joint work of** Aupetit, Michael; Lespinats, Sylvain
> **Main reference** S. Lespinats, M. Aupetit, "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings," Computer Graphics Forum 30(1):113–125, 2011.
> **URL** http://dx.doi.org/10.1111/j.1467-8659.2010.01835.x

Embedding techniques are used for multivariate data analysis. They provide a planar set of points whose relative distances estimates the original similarities.

We argue that this set of points alone is not enough to make sense out of it. We present CheckViz [2] and ProxiViz [1] as two ways to make the set of points interpretable by the user. CheckViz overload distortions straight into the map, it can be used as a sanity check and also provides inference rule which help to recover the original data topology. ProxiViz overload the true original similarity measure between a selected point and each of the other points which makes possible to reconstruct the original data structure. The embeddings appear not to be an end, but just a mean to display a complementary information which make them usable and useful for multivariate data exploration.

References
1 Michaël Aupetit, *Visualizing distortions and recovering topology in continuous projection techniques.* Neurocomputing 70(7-9):1304-1330, March 2007.
2 Sylvain Lespinats, and Michaël Aupetit, *CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings.* Computer Graphics Forum 30(1):113-125, 2011.

3.4   Matrix relevance learning and visualization of labeled data sets

Michael Biehl (University of Groningen, NL)

A brief introduction is given to Learning Vector Quantization (LVQ) as an intuitive, flexible, and very powerful prototype-based classifier.

The focus is on the recent extension of LVQ by Matrix Relevance Learning. In this scheme, one or several matrices of adaptive relevances are employed to parameterize a distance measure.

Matrix Relevance Learning makes use of a low-dimensional linear or locally linear representation of the data set, internally. This fact can be exploited for the discriminative visualization of labelled data sets.

In terms of a few application examples from the life sciences it is argued that these visualizations facilitate valuable insight into the nature of the problems.

Possible routes to extend the schemes to explicitly non-linear visualizations are briefly discussed. This leads to the question what the goal of visualizing labeled data should be.

The following references may serve as a starting point to get acquainted with Matrix Relevance Learning in the context of visualization.

References

1   P. Schneider, M. Biehl, and B. Hammer. *Adaptive relevance matrices in Learning Vector Quantization.*
Neural Computation 21(12): 3532-3561, 2009
2   K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl. *Limited rank matrix learning, discriminative dimension reduction, and visualization.* Neural Networks 26: 159-173, 2012

3.5   Supervised dimension reduction – A brief history

Kerstin Bunte (Universität Bielefeld, DE)

Due to improved sensor technology, dedicated data formats and rapidly increasing digitalization capabilities the amount of electronic data increases dramatically since decades. As a consequence the manual inspection data sets often becomes infeasible. In recent years, many powerful non-linear dimension reduction techniques have been developed which provide a visualization of complex data sets. Using prior knowledge, e.g. in form of supervision might provide more informative mappings dependent on the actual data set.

3.6   Overview of Visual Inference

Dianne Cook (Iowa State University, US)

Main reference A. Buja, D. Cook, H. Hofmann, M. Lawrence, E.-K. Lee, D.F. Swayne, H. Wickham, "Statistical
Inference for Exploratory Data Analysis and Model Diagnostics," Royal Society Philosophical
Transactions A, vol. 367, no. 1906, pp. 4361–4383, 2009.
URL http://dx.doi.org/10.1098/rsta.2009.0120

Implicitly detection of patterns in a plot of data is a rejection of some null hypothesis. What patterns might we see in the plot if the data was sampled in a manner consistent with the null hypothesis? This research area provides methods for assessing whether what we see in plots is "real", and obtaining levels of significance for findings based on visualization. Two protocols are used, a lineup and a rorschach. In the lineup, the plot of real data is embedded in a field of plots of data generated in a manner consistent with the relevant null hypothesis. In a rorschach, all plots are null plots, and the approach is a way to examine how much variability can occur purely by chance.

3.7   Eye-tracking Experiments for Visual Inference

Dianne Cook (Iowa State University, US)

Visual inference provides methods for assessing whether what we see in plots is real. The primary method is a lineup, where a plot of the actual data is embedded in a field of plots of data generated in a manner consistent with the null hypothesis. For example, to assess the relationship between two variables with a scatterplot, the null plots may show the same data, with one of the variables having its values permuted, thus breaking any real association between the avriables. If the observer picks the actual data plot from the lineup it lends significance to the conclusion of a real relationship between the two variables. Following a series of Amazon Turk experiments where the lineup protocol was evaluated under controlled simulated data experimental conditions, we selected a handful of lineups for detailed assessment. Here subjects were recorded with an eye trackers to examine (1) how long they looked at their selection, (2) which plots caught the subjects attention and (3) how subjects scanned the lineups to make their selections.

3.8   Future Analysis Environments

Jean-Daniel Fekete (Université Paris Sud – Orsay, FR)

What is the future of Analysis environments allowing machine learning and visualization to interoperate seamlessly? Should we design a new system that will solve all the problems, reuse already existing systems or in-between? These slides summarize a possible way to address the issue that might address the problem is a simple enough way.

3.9 Psychology of Visual Analytics

Brian D. Fisher (Simon Fraser University – Surrey, CA)

This talk explores the larger implications of visual analytics – the science of analytical reasoning facilitated by interactive visual interfaces for cognitive science and informatics. The visual analytics approach emphasizes the design of technologies to support the ability of trained human analysts to understand situations, make decisions, generate plans, and put them into action. The resulting visual information systems succeed when they enable analysts to more effectively work with complex "big data" from sensors, archives, computational and mathematical models, alone and in collaboration with other analysts.

My laboratory begins by building field study methods that characterize human and computational cognitive capabilities as they are used for decision-making in specific situations in flight safety, public health, and emergency management analysis. These field studies generate research questions and experimental protocols that are used to investigate human-computer cognitive systems in the laboratory. My talk will briefly discuss our "pair analytics" methods derived from H. Clark's Joint Activity Theory and two laboratory studies of perceptual cognition in display environments similar to those proposed for air traffic control and collaborative aircraft CAD.

3.10 BCI-based Evaluation in Information Visualization

Hans Hagen (TU Kaiserslautern, DE)

Evaluations have been the key factor for validating different visualization and interaction approaches. But while experts agree on their importance, the evaluation techniques currently used in Information Visualization focus mostly on objective measurements like performance and efficiency, and only rarely investigate subjective factors (states of mind and emotions that the users experience).

As the ideal evaluation should be non-intrusive and executed in real-time, many researchers turn to novel brain-computer interfaces (BCI) for directly investigating the users' affective and mental states. While current portable BCI systems are employed overwhelmingly in control tasks (e.g. moving a robotic Arm), many of them have proven useful in supporting subjectivity measurements and, thus, evaluations in real-time.

But what would an ideal BCI system detect and how would it process it in order to support the evaluation of Information Visualization approaches? Could a framework specifically designed for InfoVis evaluation with BCI systems enable researchers to obtain the answers they seek? These are a couple of specific topics that need to be addressed when looking at the potential of BCI systems as an alternative evaluation method for Information Visualization techniques and systems.

3.11 Including prior knowledge into data visualization

Barbara Hammer (Universität Bielefeld, DE)

In this presentation, the question of how data visualization and dimensionality reduction are linked to prior knowledge will be investigated. First, it will be motivated, that visualization and prior knowledge are closely connected.

Afterwards, technical possibilities how to integrate different kind of prior knowledge into dimensionality reduction will be discussed. Four major principles will be identified and demonstrated by examples: (i) change of the prior in a Bayesian model. For a cost-function based techniques, possibilities are given by a (ii) change of the data representation or metric, (iii) change of the cost function used for training, (iv) change or bias of the mapping of data to low dimensions.

3.12 Automated Methods in Information Visualization

Helwig Hauser (University of Bergen, NO)

Visualization and Machine Learning have related goals in terms of helping analysts to understand characteristic aspects of data. While visualization aims at involving the user through interactive depictions of data, machine learning is generally represented by automatic methods that yield optimal results with respect to certain initially specified tasks. Not at the least within the research direction of visual analytics it seems promising to think about opportunities to integrate both methodologies in order to exploit the strengths of both sides. Up to now, examples of integration very often encompass the visualization of results from automatic methods as well as attempts to make originally automated methods partially interactive. A vision for the future would be to integrate interactive and automatic methods in order to solve problems. A possible realization could be an iterative process where the one or other approach is chosen on demand at each step.

3.13 Distance concentration and detection of meaningless distances

Ata Kaban (University of Birmingham, GB)

Main reference A Kaban, "Non-parametric Detection of Meaningless Distances in High-Dimensional Data,"
Statistics and Computing. 22(1): 375–385.
URL http://dx.doi.org/10.1007/s11222-011-9229-0

Distance concentration is a counter-intuitive aspect of the curse of dimensionality, the phenomenon that in certain conditions the contrast between the nearest and the farthest neighbouring points vanishes as the data dimension increases. This makes distances meaning-less, exponentially slows down data retrieval, and risks to compromise our ability to extract meaningful information from high dimensional data sets. First, we show that the known

sufficient conditions are also the necessary conditions of distance concentration in the limit of infinite dimensions. We then quantify the phenomenon more precisely, for possibly high but finite dimensional settings in a distribution-free manner, by bounding the tails of the probability that distances become meaningless. We show how this can be turned into a statistical test to assess the concentration of a given distance function in some unknown data distribution solely on the basis of an available data sample from it. This can be used to test and detect problematic cases more rigorously than it has been possible previously, and we demonstrate the working of this approach on both synthetic data and ten real-world data sets from different domains.

References
1   A Kaban. *Non-parametric Detection of Meaningless Distances in High-Dimensional Data.* Statistics and Computing. 22(1): 375-385.

3.14   Visual Analysis of Multi-faceted Scientific Data: a Survey

Johannes Kehrer (VRVis – Wien, AT)

Interactive visual analysis plays an important role in studying different kinds of scientific data (e.g., spatial, temporal and/or multi-variate data). The talk is based on a thorough literature review, which investigates to which degree methods for 1) visual representation, 2) user interaction and 3) computational analysis are combined in such an analysis. A task-based categorization of approaches is proposed and different options for the visual analysis are discussed. This leads to conclusions with respect to promising research directions, for instance, to pursue new solutions that combine supervised machine learning with interactive feature specification via brushing.

3.15   Towards Visual Analytics

Daniel A. Keim (Universität Konstanz, DE)

Many of the grand challenges require not only automatic methods, but also exploration to find appropriate solutions. Visual Analytics as the tight integration of visual and automatic data analysis methods for information exploration and scalable decision support aims at integrating machine capabilities (e.g., data storage, numerical computation or search) with human capabilities (such as perception, creativity and general knowledge). Besides giving an introduction to Visual Analytics and information visualization, this talk describes common evaluation approaches and outlines the relation between visualization and machine learning.

3.16 Visualization of Network Centralities

Andreas Kerren (Linnaeus University – Växjö, SE)

The use of network centralities in the field of network analysis plays an important role when the relative importance of nodes within the network topology should be rated. A single network can easily be represented by the use of standard graph drawing algorithms, but not only the exploration of one centrality might be important: the comparison of two or more of them is often crucial for a better understanding. When visualizing the comparison of several network centralities, we are facing new problems of how to show them in a meaningful way. For instance, we want to be able to track all the changes of centralities in the networks as well as to display the single networks as best as possible. In the life sciences, centrality measures help scientists to understand the underlying biological processes and have been successfully applied to different biological networks. The aim of this talk was to briefly present a system for the interactive visualization of biochemical networks and its centralities. Researchers can focus on the exploration of the centrality values including the network structure without dealing with visual clutter or occlusions of nodes. Simultaneously, filtering based on statistical data concerning the network elements and centrality values supports this.

3.17 Embedding from high- to low-dimensional spaces; how can we cope with the phenomenon of norm concentration?

John A. Lee (Université Catholique de Louvain, BE)

Dimensionality reduction aims at representing high-dimensional data in low-dimensional spaces, mainly for visualization and exploratory purposes. As an alternative to projections on linear subspaces, nonlinear dimensionality reduction, also known as manifold learning, can provide data representations that preserve structural properties such as pairwise distances or local neighborhoods. Very recently, similarity preservation emerged as a new paradigm for dimensionality reduction, with methods such as stochastic neighbor embedding and its variants. Experimentally, these methods significantly outperform the more classical methods based on distance or transformed distance preservation.

This talk explains both theoretically and experimentally the reasons for these performances. In particular, it details (i) why the phenonomenon of distance concentration is an impediment towards efficient dimensionality reduction and (ii) how SNE and its variants circumvent this difficulty by using similarities that are invariant to shifts with respect to squared distances. The paper also proposes a generalized definition of shift-invariant similarities that extend the applicability of SNE to noisy data.

3.18 Visual Analytics of Sparse Data

Marcus A. Magnor (TU Braunschweig, DE)

High-dimensional data will always constitute only sparse representations of inter-dimensional information. As a result of large voids in n-D space, even without taking noise and erroneous data into account, putative inter-dimensional relations may only be halluscinated, by humans as well as by algorithms. In contrast, suitable interpolation on the data level, guided by high-level knowledge of the data and dimensional meaning, may be able to plausibly fill the voids and to fortify subsequent interactive and automatic analysis results.

3.19 Exploration through Enrichment

Florian Mansmann (Universität Konstanz, DE)

In many visualization scenarios, visualizing and exploring data raises hypotheses that cannot be answered with the current data. Therefore, very often an enrichment phase is needed to enhance the exploration process. In this talk, I showed two prototypes, namely ClockView in which network time series can be filtered through user-defined patterns and the Animal Ecology Explorer in which bird movement can be interactively refined through machine learning methods such as clustering and classification.

References
1 C. Kintzel, J. Fuchs and F. Mansmann. *Monitoring Large IP Spaces with ClockView*. Proc. of Int. Symp. on Visualization for Cyber Security (VizSec), 2011.
2 D. Spretke, H. Janetzko, F. Mansmann, P. Bak, B. Kranstauber, S. Davidson and M. Mueller. *Exploration through Enrichment: A Visual Analytics Approach for Animal Movement*. Proc. of the 19th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pages 421–424, 2011.

3.20 Quality metrics for InfoVis

Florian Mansmann (Universität Konstanz, DE)

Quality metrics are a recent trend in the information visualization community. The basic idea is that the quality of a visualization with respect to the loaded data can be calculated

and based on this assessment the good or optimal parameter configurations for visualizations can be found.

References

1   E. Bertini, A. Tatu and D. A. Keim. *Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization.* IEEE Transactions on Visualization and Computer Graphics (TVCG), vol. 17, no. 12, pp. 2203-2212, 2011.

3.21   The Generative Topographic Mapping and Interactive Visualization

Ian Nabney (Aston University – Birmingham, GB)

Joint work of Nabney, Ian; Tino, Peter; Maniyar, Dharmesh; Schroeder, Martin

The Generative Topographic Mapping (GTM) is a probabilistic generative data model. Using Bayes' theorem, the mapping can be inverted and used for visualization. Because the model is a constrained mixture of Gaussians, an (extended) EM algorithm can be used to train models. The smooth mapping defined by GTM defines a two-dimensional manifold embedded in data space: geometric measures (e.g. magnification and curvature) can be visualized to understand the embedding and diagnose modelling flaws.

More recent advances include modelling missing data, discrete variables, and hierarchies: all of these can be handled in a consistent probabilistic framework. With a bit more analysis, it is possible to incorporate prior knowledge of variable correlation structure (with block-structured covariance models) and unsupervised feature selection (with minimum message length criteria). The talk concluded with a short demonstration of a visualization system that integrates machine learning and information visualisation (Data Visualization and Modelling System: DVMS) written in Matlab which is available from the Aston website. http://www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/downloads/

3.22   Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization

Jaakko Peltonen (Aalto University, FI)

Joint work of Venna, Jarkko; Peltonen, Jaakko; Nybo, Kristian; Aidos, Helena; Kaski, Samuel
Main reference J. Venna, J. Peltonen, K. Nybo, H. Aidos, S. Kaski, "Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization," Journal of Machine Learning Research, 11:451–490, 2010.
URL http://jmlr.csail.mit.edu/papers/v11/venna10a.html

Nonlinear dimensionality reduction methods are often used to visualize high-dimensional data, although the existing methods have been designed for other related tasks such as manifold learning. It has been difficult to assess the quality of visualizations since the task has not been well-defined. We give a rigorous definition for a specific visualization task, resulting in quantifiable goodness measures and new visualization methods. The task is information retrieval given the visualization: to find similar data based on the similarities

shown on the display. The fundamental tradeoff between precision and recall of information retrieval can then be quantified in visualizations as well. The user needs to give the relative cost of missing similar points vs. retrieving dissimilar points, after which the total cost can be measured. We then introduce a new method NeRV (neighbor retrieval visualizer) which produces an optimal visualization by minimizing the cost. We further derive a variant for supervised visualization; class information is taken rigorously into account when computing the similarity relationships. We show empirically that the unsupervised version outperforms existing unsupervised dimensionality reduction methods in the visualization task, and the supervised version outperforms existing supervised methods.

3.23 Visualization of Learning Processes – A Problem Statement

Gabriele Peters (FernUniversität in Hagen, DE)

The visual representation of the results of machine learning algorithms can be regarded as an open research topic. But rather to restrict the visualization to only the results of machine learning approaches, the discussion should be expanded to the visualization of the learning processes themselves. Whereas a visualization of results promises a better interpretation of what has been learned, the visualization of learning processes may provide a better understanding of underlying principles of learning (also in biological systems).

Maybe it can also account for general insights in the possibilities of autonomous learning at all. In my talk I present briefly the architecture of a self-learning system with two levels of hierarchy together with some results obtained in a computer vision task. From this I derive questions of general interest such as possible options to visualize the flow of information in a dynamic learning system or the visualization of symbolic data.

3.24 Learning of short time series

Frank-Michael Schleif (Universität Bielefeld, DE)

Joint work of Schleif, Frank-Michael; Gisbrecht, Andrej; Hammer, Barbara
Main reference F.-M. Schleif, A. Gisbrecht, B. Hammer, "Relevance learning for short high-dimensional time series in the life sciences," IJCNN'12, 2012

The talk presented some concepts used to learn short and high dimensional time series.

Especially I detailed a method for topographic mapping and recent extensions thereof in the line of supervised relevance learning.

Challenges in the modeling and visualization were discussed.

3.25   Comparative Visual Cluster Analysis

Tobias Schreck (Universität Konstanz, DE)

Data that is to be analyzed with cluster analysis tools may be represented by sets of feature vectors stemming from alternative feature extraction processes.

Interesting cluster structures may reside in several of the alternative feature representations, and they may confirm, complement, or contradict each other. In this talk we consider the problem of comparative visual cluster analysis in multiple features spaces (or subspaces). We first briefly review a previously proposed method for visual comparison of multiple feature spaces represented by Self-Organizing Map models. We then discuss ongoing work that aims to make use of automatic subspace selection methods. First results based using the SURFING subspace selection method are reported. The basic idea is to define a custom similarity function for the subspaces. The function currently considers the intersection of the selected dimensions as well as the agreement in clustering structures exhibited in the subspaces. Different visual representations based on MDS layouts, TreeMap layouts etc. as well as interaction techniques are investigated. Eventually, our approach should help analysts in identifying the most interesting subspaces from a potentially much larger set of subspaces reported by the subspace selection method.

3.26   Visualization of (machine) learning processes and dynamic scenarios

Marc Strickert (Universität Marburg, DE)

The title can be related to an overwhelming plenitude of aspects such as functional brain imaging, motion sensor and eye tracker analysis, neural spike train observations, phase space portraits, or time series and data stream mining. To focus the wide topic on one essential commonality, this involves the transformation of spatio-temporal multi-dimensional input data into representations that are compatible with analysts' world view. This requires a compatibility between the data model and the world model mainly constituted by three spatial coordinates, color, intensity, and experience of spatio-temporal contiguity.

In machine learning methods sequential signals are often recursively mixed with a representation of the most recent internal state for modeling first-order context. The current model state is thus a representation of possibly unifying encodings of external dynamics. Depending on different readout functions applied to the model parameters different aspects of the input stream are focused on.

After all, structure detection, including ordering and convergence trends, is considered as crucial component for extracting aspects which are potentially relevant for visualization.

3.27 Prior knowledge for Visualization or prior visualization results for knowledge generation – a chicken-egg problem?

Holger Theisel (Universität Magdeburg, DE)

It seems that both communities – Machine Learning and Visualization – use different words for the same concept, and even use the same words for different concepts. This holds both for "prior knowledge" and "visualization". Being aware of this, the following questions are discussed: Are there new ML algorithms when the goal is preparation for a interactive visual analysis? Are there new Vis approaches when the goal is not complete insight but preparation of an automatic analysis?

3.28 Interactive decision trees and myriahedral maps

Jarke J. Van Wijk (TU Eindhoven, NL)

Joint work of Van Wijk, Jarke J.; van den Elzen, Stef;
Main reference S. van den Elzen, J.J. van Wijk, "BaobabView: Interactive construction and analysis of decision trees," IEEE VAST 2011: 151–160.

In my talk, I present BaobabView, a system developed by Stef van den Elzen. It enables users to construct, inspect, and evaluate decision trees via a wide range of features. Next, myriahedral projections are presented via a video. These are mappings of the sphere to the plane using an approximation of the sphere with a large number of facets, which are cut and folded out. Finally, a short demo of SeifertView is given. Seifert surfaces are orientable surfaces that are bounded by knots or links. They illustrate that 2-manifolds embedded in 3D can take complex shapes.

3.29 Clustered graph, visualization, hierarchical visualization

Nathalie Villa-Vialaneix (Université Paris I, FR)

Joint work of Villa-Vialaneix, Nathalie; Rossi, Fabrice
Main reference F. Rossi, N. Villa-Vialaneix, "Représentation d'un grand réseau à partir d'une classification hiérarchique de ses sommets," Journal de la Société Française de Statistique, 152, pp. 34–65, 2011.
URL http://hal.archives-ouvertes.fr/hal-00651577/

Clustering is a useful approach to provide a simplified and meaningful representation of large graphs. By extracting dense communiites of nodes, the "big picture" of the network organizatin is enlighten. Moreover, hierarchical clustering may help the user to focus on some parts of the graph which is of interest for him and which can be displayed with finer and finer details.

This talk will try to present some open issues with graph visualization based on a hierarchical nodes clustering. These issues include displaying the clusters in a coherent way between the different layers of the hierarchy or integrating information about the clustering evaluation in the visualization. It is related to the article [1].

References
1   Rossi, F. and Villa-Vialaneix, N. (2011) Représentation d'un grand réseau à partir d'une classification hiérarchique de ses sommets. Journal de la Société Française de Statistique, 152, 34-65.

3.30   Perceptual Experiments for Visualization

Daniel Weiskopf (Universität Stuttgart, DE)

I briefly describe and discuss a few examples of user experiments that investigate the visual perception of visualization results, including studies that use methods from vision research, eye-tracking experiments in the context of the visualization of node-link diagrams of graphs and trees, as well as learning attention models for video visualization by utilizing eye- tracking data.

References
1   Michael Burch, Corinna Vehlow, Natalia Konevtsova, Daniel Weiskopf: Evaluating Partially Drawn Links for Directed Graph Edges. *Graph Drawing 2011*, 226-237 (2011)
2   Michael Burch, Corinna Vehlow, Fabian Beck, Stephan Diehl, Daniel Weiskopf: Parallel Edge Splatting for Scalable Dynamic Graph Visualization. *IEEE Trans. Vis. Comput. Graph.* 17(12): 2344-2353 (2011)
3   Benjamin Höferlin, Hermann Pflüger, Markus Höferlin, Gunther Heidemann, Daniel Weiskopf: Learning a Visual Attention Model for Adaptive Fast-Forward in Video Surveillance. *Proceedings of International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, 25-32 (2012).

3.31   Introduction to embedding

Laurens van der Maaten (TU Delft, NL)

In this talk, I presented an overview of some major embedding techniques and explained their strengths and weaknesses. In particular, I explained principal components analysis, locally linear embedding, and t-distributed stochastic neighbor embedding. In addition, I showed examples of embedding techniques that go beyond traditional dimensionality reduction and multidimensional scaling.

In particular, I covered embedding techniques thta learn representations from non-metric similarities such as word associations, co-occurrences, and partial order rankings.

4 Working Groups

4.1 Results of Working Group: Visualization of Dynamic Learning Processes

Michael Biehl, Kerstin Bunte,Gabriele Peters, Marc Strickert, and Thomas Villmann

We took the learning system [1] proposed by G. Peters in her talk "Visualization of Learning Processes - A Problem Statement" as example for a dynamic learning process and figured out which components can be visualized and by which means. The system has two learning levels: one with if-then rules (boolean expressions) and one with qualities of state- action pairs. Relevant questions to ask are: What are parameters of the system? Which parameters define states to be visualized? How to visualize the list of rules and the state- action table? How to visualize dynamic processes between the hierarchy levels? We proposed several means and solutions to answer these questions and intend to submit these considerations as position paper to a suitable conference or workshop.

References
1 T. Leopold and G. Kern-Isberner and G. Peters. *Combining Reinforcement Learning and Belief Revision – A Learning System for Active Vision.* BMVC, 2008

4.2 Results of Working Group: Model Visualization – Towards a Tight Integration of Machine Learning and Visualization

Florian Mansmann, Tobias Schreck, Etienne Come, and Jarke J. Van Wijk

Choosing and configuring an appropriate Machine Learning model to solve a given analysis task is crucial for arriving at useful results. Models in Machine Learning are potentially complex and sometimes hard to understand for non-experts, and often regarded and applied as black boxes. In this working group we discussed about approaches to 'opening' the black boxes by visualizing not only the data space, but also the space of model parameters. Our goal is to eventually arrive at better selection and configuration of Machine Learning models using interactive visualization. We started our discussion with the question 'What is a model?' and developed a draft reference model for Model Space Visualization. To this end, we built on existing process models, including the Information Visualization model of Card, MacKinley and Shneiderman and the Visual Analytics model proposed by Thomas and Keim. Our model adds one level of detail to the formalism and distinguishes between expert and user roles. In particular, this new process model makes the integration of Machine Learning and Visualization explicit. Consideration of model instances and parameter sets as part of the workflow in our model aims at a tighter integration of machine learning into the interactive analysis process. Also, the model is aiming as a reference structure to survey and classify existing works in Visual Analytics and Visual Data Mining. These latter points are seen as interesting future work.

4.3 Results of Working Group: Embedding techniques at the crossing of Machine Learning and Information Visualization

Michael Aupetit and John Lee

4.3.1 Attendees

- Information Visualisation: D. Keim; L. Zhang
- Machine Learning: M. Verleysen; J.A.Lee; M. Aupetit; S. Kaski; J. Peltonen; L. van der Maaten; F.-M. Schleif

4.3.2 Emerging topics of interest

- from the Information Visualisation perspective
  Getting trust from the analyst is fondamental to make embedding common visual analytics tools. For this, the projection must not change drastically if no strong changes occur in the data distribution, so questions are:
  - How to make embeddings robust to noise and outliers?
  - How to make embeddings stable adding new data points and against local optima and different initialisation ?

  Interactivity is also a very important point in visual analytics,
  - How to deals with massive datasets in terms of speed and quantity of data to visualize?
  - How to link embeddings of different local subspaces to get better understanding of the data?

  Understandability is another main issue with non linear embeddings (axes have no sense):
  - How to connect embeddings to the meaning of the original data features?
- from the Machine Learning perspective
  Assessing Visual Analytic tools needs well defined tasks:
  - Can we define a taxonomy of tasks and data types that could benefit from embeddings?

  Raw data have to be preprocessed before embedding:
  - Which kind of preprocessing has to be done before embedding?
  - Which kind of similarity measures make embeddings more efficient for which kind of task?
  - How to deal with discrete, non-Cartesian, missing data?

4.3.3 Intended actions

- Sharing of various data sets (InfoVis) and embedding methods (ML)
- Build a joint InfoVis/ML taxonomy
- Organizing a workshop and a tutorial on embeddings at the next IEEE VisWeek 2012 conference from which we can edit a special journal issue on this topic

We thank all the contributors to this group, and all the colleagues from the Dagstuhl seminar for fruitful discussions.

4.4   Results of Working Group: Evaluation

Ian Nabney, Dianne Cook, Brian D. Fisher, Andrej Gisbrecht, Hans Hagen,
and Heike Hoffman

Our outcomes were captured by Ian Nabney in a mindmap which can be found at the URL
below as a .mm file (http://db.tt/G0Ajn0V6). These files can be opened in a number of
applications including Freemind http://freemind.sourceforge.net/

4.5   Results of Working Group: Fast Machine Learning

Jörn Kohlhammer, Di Cook, Helwig Hauser, Johannes Kehrer, Marcus Magnor,
Frank-Michael Schleif, Holger Theisel

This session discussed the topic of "Fast ML for interactive visualization" and what the
different perspectives are in ML and Infovis/VA/visualization.

It turned out that the ML community in its various sub-communities is not focused per
se on performance issues of their algorithms, at least not to the extent of trying to achieve
real-time capabilities. The InfoVis and VA community on the other hand is actively looking
for high-performance, automated methods that can be coupled with visualization techniques
to include more and more data in an interactive analysis. Response times are very important
for interactive techniques and such response times do not play a major role in many ML
approaches.

There were several thoughts about the user influence on ML methods, which might
be beneficial to the ML community. One can distinguish between an internal coupling of
methods, where the user interactively influences the automated methods during run-time, or
an external coupling, which focuses on the flexible ensemble of ML methods and visualization
methods along a structured analysis workflow.

The outcome of the discussion was that it would be highly interesting for the visualization
community to learn about the current extent of research in this direction, i.e. a more
performance-driven view on current research in ML:

- What type of ML methods do exist?
- Which sub-communities (or which research groups specifically) work on high-performance
  ML approaches?
- What are these approaches in detail? What are their characteristics, scalability constraints,
  data types, etc.?
- How could we jointly work on coupling such approaches with InfoVis and VA? Are their
  existing joint efforts, best practices, examples?
- What are the plans and future work in these Vis-relevant areas?

Next steps:

Our idea was to plan a tutorial for VisWeek 2012 with the tentative title "A performance
perspective on machine learning for visualization", to be submitted by 30 April 2012. The

tutorial could be a half day or full day tutorial, depending on the outcome of the next planning steps. There could be 4-5 speakers or even more, again depending on the structure.

The tutorial should give an overview of ML methods and go into detail on the high-performance methods (along the lines of the above questions), building a possible repertoire of ML methods for visualization.

The joint understanding is that the talks in the tutorial should be held by ML experts, but with strong involvement of visualization experts in the planning phase to make sure that the talks are targeted at and are adequate/educational for the visualization community.

4.6 Results of Working Group: Future analysis environments

Jean-Daniel Fekete

The current situation of data analysis environments can be summarized simply by saying that there are numerous environments each of which has very satisfied users that do not want to switch to another solution. The only reasonable solution to avoid duplications of efforts is therefore to have some form of interoperability between environments. This can be provided at:

- a library level, with difficulties induced by differences between programming languages;
- an export/import level using e.g. xml formats with difficulties related to encoding and similar issues;
- a component level via rpc or web services mechanisms.

While interoperability might save the day, it has its share of problems:

- speed and latency;
- data duplication;
- limitation of some of the environments.

While one environment could rule them all, this seems unlikely, and improving interoperability seems a simpler goal. This needs not only the ability to share data, but also the support of notifications (of changes) and of metadata. One possible plan would be:

- specify and implement a sharing mechanism for R, Matlab, Excel...
  - how to connect/disconnect to a shared datatable
  - how to load content lazily
  - how to emit and receive notifications
  - how to manage content consistency
  - etc.
- test it

4.7 Results of Working Group: Structured/relational data

Nathalie Villa-Vialaneix

Structured and relational data have been discussed and several issues have been extracted:

- clustering issues: evaluating the quality/relevance of a clustering/cluster

- taking into account heteregeneous data: heteregeneous data could lead to different clusterings: put the user in the loop to help find a consensual user-driven clustering
- metric came from mathematics; use the users' suggestions to try to find a consensus among the human experts and use ML to extract a relevant metric that fits the users' suggestion
- how to find a relevant labelling for a cluster: give the user hints automatic help for labelling

Participants

Daniel Archambault
University College Dublin, IE

Michael Aupetit
Commissariat a l'Energie
Atomique – Gif-sur-Yvette, FR

Michael Biehl
University of Groningen, NL

Kerstin Bunte
Universität Bielefeld, DE

Etienne Come
IFSTTAR – Noisy le Grand, FR

Dianne Cook
Iowa State University, US

Jean-Daniel Fekete
Université Paris Sud – Orsay, FR

Brian D. Fisher
Simon Fraser Univ. – Surrey, CA

Ksenia Genova
TU Dresden, DE

Andrej Gisbrecht
Universität Bielefeld, DE

Hans Hagen
TU Kaiserslautern, DE

Barbara Hammer
Universität Bielefeld, DE

Helwig Hauser
University of Bergen, NO

Heike Hofmann
Iowa State University, US

Ata Kaban
University of Birmingham, GB

Samuel Kaski
Aalto University and University
of Helsinki

Johannes Kehrer
VRVis – Wien, AT

Daniel A. Keim
Universität Konstanz, DE

Andreas Kerren
Linnaeus University – Växjö, SE

Jörn Kohlhammer
Fraunhofer Institut –
Darmstadt, DE

Bongshin Lee
Microsoft Res. – Redmond, US

John A. Lee
Université Catholique de
Louvain, BE

Marcus A. Magnor
TU Braunschweig, DE

Florian Mansmann
Universität Konstanz, DE

Ian Nabney
Aston Univ. – Birmingham, GB

Jaakko Peltonen
Aalto University, FI

Gabriele Peters
FernUniversität in Hagen, DE

Nathalie Riche-Henry
Microsoft Res. – Redmond, US

Fabrice Rossi
Université Paris I, FR

Frank-Michael Schleif
Universität Bielefeld, DE

Tobias Schreck
Universität Konstanz, DE

Marc Strickert
Universität Marburg, DE

Holger Theisel
Universität Magdeburg, DE

Peter Tino
University of Birmingham, GB

Laurens van der Maaten
TU Delft, NL

Jarke J. Van Wijk
TU Eindhoven, NL

Michel Verleysen
Université Catholique de
Louvain, BE

Nathalie Villa-Vialaneix
Université Paris I, FR

Thomas Villmann
Hochschule Mittweida, DE

Daniel Weiskopf
Universität Stuttgart, DE

Hadley Wickham
Rice University – Houston, US

Leishi Zhang
Universität Konstanz, DE

24th International Conference of the European Federation for Medical Informatics
Quality of Life through Quality of Information – J. Mantas et al. (Eds.)
MIE2012 / CD / Short Communications (Poster)

Visual Analytics to Support Medical Decision-Making Process

Samar Al-HAJJ[a,1], Dr. Ian PIKE[b] and Dr. Brian FISHER[c]

[a,c] *Simon Fraser University, SCIENCE Lab, Canada*
[b] *University of British Columbia, BCIRP Unit, Canada*

Abstract. Massive and complex data impose a challenge on the medical community. This study explores the use of Visual Analytics in a collaborative Paired and Group Analytics sessions to enable medical professionals to efficiently explore complex data to synthesize valuable information and make informed decisions about dynamic medical events and situations. Based on the conducted Analytics sessions, Visual Analytics facilitated stakeholders' collaborative data exploratory analysis process and enabled stakeholders to generate insights, build knowledge and potentially decide on critical medical situations.

Keywords. Visual Analytics, Paired Analytics, Group Analytics, Decision-making.

Introduction

Advanced technologies enable the production and collection of massive amount of complex and multidimensional medical data. The medical community is faced with the challenge of synthesizing valuable information from these complex and dynamic data to make informed decisions. Visual Analytics exploits humans' visual capabilities to amplify their perceptual skills to facilitate the analysis of massive data. Visual Analytics (VA) defined as the "science of analytical reasoning facilitated by interactive visual interface" [6], enables medical professionals to visually explore large databases in order to expedite data analysis, accelerate knowledge translation and dissemination as well as support decision-making. Visual Analytics offers medical professionals interactive and intuitive visualization tools and techniques to amplify their cognitive skills and enhance their initial understanding of massive data [4].

In this study, we exploited Visual Analytics in Paired and Group Analytics sessions to enable stakeholders to collaboratively and productively build knowledge and make informed decisions. Previous studies addressed social aspects of Visual Analytics. Heer et al. argued that social interactions contribute to the perceptual and cognitive process and improve the data analytical process [3]. Other research presented a framework for Visual Analytics multiple-analysts collaboration to conduct exploratory data analysis through interactive visual displays. To supplement previous research on the impact of collaboration for the advancement of Visual Analytics and the way analysts collaboratively explore data visualizations, synthesize information and build knowledge, this paper examines how Paired and Group Analytics fosters collaboration to improve exploratory data analysis and support decision-making.

---

[1] Corresponding Author. Samar Al-Hajj, email: samara@sfu.ca.

1. Data and Methods

We used Mortality and Morbidity injury indicators data in British Columbia, Canada to create the visualizations using Tableau Software. The injury data are segmented into categories including patients' injury types, gender, socioeconomic status as well as geographic locations. Multiple visualizations were created and integrated together to build an Interactive Visual Analytics Dashboard. The created VA Dashboard is a comprehensive visual representations of the most relevant information required for stakeholders to reach specific goals [2]. It encompasses multiple Visual Analytics displays (Bar Chart, Stacked Bar Chart, Geospatial, Temporal visualizations) to efficiently depict the injury data [Fig.1]. The Dashboard provides stakeholders with advanced visualization techniques applying Shneiderman's Visual Information Seeking Mantra: overviews first, zoom and filter, then details on demand [5].
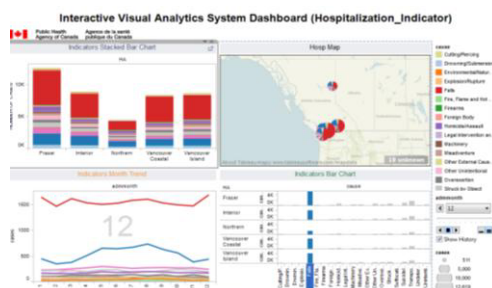
Figure 1. Interactive Visual Analytics Dashboard.

We adopted Paired Analytics and Group Analytics Methodologies. During the Analytics sessions, a Subject Matter Expert (SME) collaboratively worked with visualization Tool Expert (TE), either one-on-one or in a group setting, to solve an analytical problem using the visualization tool [1]. SME are stakeholders from diverse backgrounds including injury practitioners, researchers, epidemiologists and medical policy makers. After the Analytics sessions, stakeholders were asked to complete a questionnaire about the VA Dashboard's perceived usefulness and its potential to help stakeholders generate insights and build knowledge to support decision-making.

2. Results

We compiled the feedback data and analyzed them using JMP Software. Compiled feedback proves that stakeholders perceived the visualizations to be useful (average rating = 6.3/7, Std Dev = 0.65) and time saving (average rating = 5.6/7, Std Dev = 1.2). The three graphs in Fig 2 show a graphical summary of the computed trend lines and illustrate the strong correlations between variables. These empirical results suggest that there are strong and significant associations between stakeholders' data overview and getting insights into relevant information about the injury data ($r = 0.6$, $p \leq 0.02$), stakeholders' data exploration and generating knowledge about injury situations ($r = 0.6$, $p \leq 0.01$) and finally between stakeholders' perceived knowledge construction and making informed decisions and initiating appropriate actions($r = 0.6$, $p \leq 0.02$).
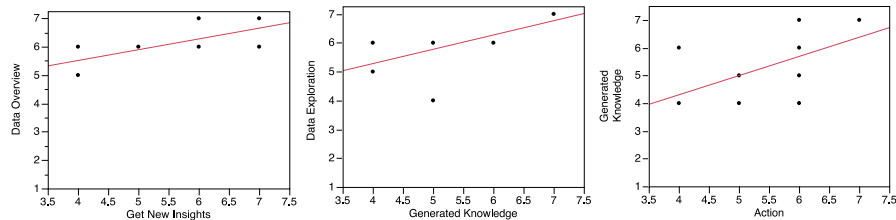
Figure 1. Correlations between variables.

The collaborative Analytics sessions promoted an engaging and cooperative environment. While the Paired Analytics served to address stakeholder's individual preferences and needs to explore specific injury data, build essential knowledge and strengthen their personal perspectives, Group Analytics enabled stakeholders to pool their expertise and fuse various perspectives and scenarios about the injury situations to collaboratively explore and accurately analyze the data.

3. Discussion

Visual Analytics proved to be powerful in helping medical stakeholders reveal valuable information about massive injury data and construct fundamental knowledge. It offered stakeholders advanced visualization techniques such as zooming and filtering, details on demand and brushing and linking capabilities to enhance their cognitive and perceptual skills and allow them to competently understand the injury data. The empirical results showed that there are strong correlations between variables, which emphasized existing associations between exploring and understanding data and the ability to get insights, generate knowledge and make informed decisions.

Collaborative Visual Analytics should be incorporated into the medical community to enable medical researchers and analysts with limited time to explore multidimensional and complex medical data, synthesize valuable information and generate decisive knowledge that is critical to initiate appropriate actions. These empirical results provide insightful approaches for future work related to collaborative Visual Analytics and its implications for the medical community to help medical professionals and researchers optimize the decision-making process.

References

[1] Arias-Hernandez R, Kaastra LT, Green TM, Fisher B. Pair Analytics: Capturing Reasoning Processes in Collaborative Visual Analytics. System Sciences (HICSS), 2011 44th Hawaii International Conference on. 2011. p.1–10.
[2] Few S. Dashboard Confusion Revisited. Perceptual Edge. 2007;
[3] Heer J, Agrawala M. Design considerations for collaborative visual analytics. Information Visualization. 2008;7(1):49.
[4] Keim D, Andrienko G, Fekete JD, Görg C, Kohlhammer J, Melançon G. Visual analytics: Definition, process, and challenges. Information Visualization. 2008;154–75.
[5] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. Visual Languages, 1996. Proceedings., IEEE Symposium on. 1996. p. 336–43.
[6] Thomas JJ, Cook KA. Illuminating the path: The research and development agenda for visual analytics. IEEE; 2005.

University

of

North Carolina

at Charlotte

Discover Diamonds-in-the-Rough using Interactive Visual Analytics System: Tweets as a Collective Diary of the Occupy Movement

Wenwen Dou and Derek Xiaoyu Wang and Zhiqiang Ma and William Ribarsky

University of North Carolina at Charlotte
9201 University City Blvd,
Charlotte, NC, 28262

Abstract

The phenomenally wide-adoption of social media has stimulated a new means in organizing and carrying-out modern social movements. Exemplified by the Occupy Movement (OM), rich information, including protest-related events and people's responses to those events, is posted and shared through social media sites such as Twitter. However, it is quite challenging to make sense of such valuable information in a collective manner, as it is often submerged by all the other content on Twitter. In this case study, we demonstrate the combination of computational methods (e.g., topic modeling and event detection) and interactive visual analytics in facilitating users to examine how relevant tweets can reflect a collective view of a social movement. In particular, we focus on discovering and associating key events throughout the OM. Based on the event frequencies, our system helps users to divide the movement into three distinct stages. Information regarding "what" the events were about, "when" and "where" the events occurred, and "who" were involved is extracted from the tweets to describe each stage of the movement. The resulting case studies show that we can indeed construct a collective diary of the social movement by analyzing events extracted from the content of the tweets.

Many discussions have been generated recently on the topic of social media and the effect it may play on the formation and mobilization of social movements. Social movement is a type of group action that involves individuals and organizations who focus on bringing social changes. In this process, social media serves as a medium for the masses to make their voices heard and to initiate and organize social movements. Recently, we start to see the impact of the pervasive use of social media on organizing large-scale social movement. Such usage is illustrated by the Bank Transfer Day (BTD), a Facebook-launched call aimed to move money from big financial institutions to local credit unions. While the BTD lasted for only two weeks (Christian 2011), it produced significant results, including an estimated 1 million consumers moving their accounts to credit unions (CSMonitor 2011). Through assessing whether the self-proclaimed goals have been met, it is relatively straightforward to determine the impact of such bursty and focused protests.

However, evaluating the impact of large-scale social protests with massive participation and a wide range of goals, exemplified by the Occupy Movement (OM), is an extremely challenging task. Even though such movement has subsided significantly in recently months, it is still on-going a year after the proclaimed start date–Sept 17, 2011. Although traditional media report selective excerpts about Occupy protests, they often do not adequately capture the scale, the response, or the opinions of the protestors' and citizens'. Understanding the overall Occupy Movement and the impact of its protest activities can shed light not only on how the movement was organized and evolved, but also on the effect of public policies dealing with the protests. Since the Occupy Movement is known to use social media to advertise, organize, and attract protesters (Wikipedia 2012b), we believe social media data such as tweets can be used as a direct source to depict all major events.

In this paper, we demonstrate the use of our visual analytics system, which combines interactive human-discoveries and a set of computational methods (including topic modeling and event detection), to enable domain users to effectively extract major events throughout the Occupy Movement. An interactive visual interface is designed to present the events in an intuitive manner for making sense of the causes, the actions and the impact of the events. More importantly, through presenting information regarding "who, what, when and where" of the protests, our system facilitates the domain users (e.g., law enforcement officers) to the construction of a collective diary of the movement using tweets.

DATA AND COMPUTATION ARCHITECTURES

Different from previous work (Marcus et al. 2011; Shamma, Kennedy, and Churchill 2010), in this paper, our system focuses on extracting event-related insights from the content of the tweets. Such analysis is conducted based on the tweets collected from Twitter's GardenHose API from 07-11-2011 to 09-18-2012. To provide more focused analyses on the movement, this tweet collection is further filtered with a general hashtag #occupy*. The resulting dataset captures a wide range of relevant tweets (∼430,000) and hashtags, including ones created to represent ideas and concepts exemplified by #occupydemocracy, #occupyteaparty, #occupybank, etc.;
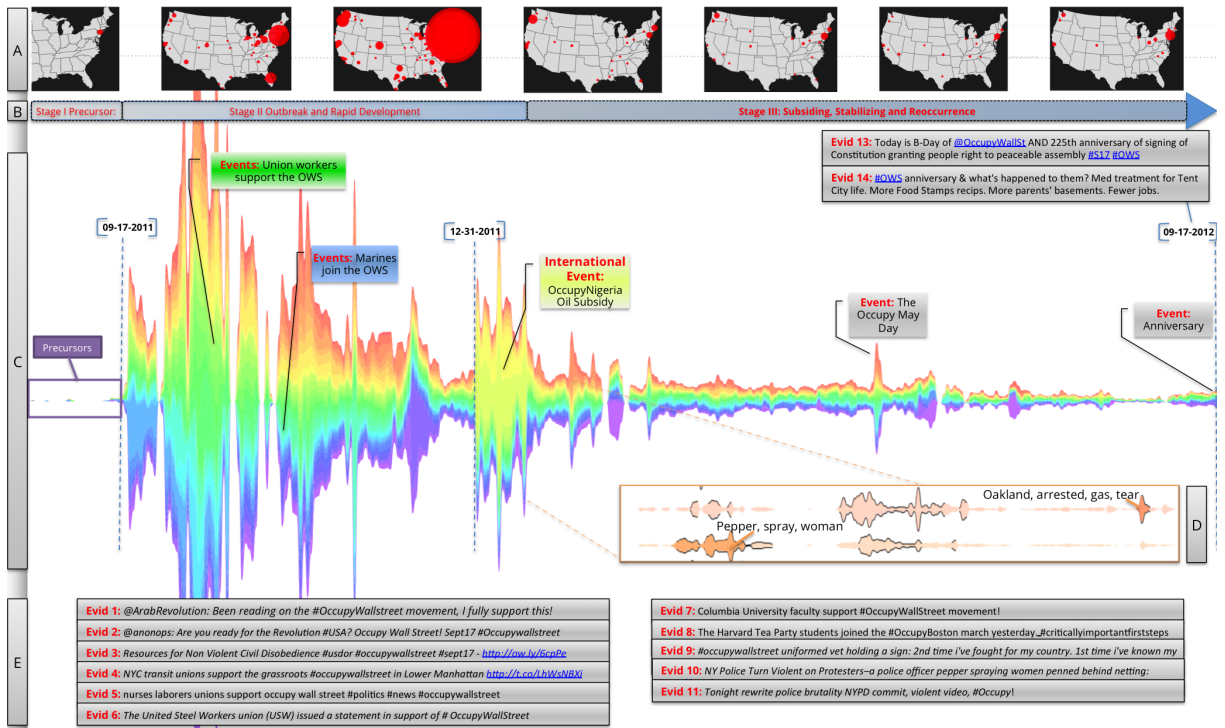
Figure 1: Overview of the OM in our VA system. *A*: Occupy hotspots over time. *B*: Three stages of the OM divided based on the rise and fall of the overall activities. *C*: Visual summary of the Occupy activities. The x-axis represents time; each color-coded ribbon represents a topic extracted from the tweets. Event detection is performed on individual topic to identify bursts as indicators of events. *D*: Sample events labeled with corresponding keywords. *E*: Evidence [Evid.X] mentioned in the paper. Note, here multiple visualizations are placed together due to the page limit. Detail about other views refer to (Dou et al. 2012)

and others that denote specific protest locations such as #OccupyBoston, #OccupySeattle, #OccupyMadrid, and etc.

The combination of the automatic data analytics and human-discovery is at the heart of our visual analytics system. Using the computational architecture introduced by Wang et al. in (Wang et al. 2012), we first applied data cleaning-and-aggregation processes to remove the contextual noise in the tweets. Then we performed topic modeling and event detection to extract meaningful insights, such as topical trends and patterns (Dou et al. 2012). Furthermore, we also sifted through features like Hashtags to extract key geospatial and temporal features.

The gist of the analysis environment provided by our system can be thought of as first organizing all tweets based on topic and time, which is visualized in Figure 1 B with each stream representing all tweets that focus on the specific topic. With the X-axis being time, the volume (or the thickness) of the stream denotes the amount of tweets discussing the topic on a given day. Then an event detection method (Dou et al. 2012) was applied to identify bursts within each topic stream as indicators of events. Figure 1 D illustrates the resulting events highlighted by black contours and labeled with a few keywords that best describe the event. In summary, our computational architecture extracts events in the form of topic bursts comprised of tweets discussing the event. The visual interface further presents the

extracted events throughout the OM in a temporal flow, with keywords indicating what each event was about.

Human-discovery is involved in this process at any given time. Users can then interactively make sense of the individual events through the visual interface, examining the tweets discussing a specific event to discover information regarding "who, what, when, and where", as well as analyzing people's responses to certain events. Altogether, we provide users a interactive visual analytics system that helps them effectively construct a diary about a social movement.

CASE STUDY: DEVELOPMENTAL OVERVIEW OF THE MOVEMENT

Based on the major rise and fall of the overall Twitter discussions of the Occupy protests, we as users can divide the movement into three major stages as shown in Figure 1.

Stage I: Precursor and Preparation

Are Occupy protests orchestrated or spontaneous?

In Figure 1, our system confirms the self-proclaimed start date of the OM and clearly illustrates its outbreak (significant boost in volume) on 09-17-11. However, the more interesting and provoking pattern (highlighted in purple) lies in the activities before the official start date. Such pattern is picked up by our computational methods and strongly signals a precursor to the protest. This precursor let us believe

that the OM is an orchestrated event rather than a spontaneous gathering. It further provides us insights on how the movement was organized and the major players involved in this preparation stage of the movement.

Specifically, the earliest tweet in our dataset was posted on 07-21 with content showing strong support of the movement [Evid.1]. With our interactive visualization, one can directly link to a tweet pointed to a blog published by Adbusters (a Canadian-based anti-consumerist organization) on 07-13. This tweet proposed a peaceful occupation of Wall Street to protest both corporate influence on democracy and the increasing inequality in wealth distribution (Adbusters 2011). In addition, our geographical visualization suggested that, at this planning stage of the movement, all geo-entities extracted from the tweets are directly concentrated on New York City (Figure 1 map left), using both GPS location and our Named Entity Recognition techniques. Based on these findings, one can infer that the Adbusters was involved in the early organization of the movement, which is later verified by other sources such as major news media (e.g., LATimes).

Our VA system further helped extract evidence that signaled the diverse contributions from other major organizers that eventually lead to the outbreaks of the movement. For example, the hacker group Anonymous posted tweets to motivate the crowd on Aug 19 [Evid.2]. Meanwhile, other contributing groups such as US Day of Rage (#usdor) advocated a non-violent movement by providing legal info for the movement on 08-27-2011 [Evid.3]. Just by quickly glancing through our visualization of tweets from this stage, one can already identify these three major groups that helped organizing the Occupy Movement, and further determine the orchestrated nature of the movement.

Stage II: Outbreak and Rapid Development

Who helped OM jump on the public bandwagon? Who joined the protest? And for what cause?

The organization of the OM has led to a successful outbreak on the 09-17. As suggested by the significant traffic jump on the #occupy topic in the following two months, one can tell that the movement entered a rapid development stage. As further evidenced in the maps in Figure 1, the rapid, almost instantaneous, geospatial spread was another contributing factor to the successful outbreak. The key to decode such rapid growth and to assess its impact lies in users' depiction of the demographics and the motivations of the people contributing to the movement.

Through interacting with the event results from our computation methods, one can quickly identify a few representative groups that participated and influenced the Occupy Movement. Through examining tweets discussing the events, the cause for their endorsement of the movement can be derived. For example, our visualization clearly suggested (Figure 1) that, on 09-28, the New York Transit Workers Union voted to support the OM, bringing hundreds more protestors on the street [Evid.4]. Furthermore, our Hashtag analysis revealed that this also led to a much larger endorsement from several more labor unions supporting the OM, only within days from prior move [Evid.5,6]. Only a few days later, our visualization (the blue event rectangle) further

indicated that burst topics labeled with keywords "marine, country, fought" emerged on 10-02. This captured the facts that veterans joined the movement to act as a first line of defense between the police and protesters (DailyKosGroup 2011)[Evid.9]. One week later, university faculties and students showed their support to the OM on 10-10. [Evid.7&8], where they banded together to make their voices heard, citing the rising amount of student loan debt and the increasing cost of college (HuffingtonPost 2011).

The massive endorsements from these diverse groups, in such a short time period, suggested that the success of Occupy protest largely relies on the wide spectrum of citizen participation. Some came to protect the protestors, while others wanted to voice their frustrations. Using the analysis and visual interaction, one was able to quickly identify the groups and infer their motivations for endorsing the OM.

Is there violence?

Although the OM was set to hold peaceful demonstrations as advocated in Stage I, a few events with violence-related keywords raised our suspicion. With large numbers of confrontations between the protestors and the police throughout the movement, however, some protests inevitably involved violent acts. Judging by the keywords associated with the events, one could quickly identify a few events that were associated with violence. On Sep 24, an event associated with keywords "pepper, spay, officer" (Figure 1 D) emerged, and the discussion of such violent acts lasted for at least two weeks on Twitter. There were tweets accounting for and condemning the violence that police used in stopping protestors and voiced for justice [Evid.9,10]. A week after that incident, people started to report more violence via Twitter, as shown in [Evid.11]. These kind of events attracted lots of public attention and media coverage (Observer 2011), which pressured NYPD to speak publicly about and investigate the incident. More recently, on Aug 3, 2012, New York City has reported to opt out to defend the officer pepper-spraying two female protesters.

Another event labeled with "oakland, arrested, tear, gas" (Figure 1 D) occurred on Jan 28, 2012 also suggested the involvement of violence. The message *"#OccupyOakland being teargassed smoked bombed & shot at w rubber bullets."* has been retweeted multiple times. Later on during the event, people tweeted *"It appears that an Iraq War Veteran was arrested #occupyoakland"*, and expressed dismay in *"Mass arrests in #Oakland. SO un-American. Sickening to watch."*

Through interactive examination of the activities throughout the movement, one can quickly depict events that occurred at different times but of similar violent nature, ascertaining the involvement of violence in the protests.

Stage III: Subsiding, Stabilizing and Reoccurrence

After the outbreak and rapid rise of the OM, the discussion of domestic (U.S.) occupy activities on Twitter started to subside significantly after the first week of Jan 2012 (shown in Figure 1). Multiple US-based events before the sudden decrease were related to evictions of the Occupy encampment. It is, therefore, reasonable to infer that the sudden drop in tweets was due to such eviction policies. This finding is validated by evidence in a recent article on a brief look back

at the Occupy Movement. After a series of evictions started in New York City on 11-15-2011, Occupy lost its ability to organize without places to gather (ABCNews 2012).

After the evictions, domestic Occupy activities started to stabilize, as the general tweet volume shrunk (Figure 1). One small climax of the Occupy activities returned on May 1, 2012 as the Occupy May Day was organized as an effort to re-energize the movement (ABCNews 2012). More recently, activities during late August and early September were protests during the Republican National Convention and the Democratic National Convention (Charlotte, NC shown as a hotspot in the map to the right). Although the protests were organized weeks before, some were disappointed in the low turnout of the participation. When the Occupy Movement marked it first anniversary on Sep 17, 2012, many were cheerful [Evid.13], while others were not quite happy with the progress so far [Evid.14].

DISCUSSION

To analyze the overall OM, our visual analytics system first facilitated users to divide the movement into three major stages based on the rise and fall of the activities. Within each stage, users can visually select a few representative events that are automatically discovered using the computational methods (Wang et al. 2012; Dou et al. 2012) . Due to the page limit, we focus the discussion mainly on domestic events throughout the OM. But it is worth noting that the movement is widely spread across more than 30 countries (as discovered in our analysis) in addition to the US. As a matter of fact, one significant sub-event (in terms of strength and length) detected by our system was the OccupyNigeria (Figure 1 light green) that was widely participated in Nigerian because of the end of government oil subsidies.

Precision and Recall of our VA System

To validate the accuracy of our automatically extracted events, we compared our results against the only comprehensive timeline of the OM available through Wikipedia (Wikipedia 2012a; 2012b). The Wiki timeline listed 41 major events while we have 60 events extracted from our Occupy related tweets. Two experimenters independently graded the accuracy of extracted events against the Wiki timeline, producing average recall of 79%. In terms of precision, we did not limit the baseline to the Wikipedia Occupy timeline. Instead we used news media to validate the existence of our identified events, which produced a high precision at 95%. Despite the precision and recall results, we acknowledge that there may be other events missed by our detection, should there be a well-compiled timeline of events throughout the OM. There are also cases that multiple co-occurring events appearing to be one big event in the detection results. Therefore, more user interactions need to be supported to allow users to refine the automatically discovered event results. However, we believe our approach provides a novel method to construct post-hoc diary of the OM and lead to comprehensive understandings of the movement. The system not only enables the depiction of meta-information of the "when" and "where", but also provides content information regarding the scale of the event, its lasting time period, and the responses to the event from protestors and citizens, etc. Through presenting the case study, we hope to showcase the capability of our computational methods and interactive visual interfaces in analyzing and reconstructing social movements in a collective manner.

Conclusion

In this paper, we showcase a novel visual analytics system that facilities the users to construct a collective understanding of the rise and the fall of the social movements. Exemplified in studying the OM on Twitter, our system can effectively help users to identify different stages of the movement. Through studying the events near the shifts between the stages, it can further lead to the discovery of the probable cause for the changes. Last, examining tweets discussing certain events, the system helps users to derive information including who were involved, where were the hotspots, people's opinions, and many others alike.

References

ABCNews. 2012. A look back at the rise and fall of occupy. http://abcn.ws/OYiVo2.

Adbusters. 2011. Occupywallstreet. http://bit.ly/n4va4T.

Christian, K. 2011. Bank transfer day. http://www.facebook.com/Nov.Fifth.

CSMonitor. 2011. Bank transfer day: How much impact did it have? http://bit.ly/vWaOgv.

DailyKosGroup. 2011. Marines join ows movement. http://bit.ly/S3PuQB.

Dou, W.; Wang, X.; Skau, D.; Ribarsky, W.; and Zhou, M. X. 2012. Leadline: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*.

HuffingtonPost. 2011. College sympathizers of ows walk out of class in support. http://huff.to/pnWOmS.

Marcus, A.; Bernstein, M. S.; Badar, O.; Karger, D. R.; Madden, S.; and Miller, R. C. 2011. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, 227–236. New York, NY, USA: ACM.

Observer. 2011. Is ray kelly's nypd spinning out of control? http://bit.ly/S4v6Ev.

Shamma, D.; Kennedy, L.; and Churchill, E. 2010. Conversational shadows: Describing live media events using short messages. *International AAAI Conference on Weblogs and Social Media*.

Wang, X.; Dou, W.; Ma, Z.; Villalobos, J.; Chen, Y.; Kraft, T.; and Ribarsky, W. 2012. I-SI: Scalable Architecture of Analyzing Latent Topical-Level Information From Social Media Data. *Computer Graphics Forum* 31(3):1275–1284.

Wikipedia. 2012a. Occupy movement. http://bit.ly/tMOGfp.

Wikipedia. 2012b. Occupy wall street timeline. http://bit.ly/qRcMEE.

HierarchicalTopics: Visually Exploring Large Text Collections Using Topic Hierarchies

Category: Research
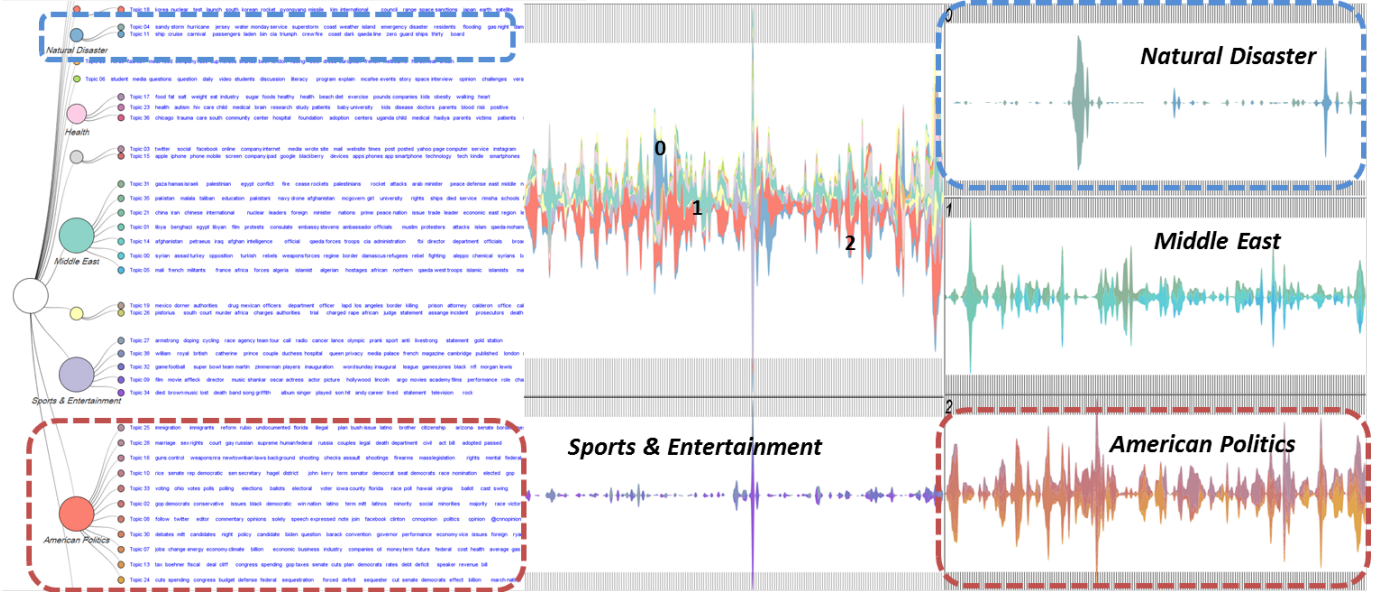
Fig. 1. Overview of the HierarchicalTopics system. The Hierarchical Topic structure is shown on the left in a tree visualization. The Hierarchical ThemeRiver view on the right presents the temporal pattern of topics in a hierarchical fashion. The dataset being visualized is the CNN news corpus. Topics are organized into 5 categories and annotations are attached to describe each news category. The corresponding categories in both view are outlined with same colors.

Abstract—Analyzing large textual collections has become increasingly challenging given the size of the data available and the rate that more data is being generated. Topic-based text summarization methods coupled with interactive visualizations have presented promising approaches to address the challenge of analyzing large text corpora. As the text corpora and vocabulary grow larger, more topics need to be generated in order to capture the meaningful latent themes and nuances in the corpora. However, it is difficult for most of current topic-based visualizations to represent large number of topics without being cluttered or illegible. To facilitate the representation and navigation of a large number of topics, we propose a visual analytics system - HierarchicalTopic (HT). HT integrates a computational algorithm, Topic Rose Tree, with an interactive visual interface. The Topic Rose Tree constructs a topic hierarchy based on a list of topics. The interactive visual interface is designed to present the topic content as well as temporal evolution of topics in a hierarchical fashion. Rich user interactions are provided for users to make changes to the topic hierarchy based on their mental model of the topic space. To qualitatively evaluate HT, we present a case study that showcases how HierarchicalTopics aid expert users in making sense of a large number of topics and discovering interesting patterns of topic groups. We have also conducted a user study to quantitatively evaluate the effect of hierarchical topic structure. The study results reveal that the HT leads to faster identification of large number of relevant topics. We have also solicited user feedback during the experiments and incorporated some suggestions into the current version of HierarchicalTopics.

Index Terms—Hierarchical Topic Representation, Topic Modeling, Visual Analytics, Rose Tree

---

1 INTRODUCTION

Digital textural content is being generated at a daunting scale, much larger than we can ever comprehend. Vast amounts of content is accumulated from various sources, diverse populations, and different times and locations. For example, 1.35 million scholarly articles were published in 2006 alone [17]. With an average annual growth rate of 2.5% [29], research articles are currently being published at the pace of approximately 4400 titles per day. In the social media world, people are contributing to the accumulation at an even faster pace. By June 2012, Twitter is seeing 400 million tweets per day [30]. Meanwhile, 900 million active Facebook users have been busy sending 1 million messages every 20 minutes [27]. Today, part of the content (e.g, tens of thousands of different sites, Twitter, digitized books) is archived in the US Library of Congress with more than 300 terabytes in size,

which keeps on growing [11].

It is generally agreed in government and industry that valuable but latent information is hidden in the vast amount of digital textual content. For instance, in scientific research, one of the crucial investigations is on the development of science. To this aim, researchers have created maps of science [24, 26] and evaluated the impact of science funding programs [14] by analyzing research publications and proposals. For emergency response agencies, sifting through massive amount of social media data could help them monitor and track the development of and response to natural disasters, as illustrated in the use of Twitter to reach victims from Hurricanes [33]. Last but not least, the emergence of numerous social media startups shows that profitable marketing and business analytics insights that can be extracted from

such content. To extract insights and make sense of large amounts of textual data, efficient text summarization is therefore much needed.

In this regard, topic models have been considered as the state-of-the-art statistical methods to extract meaningful topics/themes for summaization. Although powerful, topic models do not provide meanings and interpretation; human must be involved [7]. To enhance the interpretations of topical results, visual text analytics researchers have designed algorithms and visual representations that make the probabilistic topic results legible and exploratory to a broader audience [8, 9, 10, 14, 15, 25, 32]. Examples of the utility of these topic-based visualization interfaces include the analysis of social media users based on the content they generated [21], depiction of the temporal evolution of topics [14, 25], and identification of interesting events from news and social media streams [8, 15]. Many of these topic-based visualization systems have been studied through use cases and regarded powerful in aiding text analysis processes.

However, current visual text analytics systems have limitations. In contrast to the common practice of extracting hundreds of topics from large document corpora in the topic model community [2, 4, 20, 28, 31], current systems usually only manage to effectively represent a small number of topics. As more textual data becoming available, the number of necessary topics for interpretable text summarization will grow inevitably. Only extracting a small number of topics, therefore, won't capture the nuances in the corpora. As the number of topics increase, sifting through and comprehending all the topics becomes a time-consuming and laborious task, which will be further hampered by the visual clutter introduced when displaying the temporal evolution of hundreds of topics with no organization.

In particular, three challenges must be met to effectively analyze document collections that are summarized by large number of topics:

1. **How to organize the topics to facilitate the navigation and analysis within the topic space?** Without organization, sifting through a hundred topics with each topic consisting of 20 or more keywords could be intimidating. One example that highlights the problem is that when developing the NSF Portfolio Explorer, it took days for a researcher to manually examine a thousand topics to select 30 topics for further analysis and visualization [12]. Since certain topics are closer in meaning than others, organizing semantically similar topics into topic groups will ease the navigation in the topic space. Having an automated classification of topics could potentially jumpstart the analysis of text collections based on large number of topics, however, the automated classification may not always conform to individual users' mental model of the topics space.

2. **How to visually convey and permit user interactions with the organized topic results so that users can classify the topics based on their interests?** It is essential to place users in the center of the topic analysis process, allowing users to leverage and modify the topic classification results. For example, when analyzing a news corpus, a user may want to organize the topics into a hierarchical structure through first categorizing the news topics into either domestic or foreign news. In addition, for domestic news topics, the user may want to further divide the topics into groups such as politics, sports, entertainment, etc. Similarly, when analyzing topics from Twitter streams, a business analyst may be interested in grouping all topics related to sales and customer services and further divide them into more refined categories. Therefore, intuitive topic visualizations and user interactions are needed to support the analysis and modification from an initial topic organization provided by an automated algorithm.

3. **How to modify existing visual metaphors to accommodate the organization of a large number of topics?** After a user has identified a desirable hierarchical topic structure, the third challenge lies in tailoring existing visual representations. Visualizing temporal evolution of topics has been considered essential to understanding various domains (e.g. scientific fields, breaking news, etc.) over time. However, ThemeRiver [16] and stack graph that are commonly used to present the temporal trends of the topics do not convey hierarchical information. To enable the analysis and comparison of temporal behavior of topic and topic groups, it is essential to extend the current visual metaphors to incorporate hierarchical structure of topics.

To tackle the three challenges, we propose HierarchicalTopics (HT), a visual analytics system [1] that supports scalable exploration and analysis of document corpora based on a large number of topics. HierarchicalTopics *addresses the first challenge* by integrating a novel algorithm that automatically classifies topics into a hierarchical structure. Through joining similar topics into the same group, the new organization of topics provides scalable representation and navigation in the topic space. HierarchicalTopics further incorporates visual representations and interactions that embrace the hierarchical organization of the topics, and enables the users to depict the temporal evolution of topics or topic groups. In addition, rich user interactions are provided in HT to *address the second challenge*. Along with the visual representations of the topic hierarchy, HT allows users to modify and update the automatically computed topic groups. It therefore supports the customization of the visualizations based on the users' analytical interests. *To address the third challenge*, a new Hierarchical ThemeRiver has been designed to accommodate the hierarchical organization of the topics. The Hierarchical ThemeRiver eases the exploration of temporal behaviors of topic groups, and enables the comparison of topic groups on a temporal dimension. Through tight coordination between the visualizations of topic hierarchy and hierarchical temporal trends, we intend to provide an inviting interface that supports making sense of large document collections via navigating through large number of topics and their temporal evolution.

We have assessed the HT through both qualitative and quantitative evaluations. To evaluate the system in a qualitative manner, we present a case study in which an expert user performed in depth analysis on a collection of 11,961 NSF awarded proposal abstracts. To evaluate HT in a quantitative fashion, an 18-participant user experiment is conducted to compare the HierarchicalTopics system to a non-hierarchical representation based on a CNN news corpus that contains 2453 recent news articles. The experiment results reveal that the hierarchical topic visualization leads to faster identification of a large number of relevant topics. Constructive user comments were also collected during the experiment. After the user study, some suggestions on improving the visualization and interactions from the participants have been incorporated into the current version of the HierarchicalTopic system.

The rest of the paper is structured as follows: we introduce the previous work that inspired the design of HierarchicalTopics in Section 2. Section 3 focuses on introducing the HierarchicalTopics, including its system architecture and interactions. We present a case study in Section 4, followed by descriptions of a user study in Section 5.

2 RELATED WORK

Two lines of work inspire the design of HierarchicalTopics, namely topic models and topic-based visualizations.

2.1 Topic Models

Topic models can be effective tools for text summarization and statistical analysis of document collections [2]. The number of topics needed is typically determined by the size of the text corpora. The larger the size the more topics are preferred to ensure topic comprehension and human interpretability, typically tens of thousands of articles will require topics in the scale of hundreds. Specifically, one school of topic models is based on a human-defined number of topics. Here researchers and practitioners usually generate a large number of topics to capture the themes that pervade the text collection as well as the nuances. For instance, in the experiment of evaluating the collaborative topic model [31], the authors extracted 200 topics from a paper-abstract collection with 16,980 articles and a vocabulary size of 8000. In other non-parametric Bayesian topic models, such as the hierarchical Dirichlet process (HDP) [28] and the discrete infinite logistic

---

[1]A video of the HierarchicalTopics can be found at http://youtu.be/Vi1FP5kAbOU.

normal distribution (DILN) [20], the number of topics is determined by the model. However, it is evidenced that such algorithmically generated number of topics is typical rather large. For example, in the experiment evaluating DILN, the model produced 50 to 100 topics given a fairly small dataset with only 3000 to 5000 news articles.

Such large number of topics are prone to be challenging to the human interpretations and sense-making process. Much research has been focused on revealing the correlations between latent topics and organizing topics into more human interpretable structures. Work in this area aims to facilitates the navigation through the topic space and enables the discovery of documents exhibiting similar topics. While most of the existing topic models do not explicitly model correlations between topics, a few exceptions have directly accounted for relationships between latent topic themes. For example, both correlated topic model (CTM) [4] and DILN [20] have demonstrated better predictive performance and have uncovered interesting descriptive statistics for facilitating browsing and search. Although the topic correlations have been modeled, it is still difficult for users to take advantage of the descriptive statistical relationship of topics without an effective organization and visual representation of the topics.

Many researchers consider that organizing topics into a hierarchical structure presents a scalable solution to improve human-interpretability of topic. To this aim, Blei et al. have proposed a hierarchical topic model (hLDA) that learns topic hierarchies from data to accommodate a large number of topics [3]. The hLDA is a flexible, general model for extracting topic hierarchies that naturally accommodates growing data collections. However, the topic hierarchies hLDA produced are rather rigid since the depth of such hierarchies is predefined and fixed throughout the modeling process. In addition, the higher level topics generated by hLDA usually consist of stopwords, therefore less meaningful for human users.

In order to leverage the scalable hierarchical structure without enforcing rigid restrictions on the topic models, we developed an algorithm, Topic Rose Tree, to construct a multilevel hierarchical structure with any given number of generated topics. Together with interactive visualizations, our HierarchicalTopics system enables users to explore and iteratively update the topic hierarchy. In this way, our system aims to improve human-interpretability by enabling users to tailor the hierarchical topic results to their own analytical interests or mental models of the topic space.

2.2 Visualization based on Topic Models

The power of topic models in summarizing and organizing large text corpora has been widely recognized in the visualization community. A good number of visualization systems have been developed based on topic models for users to comprehend document collections.

As one of the pioneer visual text analysis systems, TIARA [32] combined topic models and interactive visualization to help users explore and analyze large collections of text. Specifically, TIARA utilized a stack graph metaphor to represent temporal change of topics over time. Similarly, another system ParallelTopics was also developed to depict both temporal changes of topics using ThemeRiver and the characteristics of documents based on their topic proportions via Parallel Coordinates [14]. Since temporal evolution of the topics has been considered one of the most useful features of the topic-based visualizations, researchers have extended a great deal in this direction. TextFlow [13] presented a novel way to visualize topic birth, death, and merge that signify critical events. In a similar vein of identifying events, LeadLine [14] applied event detection methods to detect "bursts" from topic streams and further associate such bursts with people and locations to construct meaningful events. Furthermore, Chae et al. proposed a visual analytics approach that supports the analysis of abnormal events detected from topic time series [8]. Instead of representing and analyzing topics along the temporal dimension, Lee et al. proposed a visual analytics system for document clustering based on topic modeling [18]. Users could guide the clustering process through adjusting term weights in the topics.

These topic-based systems have demonstrate the effectiveness of combining topic models with interactive visualizations in facilitating

analysis of text corpora. As indicated in in most of their reported case studies, however, these systems only dealt with a fairly small number of topics. This is quite contrary to the common practice in the topic modeling community, where a lot more topics are generated for a text collection of similar size (Section 2.1).While a greater number of topics will inevitably introduce visual clutter and legibility issue to the visualization systems, limiting the topic number may also hamper users' ability in comprehending the text collection.

Therefore, more scalable approaches to organizing the topics and visual representations based on the topics are much needed to support real-world challenges of analyzing large text corpora. To meet this need, HierarchicalTopics provides a scalable solution that allows iterative analysis of document collections with a large number of topics and further supports the exploration of temporal evolution of those topics in a hierarchical fashion.

3 HIERARCHICALTOPICS

3.1 System Pipeline

As illustrated in the overall system architecture in Figure 2, HierarchicalTopics is a user-centered analysis system that integrates computational methods with interactive visualizations. HT systematically incorporates both online and offline computations and utilizes scalable infrastructures described in [withhold for anonymity], including MapReduce and Parallel Processing. There are four key processing stages in the HT architecture including two offline computation modules (e.g., Data collection, and preprocessing and Parallel Topic Modeling) and two online components (e.g., Topic Rose Tree and Hierarchical Visualizations).

In particular, HT accommodates digital text content from various sources such as social media, research publications, news, etc. Once the data is collected, it will be streamlined into HT's data cleaning and preprocessing step, as shown in Figure 2A. In this process, HT first unifies the formats of input data and converts certain documents (PDFs) to proper topic-model-readable text files. It further prepares the documents for parallel topic models by removing stopwords, punctuation, and emojis.

The cleansed data then goes through the topic modeling stage (Figure 2B), which extract topics from the document collection. It is worth noting that the choice of the topic model component in HT is rather flexible. The architecture of HT is set to utilize a variety of topic models and can leverage their unique strengths such as interpretability [7], convenience of non-parametric models [20, 28], and accounting for additional metadata [22, 23], etc. As reported in paper, HT has successfully incorporated both the vanilla LDA [5] and the Author Topic Model (ATM) [23] to handle the natures of different text corpora.

After the first two stages are accomplished offline, the rest of the computation and visualization are computed online. The Topic Rose Tree (TRT) shown in Figure 2C organizes the probabilistic topic results into a hierarchical structure, as detailed in next section. Based on the hierarchical topic organization, two coordinated interactive visualizations (Figure 2D) are designed to present and support interactive analysis of topics and temporal evolution of the topics.

The TRT and the visualizations are closely coupled through the rich interactions provided by the HierarchicalTopics system. In particular, the three essential operations in the TRT algorithm (e.g., join, absorb, and collapse) are directly incorporated in the visualizations and interactions. Through direct visual manipulations, HT allows the users to perform the same operations to modify the initial topic hierarchies and iteratively derive the most interpretable topics groups based on their analytic interest.

In the rest of this section, we will focus on presenting details of the online components of HierarchicalTopics.

3.2 Topic Rose Tree

Our goal in designing the Topic Rose Tree is to support scalable visual representation and exploration. TRT is an automated method that can meaningfully organize a list of topics into a hierarchical structure. Its core algorithm is built upon key concepts from the Bayesian Rose Tree (BRT), which constructs a hierarchy using hierarchical clustering
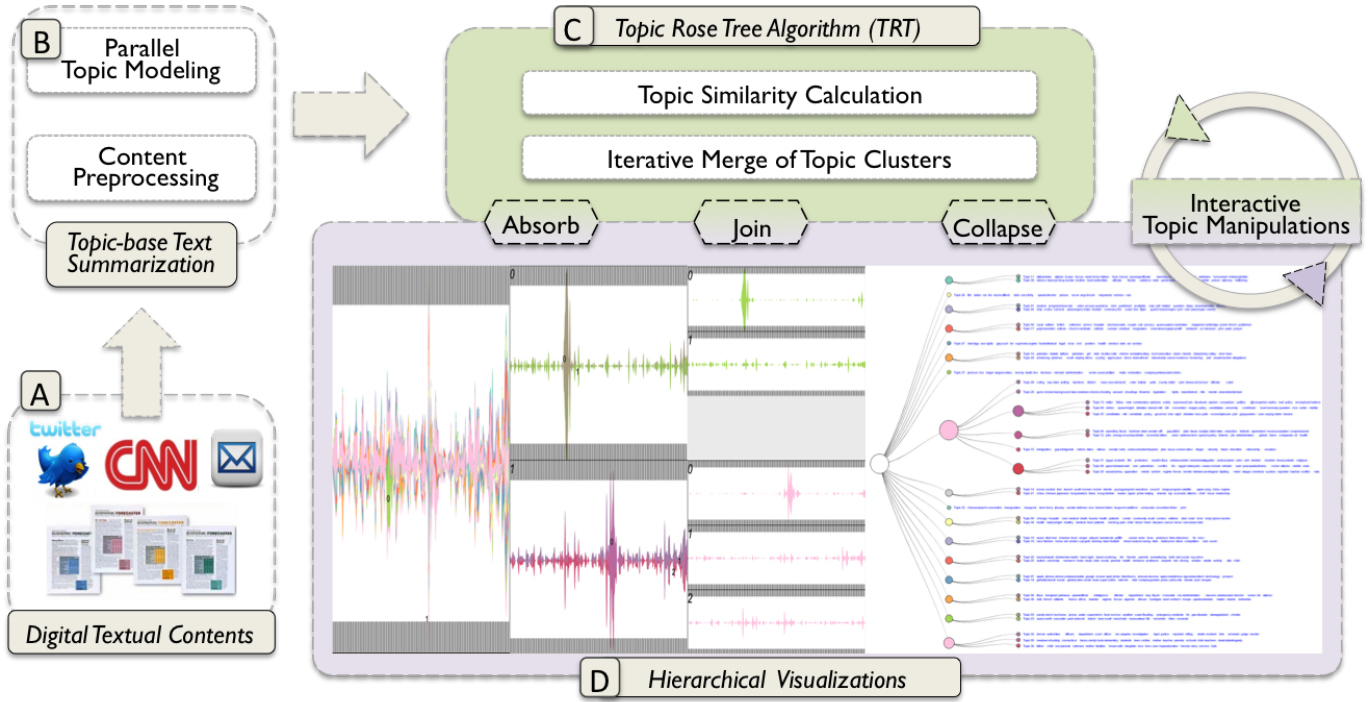
Fig. 2. System Architecture of HierarchicalTopics. Starting from bottom left, textual data is first harvested (A). The data then goes through a preprocessing stage before entering the topic model component (B). These two steps are completed offline. The resulting statistics from topic models then serve as input to the Topic Rose Tree (C), which constructs a hierarchy given a list of topics. The topic hierarchy is then visualized in the interactive visual interface (D) for users to analyze the topics and temporal trends in a hierarchical fashion to derive understanding of the text collection.
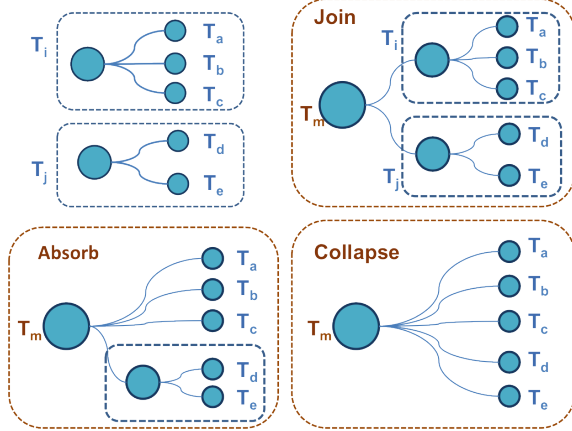
Fig. 3. The three essential operations of our Topic Rose Tree algorithm.

methods [6]. Compared to previous hierarchical clustering methods that limit discoverable hierarchies only to those with binary branching structures, BRT produces trees with arbitrary branching structure at each node, known as rose trees [6]. We consider such characteristic more natural in organizing topics, since any number of topics could be similar and should be grouped into one partition in a hierarchical structure. The essence of generating a rose tree is support of the three operations, namely join, absorb, and collapse (shown in Figure 3).

Unfortunately, simply borrowing BRT and directly applying it to topic models is unfit based on our experiments. This is primarily caused by the large number of features (words in the vocabulary) from topic models. In addition to the vocabulary size of a text corpus, which is usually in the thousands, the binarized matrix of topic distributions over the vocabulary is extremely sparse, causing problems for calcu-

lating the marginal probability of the topic groups in a tree.

Therefore, we developed TRT, an algorithm that built upon the three operations to construct hierarchies specifically from topic modeling results. TRT is a one-pass, bottom up method which initializes each topic in its own cluster and iteratively merges pairs of clusters. To construct the hierarchical structure, we first compute the similarity between any pair of clusters (topics/topic groups). TRT then merges the most similar clusters using one of the three operations. In this process, the Hellinger distance, which is a symmetric measure of the similarity between two probability distributions, is used to calculate the similarity of a pair of clusters. Intuitively, topics or topic groups that share similar distributions over the vocabulary yield lower distance. To construct the hierarchy, the most similar topic (group) clusters will be merged at each step.

In particular, each topic from the topic modeling results is represented as a probabilistic distribution over the entire vocabulary given a text collection, denoted by $X_{i,v}$, with i representing the ith topic and v representing the vocabulary of size N. To represent the probabilistic distribution of a node that contains multiple topics (children), we simply compute an average of all distributions of the children's. Details of the TRT are shown in Algorithm 1.

The complexity of the topic rose tree is the same as the BRT algorithm. First, the distance for every pair of data items needs to be computed-there are $O(n^2)$ such pairs. Second, these pairs must be sorted in order to find the smallest distance requiring $O(n^2 \log n)$ computational complexity.

To showcase how the topic rose tree algorithm could group similar topics together, Figure 4 shows a partial result from the initial grouping. In this case, we used the 2011 VAST mini challenge 1 microblog data, which contains an embedded scenario of an epidemic spread. This data is good for qualitatively evaluating the algorithm since we expect similar topics regarding the epidemic spread should be grouped together. The topic group shown in Figure 4 (top) contains three topics highlighting the flu-like symptoms for the first two days of the

Algorithm 1 Topic Rose Tree

Input: Data $\mathbf{D} = \{\mathbf{X}_i, v\}, i = 1, 2, ...n$; v is the vocabulary of the corpus
Output Topic rose tree T_{n+1}, a hierarchical structure with all topics
Initialize: $\mathbf{T}_i = \{\mathbf{X}_i, v\}, i = 1, 2, ...n$
Steps:

    Denote c as cluster count
    while c $>$ 1 **do**
        for each pair of trees Ti and Tj **do**
          Calculate cost D(i,j) for 3 possible operations (join, absorb, or collapse):
$$D(i,j) = 1/2 * \sum_{v=1}^{N} (\sqrt{t_{i,v}} - \sqrt{t_{j,v}})^2, t_{i,v} \text{ denotes the prob-}$$
ability distribution of tree node T_i over the vocabulary of size N
          Find operation m which yields lowest cost for T_i and T_j
          Merge T_i and T_j into T_m using operation m
          Delete T_i and T_j, $c = c - 1$
        end for
    end while

---

epidemic (each tick on the x axis denotes a day). Another topic shown in Figure 4 (bottom) highlights evolved symptoms such as pneumonia for the third day of the epidemic. Note that since the words that were tweeted to describe the symptoms have changed a great deal, the topic rose tree did not put topic 31 into the first topic group. However, combining with the temporal patterns, one can identify when the epidemic spread started, and how the symptoms evolved over time. This example illustrates that the topic rose tree is able to group similar topics together, and the result is very much interpretable by human users.
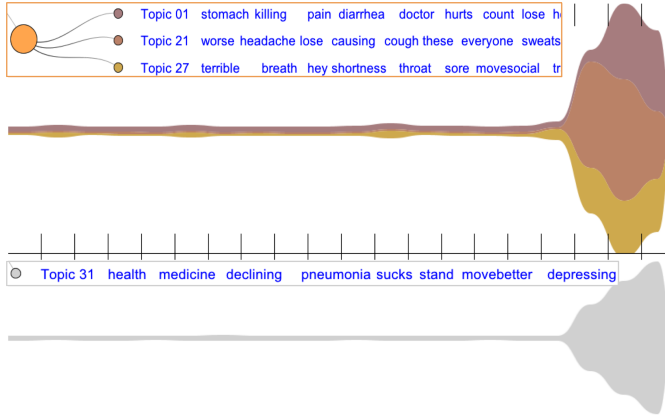
Fig. 4. An example showcases the capability of TRT grouping to group topics together. The top three topics (grouped by TRT) describe all flu-related symptoms on the first two days of the disease outbreak. The bottom topic (in grey) was not grouped in to the first group by TRT since it describes different symptoms on the third day.

3.3 Visual Components

After applying the Topic Rose Tree to the topic modeling results, a hierarchical organization of the topics is generated. To facilitate the topical analysis of the text collection, we present a visual interface that is tailored to the hierarchical organization of the topics. The visual interface consists of two coordinated views, namely Hierarchical Topic View and Hierarchical ThemeRiver. The two views are coordinated through user interactions with a focus on correlating the hierarchical information.

3.3.1 Hierarchical Topic view: Depicting topics in a hierarchical fashion

While TRT computationally alleviates the topic organization issue, the Hierarchical Topic view is designed to visually address *Challenge 1* by presenting the topic contents in a hierarchical fashion. Such representation not only offers a scalable solution as it allows the number of

topics to accrue, but also supports better navigation by grouping similar topics together. Figure 1 shows the Hierarchical Topic view with 40 topics extracted from the CNN news corpus. To provide user a familiar visual environment, we adopt straightforward tree visual representation. In this view, each leaf node represents a topic, while the non-leaf nodes denote topic groups. The first node on the left is the root of the topic hierarchy, with the rose tree spanning from left to right. The content of each topic (in the form of a group of keywords) is presented to the right of each leaf node. The size of the node is drawn proportionally to its number of children (shown in figure 1).
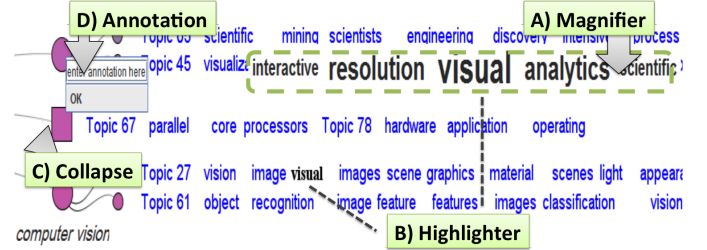
Fig. 5. Interactions provided by the Hierarchical Topic view. A) Magnifier: enlarges keywords near mouse cursor. B) Highlighter: highlight all occurrences of a selected keyword. C) Node collapsing: details of the collapsed children nodes are no longer shown. The shape of the node turns rectangular when collapsed. D) Annotation: allows users to enter annotation.

User interactions. The Hierarchical Topic view provides a rich set of user interactions to help users effectively explore and navigate through large numbers of topics. In addition to standard panning and zooming, this view employs both an on-demand magnifier and highlighter to facilitate the examination of the topic contents, as shown in figure 5 A and B. The magnifier is designed to help users to better read the topic keywords through enlarging the font near the mouse cursor, while the highlighter aims to reveal the associations between topics by highlighting all occurrences of a certain keyword in the other topics. To further help users concentrate on the topics of interests, the Hierarchical Topic view supports interactive collapsing and expanding topic groups, shown in the square node in Figure 5C. More importantly, the Hierarchical Topic view further allows users to annotate on the nodes to attach semantic meanings to topic groups (Figure 5D).

Interactive modification of the topic hierarchies. In addition to facilitating topic exploration, the Hierarchical Topic view also aims to provide an intuitive way to visually classify the topics based on users' interest. Particularly in the process of analyzing a text corpus, only human users can attach semantics to the topics and provide meaningful though sometimes subjective groupings. Therefore, it is essential to allow users to interactively modify the rose tree based on their analytical interests.
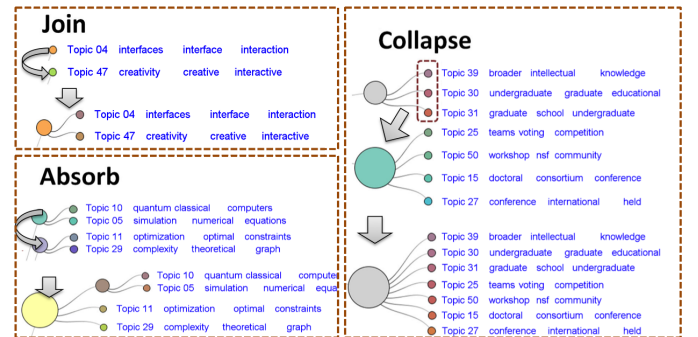
Fig. 6. Three operations supported to modify the topic hierarchy through user interactions.

To permit such modification, the three operations that are used to

construct the hierarchy in the topic rose tree algorithm are supported intuitively through drag-n-drop in the Hierarchical Topic view. As shown in Figure 6, dragging one leaf node into another constitutes the "join" operation. Drag-and-dropping any non-leaf node into another is considered as performing the "absorb" operation, while dragging multiple nodes into another node is interpreted as the "collapse" operation.

As observed in both the case study and user experiments (Section 4 and 5), the ability to iteratively refine and manipulate topic groups has demonstrated significant utility when analyzing text collections. Especially when HierarchicalTopics embodies the above three essential operations into intuitive mouse interactions, it creates a flexible text analytics environment for users to categorize, modify, and update topics and topics groups. For example, as illustrated in Figure 1, participants in our user study have used these three operations to effectively group topics into five news categories based on the initial TRT hierarchy. In addition, the annotation interaction in HT view permits the users to attach semantic interpretations of the topic groups, and further helps them to connect the dots of a large number of topics. Many of our participants agreed that such user interactions served as a potential solution to the *Challenge 2* (see Section 1).

In summary, the Hierarchical Topic view provides both a visual representation of the topic hierarchy and a rich set of user interactions to serve as the first step to effectively analyze text collections.

3.3.2 Hierarchical ThemeRiver: Representing the temporal trends of topic groups

In addition to visually representing the topics which serve as a summarization of the document collection, visualizing the temporal evolution of the topics brings a unique contribution; it permits the discovery of the rise and fall of different topic themes, as well as identifying possible critical events [13, 15].

To this aim, we extend the widely adopted temporal visualization, ThemeRiver [16], to further incorporate hierarchical information. Our goal in designing the Hierarchical ThemeRiver is to provide users the ability to analyze and compare temporal behaviors of topic and topic groups, which address the core issue in *Challenge 3* (Section 1).

As illustrated in Figure 7, the Hierarchical ThemeRiver starts with the main panel (Figure 7A), where the temporal evolutions of the highest hierarchy (children of the root node) are shown; the height of each ribbon is calculated by summing the height of its leaf nodes. Once a ribbon is hovered, a preview of the temporal evolution of the child nodes will be shown in the preview panel (Figure 7B). The panels support interactively examination of the overall temporal trends of a text corpus as well as individual topic groups.

An elastic-panel structure is built into the view to enable the users' comparison of multiple topic groups. To compare different topic groups, a user can start by selecting a topic ribbon in the main panel; such interaction will create a sub panel (Figure 7C) showing the next level of hierarchy of the currently selected node. Multiple selections can be made to view the detailed temporal evolution of different topic groups, thus enabling the comparison and association of temporal patterns. Note that sub panels are always expanded to the right of the current selection, creating a coherent look and feel of the layout as in the Hierarchical Topic view.

Color assignment. To assist user exploration as well as to keep a smooth transition between panels, we have carefully chosen 12 perceptively coherent colors for the Hierarchical ThemeRiver view. This is done in an experimental fashion using the "i want hue" system [19], with the k-Means clustering and light background option. In the Hierarchical ThemeRiver view, the 12 distinct colors are first assigned to the topic ribbons in the main panel (Figure 7A). The child ribbons of each selected parent ribbon get colors of the same hue, but with varying luminance and chorma, as shown in Figure 7C. The same color scheme is also used in our Hierarchical Topic view to provide a coherent visual cue that helps correlating the two different representations of the same topic or topic groups.

Temporal selection and details on demand. To permit the examination of documents of interests, details of the text content are shown upon selection. In any panel within the Hierarchical ThemeRiver view,
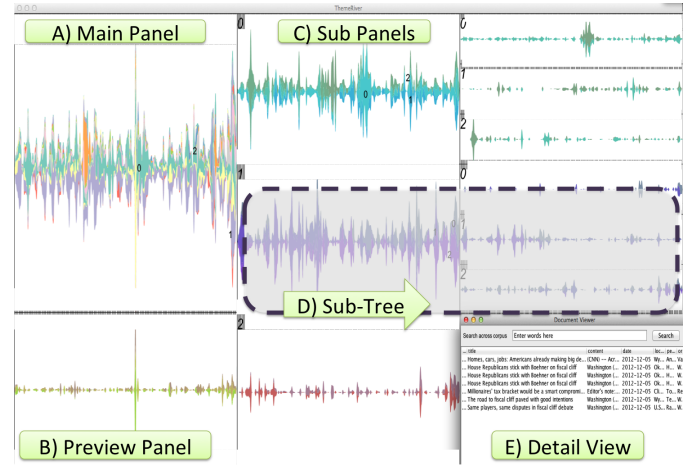
Fig. 7. Overview of the Hierarchical ThemeRiver. The dashed rectangle, in component D, highlights a sub tree created upon user interaction to view temporal patterns of child nodes.

a user can enable the "time column" mode and interactively select a subset of documents published in a certain time period. By doing so, a detail view (Figure 7 E) will be shown to help the user validate the temporal patterns and understand its cause. During the user study, for example, this operation was demonstrated useful in examining the contributing posts to a topic burst pattern.

In summary, the Hierarchical ThemeRiver view is tailored to represent temporal patterns of topic and topics groups in a hierarchical manner. The incorporation of hierarchical information is mainly achieved through user interactions and in a way that is coherent to the Hierarchical Topic view representation.

3.3.3 View Coordination

Both views in the HierarchicalTopics system are tightly coordinated. On the one hand, selecting a node in the Hierarchical Topic view would highlight a corresponding temporal panel in the Hierarchical ThemeRiver view. This helps users to examine the temporal evolution of the selected topic group. On the other hand, selecting a ribbon in the temporal view will highlight the corresponding node and its path in the topic view. More importantly, once the hierarchy is modified through user interactions in the topic view, the temporal view will also be updated accordingly to reflect the new hierarchical structure.

In summary, the HierarchicalTopics system presents both topic information and temporal evolution of the topics in a hierarchical fashion. This system is designed to aid the exploration of topic content and temporal trends of topic groups through a rich set of user interactions. In addition, our system allows users to iteratively modify, define, and annotate topic groups based on their interpretation. The HierarchicalTopics provides a flexible visual analytics environment that tightly integrates computational methods with interactive visualizations for analysis of large document collections.

4 CASE STUDY

To qualitatively access the utility of HierarchicalTopics in facilitating the analysis of text corpora with large number of topics, we recruited a senior researcher whose research interests covers HCI and Information Retrieval. This case study is set up for him to explore a collection of NSF awarded proposal abstracts to identify interesting research trends in his research domains. Eighty topics were extracted from 11,961 proposal abstracts funded by all three divisions (IIS, CCF, CNS) in the CISE (Computer and Information Science and Engineering) directorate from 2005 to 2012.

4.1 Depicting temporal portfolio of NSF programs

Using the Hierarchical Topic view, the researcher started by visually browsing all hierarchical topic groups that are produced by the TRT al-
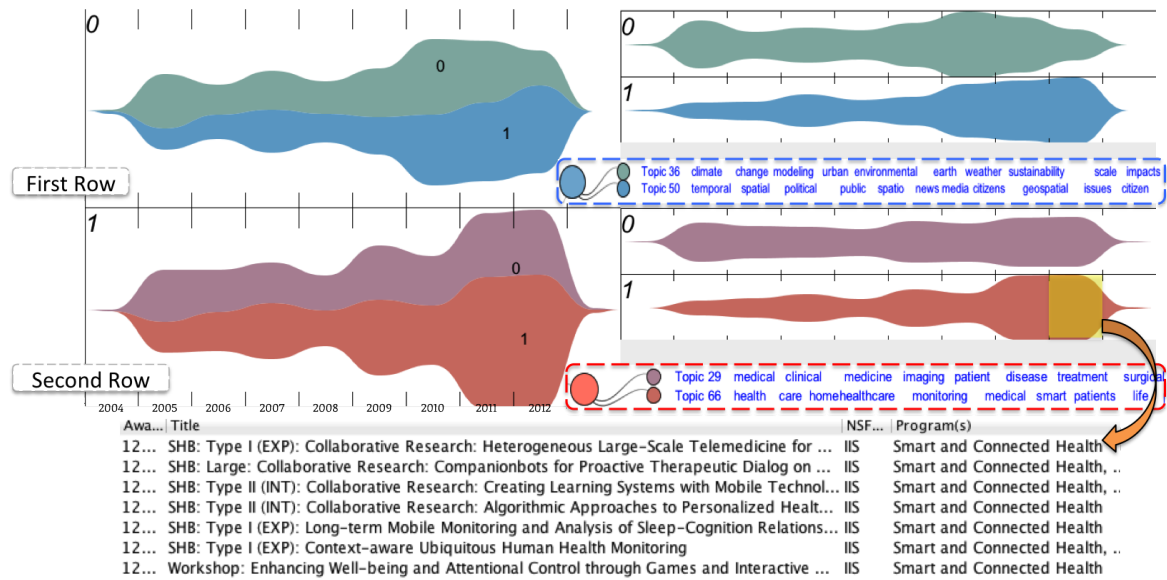
Fig. 8. Case Study: Making sense of increasing topic group trends. Top (with a blue hue): topic group of "environmental and citizen science" has seen recent growth. Middle (with a red hue): heath care related topic group exhibit growth in the past two years, with the "health monitoring" topic as the major contributor to the overall growth. Bottom: detail view showing proposals regarding the "health monitoring" topic awarded in 2012.

gorithm. He quickly identified a few topics of interest and interactively merged them into topic groups that fits his analytic goal. The result of his customized grouping and corresponding annotation is shown in the first column in Figure 9. Specifically, two groups of topics are created through the "join" and "collapse" interactions, "*HCI*" and "*Information Retrieval and Data Mining (IR)*".

With the exploration scope narrowed down to these two topic groups, the user wanted to identify and compare the trends in research funding for individual group over the years. Therefore, he turned to the Hierarchical ThemeRiver view and selected the two topic groups so that their research funding trends can be examined and compared.
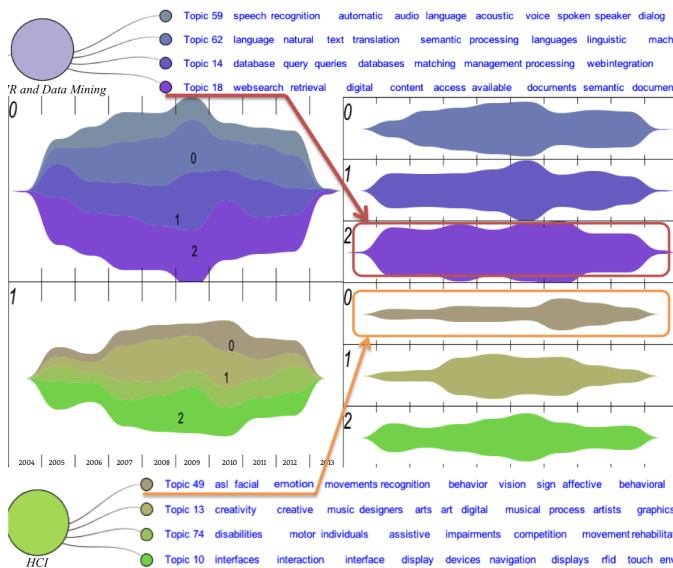
Fig. 9. Case Study: Examination of topic groups of interest. Top (with a purple hue): Topic keywords and temporal trends of the "Information retrieval and data mining" research domain. Bottom (with a green hue): Topic keywords and temporal patterns of the "Human Computer Interaction" field.

The second column in Figure 9 illustrates the overall temporal evolution of selected groups. The user noticed that the trend of proposals

awarded under the *IR* group seemed steady with a slight decline over the recent two years. To examine and compare the development of individual topics in the IR group, the users further isolated three topics that are of interest. The corresponding trends for these topics are shown next to the overall trend.

Through quickly examining the volume of each topic trend, the user confirmed his hypothesis that topic 18 on "web search and document retrieval" has continued to be a more popular subfield over the years in terms of NSF research funding (Figure 9 ribbon with red border). However, the user was also surprised when found out that the *"HCI"* group exhibits a slight decline in recent years after a steady growth around 2007. Through examining individual topic trends, more interesting patterns prevailed. Although the overall trends for other topics group have subsided slightly, the research on "*affective computing and emotion related studies*" has gone up significantly in the past two years, as outlined in Orange.

This use case illustrated that the visual interface not only enables the user to view trends for a group of topics that describe a research field, but also permits the discovery of the contributions of individual topics to the overall trends as well as anomalies. According to the user, such analysis gave him valuable insights in understanding the research trends in the areas he is interested in and could potentially help him adjust future proposal focus.

4.2 Identifying program impacts in research

Given that the above two topic groups all exhibits slight downward trending, the user wanted to identify upcoming research topics that received more funding interest in the recent years. He started by mouse hovering over each topic ribbon in the main Hierarchical ThemeRiver view, looking for increasing trends.

Two topic groups caught his attention as shown in Figure 8. Both groups exhibit increasing volume in the past three years, indicating more research proposals were awarded in the two areas. The top row illustrates a topic group related to environmental related research as well as citizen science. As shown in the individual temporal trend for each topic, the user identified that the topic on citizen science and spatial temporal analysis significantly contributed to the recent growth of the focused topic group.

The second row in Figure 8 illustrates a topic group that summarizes research on medical and healthcare related research. Through enabling the time column selection, the user selected proposals related to the health care topic that were awarded in 2012, highlighted in the

yellow rectangle. He then discovered that most of the proposals were related to health monitoring and were awarded by the only-recently launched program–Smart and Connected Health (2011).

The user was pleased to find out the impact of a newly established program on research trends and considered the HierachicalTopics a powerful tool in aiding the discovery of the contributors to the temporal changes and possibly the cause for such changes.

5 USER STUDY

To quantitatively evaluate the utility of HierarchicalTopics in aiding users analysis of a text corpus, we conducted a formal user study focusing on comparing hierarchical to non-hierarchical topic structure. Our hypothesis is that the hierarchical topic structure would yield faster identification of topics that are similar in nature.

5.1 Data and Tasks

The dataset used for the user study contains 2453 news articles published between Sept 2012 to March 2013 on CNN.com. Two conditions were designed to evaluate the effect of hierarchical topic structure versus representing them as a flat list of topics. We designed two tasks for the experiment: the first task aims to group individual topics into different news categories; the second task focuses on examining the overall temporal trends for the topics in each news category. For the second task, we required the participants to group all the topics based on their findings in task 1.

Specifically, we asked the participants to identify news topics that fall into the following five categories: American Politics, Sports and Entertainment, Natural Disaster, Health-Related Issues, and Middle-East News. An example topic grouping result produced during one of the experiments is shown in Figure 1. Note that, to control the complexity of the tasks, we extracted 40 topics from the news corpus. The reason for doing so was that the participants assigned to the non-hierarchical topic organization had to go through the topics one by one. With no initial aid of organizing similar topics together, grouping large number of topics would become laborious and require a lot of repetitions of the same operations. This implies that, if the hierarchical structure proves superior in this study, it will increase its edge relative to a flat structure as the number of topics grows

5.2 Experiment Design

Eighteen participants took part in the study (13 male, 5 female). The age of the participants ranged from 18 to 34.The study used a between-subjects design. All participants were first provided 10 minutes of training on the HierarchicalTopics visual interface. Each participant was then randomly assigned to one of two conditions (hierarchical vs. non-hierarchical topic organization). The participants were asked to write down their findings on an answer sheet, which records the identified topic numbers for each listed category for the first task and the pattern of the temporal trends for the second task. The experimenter timed the participants for completing each category while they were performing the tasks. The study was conducted in a lab setting, on a computer with two displays (resolution at 2560x1600 and 1920X1200, respectively), 2x 2.66GHz CPU and 12 GB memory.

5.3 Results

For the purpose of analyzing whether the hierarchical topic structure helps the analysis of large text corpora, we calculated the difference of average time for identifying topics for each news category. The average time is computed as the overall time to find all topics for each category, divided by the number of topics identified. The reason for using the average time is because participants identified different number of topics for a given category. In practice, determining whether a topic belongs to a certain category can be subjective. For instance, some participants consider a topic related to the trial of Conrad Murray (the physician for Michael Jackson) belonging to the "Sports and Entertainment" category since it's related to the pop singer. Other participants may consider this being a stretch since Michael Jackson is not the main subject of the news articles related to the topic.

For the same reason, we did not grade the accuracy of the identified topics, since arguments could be made for topics to be included or excluded from a news category. Although we did not grade accuracy of the identified topics, most of the identified topics for each news category did overlap a great deal among all the participants.
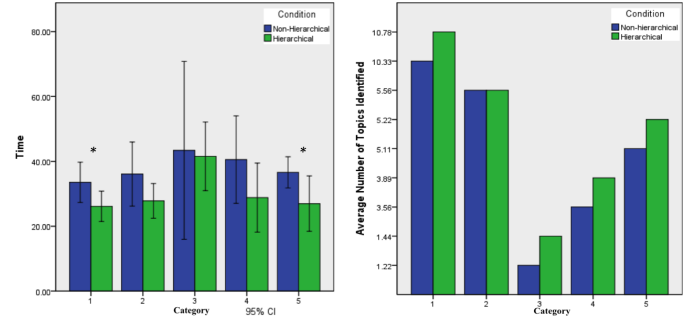
Fig. 10. Left: Average time to identify all topics for each news category during task1. Asterisk denotes significant difference. Right: Average number of topics identified for each news category.

5.3.1 Speed: hierarchical topic vs. non-hierarchical topic organization

To measure whether the hierarchical topic organizations yield faster speed for identifying topics for each news category, we performed one-way ANOVA on each category. A significant effect was found for two categories: American Politics and Middle-East News. For the American Politics category, a significant effect of hierarchical topic organization on the time for identifying relevant topics (Task 1) was found at the $p<.05$ level for the two conditions [$F(1,16) = 4.84, p = .043$]. For the same category, a significant effect was also found between two conditions [$F(1,16) = 4.79, p = .044$] in task 2, which involves grouping the identified topics and observing the temporal trends. For the Middle-East News category, the ANOVA revealed a significance between two conditions [$F(1,16) = 5.15, p = .37$]. No significance was found for the other three categories. Detailed results are shown in Figure 10 (left).

Combining with the average number of topics found in each category shown in Figure 10 (right), the results became more informative. Significant differences were found for categories with relatively large number of topics. In other words, the hierarchical topic structure lead to faster identification and grouping of large number of relevant topics.

5.3.2 Response on potential scalability of the system

The last question on the answer sheeting was regarding the potential scalability of the system. In particular, the question asked the participants to comment on if the HierarchicalTopics could scale to hundreds of topics. We tallied the participants' response and plotted them in Figure 11. The x axis shows the type of answers, including "*Yes, Maybe, and No*". The y axis illustrates the count of each answer.

None of the participants assigned to the non-hierarchical topic condition thought the system could scale to hundreds of topics, while the participants answering "*Maybe*" under the same condition further commented that some sort of automated classification such as topic groups could make the system much more scalable. The participants assigned to the hierarchical topic condition provided more positive responses toward the potential scalability of our system. Several of constructive comments were generated based on user feedback, details of which will be described in the discussion session.

In summary, the experiment results reveal that hierarchical topic structure leads to more efficient identification and grouping of larger numbers of relevant topics. After performing two tasks through interacting with the visual interface, most participants consider the hierarchical system scalable and think it can potentially handle hundreds of topics.
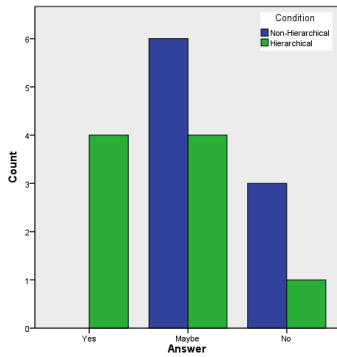
Fig. 11. Answer to potential scalability of hierarchical (green bars) vs. non-hierarchical (blue bars) topic representation.

6 DISCUSSION

In this section, we discuss possible improvements on the topic rose tree algorithm and the visual interface.

6.1 Limitation and future improvements on Hierachical-Topics system

During the user study, the participants provided constructive comments for improving HierarchicalTopics. A few users mentioned the need for annotation feature, which would allow them to annotate or bookmark a general topic group. In addition, users would also like to search for a particular word in the topic view, for the purpose of discovering all topics containing a word of interest. As mentioned in Section 3.3.1, we have actually incorporated both the annotation feature and the search function into the current system based upon this feedback.

Another interesting comment was on possibly taking advantage of spatial organization of the topics. One participant would like to organize the topics into interested vs. not interested piles and place them on different parts of the screen. Spatial organization is commonly used when working with real objects, and has been shown to aid more complex sense-making processes [1]. Thus more flexible user interactions need to be supported for users to accomplish such task in an un-laborious manner.

During the study, a few participants raised the question of what if one topic falls into two or more topic groups. For example, the topic of human robot interaction could be categorized into both HCI related topic group and Robotics related group. Therefore, we are planning to provide additional user interactions that allow users to duplicate topics and keep track of the duplicates.

Lastly, one limitation arose from the use of tree visualization to represent the hierarchical topic structure. The concern is that tree visualizations may not scale to displaying very large number of topics or multi-level hierarchies. Our HierarchicalTopics system alleviates this issue by supporting rich user interactions, including collapsing, annotating, and deleting the nodes in the rose tree. Nonetheless, we acknowledge the potential limits of this tree representation and will further explore other visual metaphors.

6.2 Future improvement on the Topic Rose Tree

As for the Topic Rose Tree algorithm, a few improvements could be added to make the algorithm more transparent and interactive to end-users. For example, when merging two subtrees in each computational step, selecting different operations would yield different results not only in terms of topic groups, but also in regard to the depth of the tree. Theoretically, both the absorb and collapse operations would lead to a rose tree with smaller depth compared to the join operation. Trees with less depth may make more sense for grouping topics, since the topics were assumed to be equally descriptive in the topic models. In the hLDA [3], topics on a higher level are usually less meaningful, comprised of mainly stopwords. Thus, it makes sense to control the tree depth to be as small as possible. A simple way to influence

the depth of Topic Rose Tree is to encourage either the absorb or collapse operation rather than the join operation. New interactions could, therefore, be designed to allow users to tweak the weight when calculating the cost of each operation. Such interactions could potentially support advanced users in influencing the topic hierarchy generation. This will certainly be one of the future directions for our visual text analytics research.

7 CONCLUSION

In this paper, we present HierarchicalTopics, a visual analytics approach to support the analysis of text corpora based on large number of topics. HT is designed to address the three challenges faced when analyzing large text corpora through topic based methods. HierarchicalTopics not only provides initial hierarchical structure of topics to facilitate exploration and navigation, it further allows users to modify topic hierarchies based on users' interest through intuitive interactions. In addition, the ThemeRiver in HierarchicalTopics is tailored to represent temporal trends in a hierarchical fashion. It enables the analysis and comparison of groups of topics as opposed to viewing the evolution of one topic at a time. Through both case study and user experiments, we have demonstrated the efficacy of HierarchicalTopics in helping users identifying topics groups, as well as interesting temporal patterns.

We have further produced a video that illustrate our Hierarchical-Topics system in action. The video can be found at `http://youtu.be/Vi1FP5kAbOU`.

REFERENCES

[1] C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 55–64, New York, NY, USA, 2010. ACM.

[2] D. M. Blei. Probabilistic topic models. *Communication of the ACM*, 55(4):77–84, 2012.

[3] D. M. Blei, T. Gri, M. Jordan, and J. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. *Neural Information Processing Systems(NIPS)*, 2003.

[4] D. M. Blei and J. D. Lafferty. Correlated topic models. *Neural Information Processing Systems*, 2006.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.

[6] C. Blundell, Y. W. Teh, and K. A. Heller. Discovering nonbinary hierarchical structures with bayesian rose trees. *Mixtures: Estimation and Applications*, April 2011.

[7] J. Boyd-Graber, J. Chang, S. Gerrish, C. Wang, and D. Blei. Reading Tea Leaves: How Humans Interpret Topic Models. In *Neural Information Processing Systems (NIPS)*, 2009.

[8] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *IEEE VAST*, pages 143–152, 2012.

[9] J. Chuang, C. D. Manning, and J. Heer. Termite: Visualization techniques for assessing textual topic models. In *Advanced Visual Interfaces*, 2012.

[10] J. Chuang, D. Ramage, C. D. Manning, and J. Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *ACM Human Factors in Computing Systems (CHI)*, 2012.

[11] CNN. Library of congress digs into 170 billion tweets. http://bit.ly/Uwqi7X.

[12] Committee on National Statistics. Science of science and innovation policy principal investigators' workshop. http://bit.ly/10o3via, Sep 2012.

[13] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2412–2421, 2011.

[14] W. Dou, X. Wang, R. Chang, and W. Ribarsky. Paralleltopics: A probabilistic approach to exploring document collections. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 231–240, 2011.

[15] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. Zhou. Leadline: Interactive visual analysis of text data through event identification and explo-

ration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 93–102, 2012.

[16] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: visualizing thematic changes in large document collections. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):9–20, 2002.

[17] A. Jinha. Article 50 million: An estimate of the number of scholarly articles in existence. *Learned Publishing*, 23(3):258–263, 2010.

[18] H. Lee, J. Kihm, J. Choo, J. Stasko, and H. Park. ivisclustering: An interactive visual document clustering via topic modeling. *Comp. Graph. Forum*, 31(3pt3):1155–1164, June 2012.

[19] Medialab Tools. i want hue web color chooser. http://tools.medialab.sciences-po.fr/iwanthue/, March 2013.

[20] J. Paisley, C. Wang, and D. M. Blei. The discrete infinite logistic normal distribution for mixed-membership modeling. *Bayesian Analysis*, 7(4):997–1034, 2012.

[21] D. Ramage, S. Dumais, and D. Liebling. Characterizing microblogs with topic models. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. AAAI, 2010.

[22] D. Ramage, C. D. Manning, and S. Dumais. Partially labeled topic models for interpretable text mining. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 457–465, New York, NY, USA, 2011. ACM.

[23] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, UAI '04, pages 487–494, Arlington, Virginia, United States, 2004. AUAI Press.

[24] D. Shahaf, C. Guestrin, and E. Horvitz. Metro maps of science. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 1122–1130, New York, NY, USA, 2012. ACM.

[25] L. Shi, F. Wei, S. Liu, L. Tan, X. Lian, and M. Zhou. Understanding text corpora with multiple facets. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 99–106, 2010.

[26] R. M. Shiffrin and K. Börner. Mapping knowledge domains. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5183–5185, 2004.

[27] Statisticbrain.com. Facebook statistics. http://bit.ly/YaAVmg.

[28] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101, 2004.

[29] The National Science Board. *Science and Engineering Indicators 2010, Chapter 5, Page 29*. National Science Foundation, 2010.

[30] The Unofficial Twitter Resource. Twitter now seeing 400 million tweets per day, increased mobile ad revenue, says ceo. http://bit.ly/JP9DXA, Feb 2013.

[31] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 448–456, New York, NY, USA, 2011. ACM.

[32] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. Tiara: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 153–162, New York, NY, USA, 2010. ACM.

[33] ZD Net. Engaging citizens the right way: Government uses twitter during hurricane irene. http://zd.net/mS0aOU, Sep 2011.

LeadLine: Interactive Visual Analysis of Text Data through Event Identification and Exploration
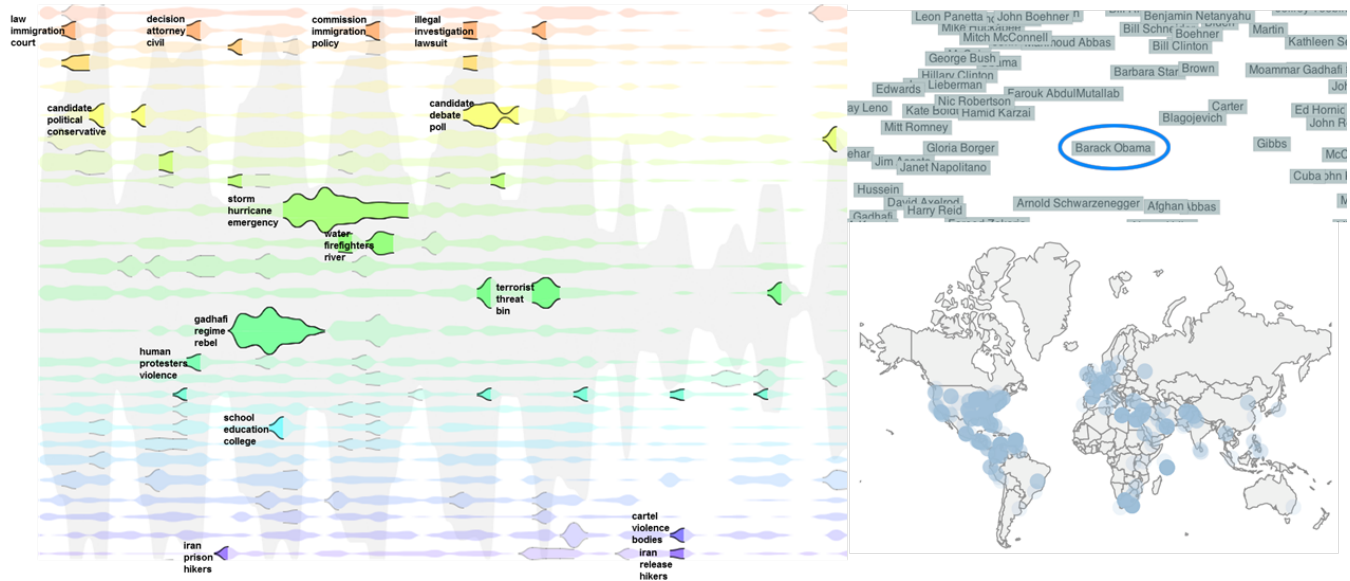
Figure 1: Overview of Leadline. Top right: people and entities related to President Obama are shown in the graph. Bottom right: locations mentioned in news articles related to the president. Left view: bursts are highlighted to indicate events related to President Obama.

ABSTRACT

Text data such as online news and microblogs bear valuable insights regarding important events and responses to such events. Events are inherently temporal, evolving over time. Existing visual text analysis systems have provided temporal views of changes based on topical themes extracted from text data. But few have associated topical themes with events that cause the changes. In this paper, we propose an interactive visual analytics system, LeadLine, to automatically identify meaningful events in news and social media data and support exploration of the events. To characterize events, LeadLine integrates topic modeling, event detection, and named entity recognition techniques to automatically extract information regarding the investigative 4 Ws: who, what, when, and where for each event. To further support analysis of the text corpora through events, LeadLine allows users to interactively examine meaningful events using the 4 Ws to develop an understanding of how and why. Through representing large-scale text corpora in the form of meaningful events, LeadLine provides a concise summary of the corpora. LeadLine also supports the construction of simple narratives through the exploration of events. To demonstrate the efficacy of LeadLine in identifying events and supporting exploration, two case studies were conducted using both news and social media data

1 INTRODUCTION

Text data such as online news stories and microblog messages contain rich real-time information about worldwide events and social phenomena. In particular, news stories report ongoing development of events; microblogs capture people's comments and reactions to these events especially from a social aspect. Overarching patterns are lost in the here and now of the constant feed. Valuable information regarding major social and news events is hidden by the details. Therefore, methods to distill text data into not only meaningful overarching topics, but more importantly into triggering events are of great help to assemble the details into summarized information.

While summarizing large text corpora based on topical themes has received much attention, few have approached the problem from an event-driven perspective. Much work in the visualization community has been devoted to summarizing text data through representing topic evolution over time. Similar to these visual text analysis systems with a temporal focus such as ThemeRiver [20], TIARA [35, 33] and ParallelTopics [16], our approach organizes text corpora based on meaningful overarching topics. Yet unlike these tools, our focus is not visually presenting topical trends over time, but rather revealing indicators of events that trigger the major changes in temporal trends.

1.1 Formulating Events

Several questions are critical to identifying events from text corpora. How does one identify meaningful events given a text collection? What are the attributes that characterize an event? How does one automatically and systematically discover attributes that char-

acterize an event from text collections? To address these questions, we first determine what comprises an event:.

Merriam-Webster provides a general definition of an event as "a noteworthy happening and a social occasion or activity" [17]. In the Topic Detection and Tracking (TDT) community and other new event detection research [9], an event is defined based on its attributes as "something that has a specific topic, time, and location associated with it". From a storytelling standpoint, McKee refers to a story event as something that "creates meaningful change in the life situation of a character" [27].

Combining these definitions with our perspective on analyzing text corpora, we define an event as:

"*An occurrence causing **change** in the volume of text data that discusses the **associated** topic at a specific time. This occurrence is characterized by **topic** and **time**, and often associated with entities such as **people** and **location**.*"

For simplicity, we refer to < ***Topic, Time, People, Location*** > as four attributes of an event. These four attributes address common questions in investigative analysis: *When* did an event start and end? *What* is the event about? *Who* was involved? And finally *where* did the event occur?

1.2   Introducing LeadLine

Given our notion of events, we identify techniques and models that allow us to use computational methods to automatically extract events from text corpora. To discover events, we extract information regarding who, what, when and where through integrating topic modeling, event detection, and named entity recognition techniques. More specifically, we first organize text data such as news stories and microblog messages based on topical themes using Latent Dirichlet Allocation (LDA) [12]. This step provides topical information for events. To identify the temporal scale for events, we applied an Early Event Detection algorithm to automatically determine the length and "bursty-ness" of events. This step provides a beginning and an end to each event in time. To discover people and location associated with each event, we first perform named entity recognition on the text corpora and associate the entities with events. With all four attributes explicitly modeled, our approach supports identification and exploration of events at the topical, temporal, and entity level.

To effectively communicate the event identification results, we developed a visual interface. The interface allows users to interactively explore events and, more importantly, to steer the event identification process to adjust the granularity of the detected events. In addition, organizing text corpora based on events provides basics for building narratives. This is a constructive format that describes a sequence of events [17] [27]. We extended LeadLine with the capability to examine narratives, which allows users to easily access and revisit their explorative findings.

Our approach provides an event-driven summarization of text data with exploratory capability for investigating the events based on who, what, when, and where. Specifically, our approach presents three contributions:

- A general process that couples topic modeling, named entity recognition, and early event detection techniques to identify meaningful events from text corpora.

- An interactive visual interface for exploring events based on the aspects of who, what, when, where, and further allowing interactive adjustment of event granularities.

- A narrative examination interface that allows users to report and revisit their findings.

2   RELATED WORK

Three areas of research, namely event detection, topic analysis, and text visualization techniques, are the main inspiration for the design of LeadLine. Another thread of research on event structure in cognition provides background for using events as a summary.

2.1   Event Structure in Perception and Cognition

As Zacks and Tversky noted, the world presents nothing but continuity and flux, yet our mind has a gift to perceive activity as consisting of discrete events that have some orderly relations [36]. An event is a segment of time at a given location that is perceived by an observer to have a beginning and an end. People make sense of continuous streams of observed behavior in part by segmenting them into events.

People seem to segment observed physical activities into events effortlessly and simultaneously at multiple timescales [23]. However, little research indicates that the same skill applies to abstract continuous streams, such as topical streams derived from text corpora. Since an event is considered the unit for making sense of continuous activities, we argue that accurately identified and conveyed events may serve as a more natural representation for making sense of the activities.

2.2   Event Detection

The Biosurveillance community has long been investigating ways to improve clinical preparedness for bioterrorism [21]. Early surveillance tasks required continuous monitoring of massive quantities of multivariate data in order to identify emerging patterns [**?**]. Considering a particular type of health data, such as over-the-counter (OTC) medication sales, as a source for detecting events indicating disease outbreaks, Goldenberg et al. described a statistical system designed for timely detection of anthrax epidemics. This approach falls into the category of univariate methods which focus on detecting events from time series [19].

Other disease surveillance systems take into account both temporal and spatial information. The system described by Sabhnani et al. [31] monitors daily data feeds from over 20,000 hospitals and pharmacies nationwide, including emergency department visits and OTC drug sales, to identify early events. Specifically, an expectation-based scan statistic approach is proposed to search for space-time regions for disease outbreaks. As an extension, Neil at al. further developed a "multivariate Baysian scan statistic" (MBSS) [29] method for faster and more accurate event detection.

As efficient as the proposed event detection algorithms are, they lack the ability to handle text corpora, which may contain rich information about when the symptoms emerge and how they evolve over time. In this paper, our formulation allows us to transform textual data into multiple meaningful time series so that we can apply ideas from the Biosurveillance community for early event detection on text corpora.

2.3   Event Extraction from News Sources

Topic Detection and Tracking (TDT) is a body of research and an evaluation paradigm that addresses event-based organization of broadcast news [9].The motivation for research in TDT is to address analysts' need to monitor new events from the high volume of broadcast news. The TDT has conducted several full-scale evaluations with research groups around the world by creating news corpora with the ground truth of events and topics. In his book that summarizes the TDT research program from 1997 to 2002, Allan James noted that much of the research on TDT has been tuning parameters to make certain information retrieval approaches more appropriate for events [9]. However, since an event is considered as something that happens at a particular time and location, James pointed out that existing approaches lack explicit modeling of temporal and geospatial dimension of the news streams. Our approach
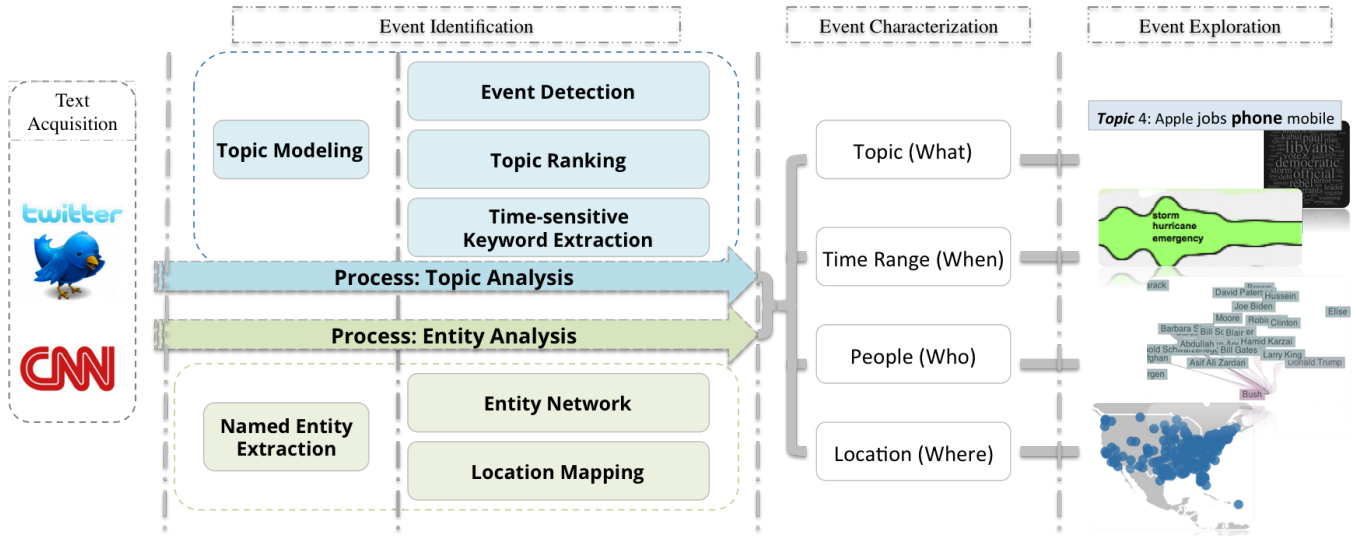
Figure 2: System Architecture of LeadLine.

not only takes into account temporal aspects of events, but also explicitly extracts named entities including people and location, and further makes use of these entities for inferring relationship between events and exploring events.

2.4 Visualization of Events and Temporal Topical Trends

Multiple visualizations have been developed to visualize events from news sources. For example, in the time-series visualization CloudLines, Krstajic et al. proposed an incremental visualization technique that allows detection of visual clusters in a compressed view of multiple time series [22]. Luo et al. described a visual analytics approach which presents events in a river-like metaphor based on event-based text analysis [24]. To explore events reflected in microblogs, Marcus et al. presented TwitInfo [26], which is a system for visualizing and summarizing events on Twitter. As opposed to a summarization approach, TwitInfo requires users to specify a Twitter keyword query to start exploring events related to the query. As opposed to the above systems, our approach provides investigative information regarding not only what (the events are about) and when, but also who is involved and where the events possibly happened.

More recently, Cui et al. [15] presented TextFlow for analyzing topical evolution patterns. In particular, TextFlow identifies critical events as topic birth/disappearance and topic merging/splitting. Our approach differs from their perspective of annotating topic evolution with critical events by centering on the identification of events, with topics as one aspect to characterize events.

Since topic and time are two essential attributes that characterize an event, previous work on providing temporal summaries of topics provided inspiration for our approach. There have been multiple examples of presenting topics along the time dimension. Wei et al. [35] introduced TIARA, a time-based interactive visualization system that presents topical themes in a time-sensitive manner. Similary in ParallelTopics [16], described by Dou et al., temporal evolvement of topical themes are also presented in a ThemeRiver visualization [20]. In our approach, we use a topic-based summarization method (topic models) to organize events over time. In addition, we allow users to interactively explore events based on who and where, as well as generate a narrative as a results of the exploration.

3 ANALYTICS ARCHITECTURE FOR EVENTS CHARACTERIZATION

To extract information regarding events from text corpora, we integrate several techniques to identify < **Topic, Time, People, Location** > for each event. To extract meaningful topics and timespan for events, we leverage topic models for topical themes and an Early Event Detection method to identify a start and an end for each event (section). To extract information regarding who (people) and where (location), we perform named entity recognition and further analyze relationships between extracted entities (Section 6).

As shown in Figure 2, we categorize the identification of topical themes and timespan as topic-based analytics, in which we first extract meaningful topics from the input textual collection using Latent Dirichlet Allocation (LDA). We then apply a 1) topical-level event detection algorithm to automatically identify "bursts" as indicators of events labeled by a timespan; 2) topic ranking to facilitate the discovery of event associations by placing bursts with similar topics nearby; and finally 3) time-sensitive keyword extraction that provides information regarding an event with a set of succinct keywords.

Complementing the topic-based analytics, our architecture also focuses on entity-based analytics by applying named entity recognition technique to identify people and location related with each event. Specifically, this process is centered on extracting key entities from the text data regarding who and where. More importantly, this process also reveals the relationship between the entities, leading to further characterization and associations of events.

As illustrated in Figure 2, visual representations are designed and tailored to both stages of topic-based and entity-based analytics processes. The interactive visual interface is an integral part of both analytics processes to bridge users and the complex analytics results. With the interactive visualization interface, LeadLine supports interactive exploration of events from various aspects (who, what, when, where), as well as allows users to interact with the underlying analytics algorithms to partially steer the process of detecting events from text streams.

In the following sections, we will detail the algorithms and processes used in both topic and entity analysis, alongside the visual designs that are tailored for identification and analysis of the events.

4 DATA ACQUISITION AND PREPARATION

To demonstrate the generalizability of our algorithms and their applicable domains, we have applied our analytics architecture to two types of text data: CNN news and microblogs from Twitter. While both text sources contain rich information reflecting major real-world events, the primary reason for selecting these two data sources is because of their different editorial styles and the time window for responding to an event. In particular, content from news media (CNN and others) are edited by professional journalists, usually centered on a specific topic with some background. Individual tweets, on the contrary, contain mostly the unedited commentaries without much contextual information [10]. These different text sources provide various benchmarks to help us evaluate our analytic architecture.

While both sources are in the public domain, there are no datasets readily available or that can be shared due to privacy policies. Therefore, we have extended our existing scalable data architecture to acquire news and tweets using customized crawling strategies. The details of our data architecture can be reviewed in previous work [withhold].

News Data Acquisition: In particular, our current work extended on the previous architecture by adding the news article crawler. Both historical news as well as up-to-the-hour news are acquired by our customized webpage crawlers and our RSS daemons, respectively. Both methods employ universal programs that attempt to crawl an entire web domain, download all the webpages, extract all textual articles, parse article time information, and finally normalize articles by removing HTML formatting and noise (e.g. advertisement and external website links). More specifically, our news crawler is built with python, and Apache Nutch [30] implementation. The data is stored into our HBase data structure for fast access and MapReduce [2] based data cleaning and processing. Using these crawlers, we could retrieve and clean news articles dated back to 2004 ($\sim$102573 articles) within several hours of the crawling process.

Twitter Data Crawling: Microblogs from Twitter are also collected from dual crawling processes. The primary process utilizes our MapReduce parallel data crawler, which is interfaced with the Internet through multiple independent crawlers. Each of the crawlers constantly collects social media data from various public domains and dumps it into HBase. Specifically, we have created crawlers to connect to Twitter's public "Garden-hose" API to collect 10% of tweets, which is a statistically significant sample of all content from Twitter [7]. In addition, we have also experimented with an Online Social Network (OSN) graph-based crawling mechanism to extend our practice using Apache Nutch. Inspired by Catanese et. al [13], the concept for this crawling process is accustomed to the nature of OSN which can be represented as graphs, with nodes representing users and edges denoting connections. We perform a breadth-first search using Nutch to acquire Twitter public user-graphs and capture the tweets through their webpage portal for broader streams. As a result, we were able to collect over 5 billion tweets from all languages over the course of 3 months, providing a reliable database for evaluation purposes.

5 TOPIC-BASED EVENT ANALYSIS AND VISUALIZATION

Topic-based analytics is crucial for event characterization in terms of revealing **topics** and **time**. In this section, we introduce algorithms to extract topical and temporal information with regard to an event, as well as visual representations that communicate the topical and temporal aspects.

5.1 Extracting Topics from Text Data

We start by organizing text streams based on topics. Given a text corpus, there are multiple ways to extract meaningful topical themes. Among them, probabilistic topic models [11] are considered to be advantageous compared to traditional vector-based text processing techniques. In LeadLine, we first employ the most widely used topic model, LDA [12], to extract semantically meaningful topics from text corpora. LDA generates a set of latent topics, with each topic as a multinomial distribution over keywords, and assumes each document can be described as a probabilistic mixture of these topics [12].

5.1.1 Topic Streams

In addition to topics, another important attribute of an event is time. In order to highlight the temporal dimension, we organize the extracted topics along the temporal axis. Considering each topic as a stream that evolves over time, the calculation of each topic stream is performed by computing a spline based on the volume of textual information associated with the topic in each time frame. A time frame is a unit that the texts are temporally aggregated upon. The time frame unit varies by datasets and tasks, ranging from minutes for social media data to days for news stories.

5.1.2 Visualization of Topic Streams

To visually represent topical themes over time, we adopt a flow-like visual metaphor in which each stream (row) represents a topic and the height of each topic changes as the amount of textual information related to the topic varies (Figure 1 left) We represent each stream in a separate row for the later discovery of events within each topic stream.

However, with each topic stream represented separately, the visualization does not capture the overall trend of how all topics evolve over time. In order to provide the overall context, a ThemeRiver representation is placed in the background of the visualization. Therefore repetitive patterns exhibited by a text stream (such as a weekly pattern for news stories) are still portrayed.

5.2 Automatically Detecting Events in Topical Streams

Algorithm 1: CUSUM
Input: topic time series TT
Steps:

1. Calculate the reference value r = mean of the time series;

2. Calculate running sum S from the starting time frame
$S_1 = TT_1$ - r
$S_2 = (TT_2$ - r$) + (TT_1$ - r$)$
$S_k = \sum_{i=1}^{k} (TT_k$-r$)$.

3. When S_k exceeds a threshold H, event alarm and reset S, go back to step2.

A key technical contribution of this paper is automatically identifying temporal peaks in topics as indicators of events. To detect events from topic streams, we consider each stream as a time series. Each time series is computed by aggregating texts related to the topic within every timespan.

To automatically discover events from these topical time series, we apply an early event detection (EED) method to look for bursts in topic streams. More specifically, we adopt the cumulative sum control chart (CUSUM) widely used for change detection [28]. CUSUM is good at detecting shifts from the mean in a time series by keeping a running sum of "surprises". We implemented our version of CUSUM for detecting changes in topical volume. For each topic stream, the algorithm keeps a cumulative sum of topic volume at each timespan that are higher than the mean topic volume. Once the cumulative sum exceeds a threshold, the event alarm triggers

and the running sum is reset to 0. Our implementation of CUSUM is shown in Algorithm 5.2.

The result is a set of automatically detected events within each of the topic streams, with each event labeled by a beginning and an end along the time dimension. Our approach of detecting events on time series is generalizable. It could be used as an extension on any time-series visualization to highlight changes as indicators of events.

5.2.1 Visualizing Detected Events

To visually represent the detected events, we use a combination of contours and highlights to indicate events within each topical stream. As shown in Figure 1 left, we provide an overview of events by drawing contours in topical streams to signal events. The time span of the contours is determined by the event detection results.

In addition to using contours to provide an overview of all events, LeadLine supports highlighting events of interest through user interaction. For example, a user could interactively highlight events related to a specific named entity such as a person or location (the extraction of entities is introduced in Section 6). To provide details on demand, LeadLine allows users to access documents (news or microblog messages) that discuss a highlighted event by clicking on the event.

5.2.2 Interactive Event Detection
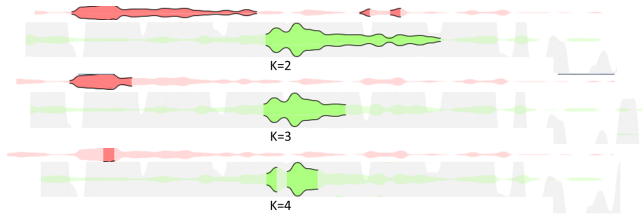
Figure 3: Comparison of different granularities of events. K indicates detecting changes K*standard deviation above the mean.

Although one advantage of our event detection algorithm is to automatically discover events that trigger topical bursts, we still want to provide end users the ability to specify how big a change warrants an event. By adjusting the scale of the "change", users will be able to view either finer or coarser-grained events.

In step 3 of Algorithm 1 , the event alarm triggers when the difference between the mean and running sum S is greater than a threshold. The threshold is usually measured by a fixed number of standard deviations (K) [6]. In LeadLine, we allow users to interactively adjust K between one and four standard deviations above the mean of one topic stream. When K is smaller, the algorithm is able to detect smaller shifts from the mean, leading to more events that cause less drastic change to be discovered. On the other end, when K is larger, the algorithm is only able to detect big shifts from the mean so only larger "bursts" are detected. Figure 3 shows three event detection results with varying thresholds. Once a user adjusts the K value using the slider in the interface, LeadLine re-runs the event detection algorithm to produce new event results. Through allowing users to interactively manipulate the fineness of events, LeadLine provides overviews with more or less details for given text corpora.

5.3 Topic Ranking

When placing the topics in our visualization, we want similar topics to be close so that the events later derived based on topics are also posited in proximity. Since LDA does not explicitly model the

relationship between topics, we rank the topics based on their similarity measured by Hellinger distance. Specifically, Hellinger distance measures the distance between two probability distributions of topics over the entire vocabulary in the text streams:

$$topic - similarity_{i,j} = \sum_{v=1}^{N} (\sqrt{\beta_{i,v}} - \sqrt{\beta_{j,v}})^2 \qquad (1)$$

β is the probability of the ith topic over term v in the vocabulary, and N denotes the vocabulary size.

5.3.1 Visualization of Topical Content
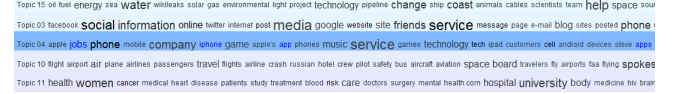
Figure 4: Topic Cloud with the topic related to mobile device/technology highlighted. Keywords in blue are the time-sensitive keywords in a certain time span. The topics are extracted from CNN news corpora (Aug 15 - Nov 5, 2011).

To represent topics in the form of a set of keywords, we develop the Topic Cloud view based on tagCloud representation (Figure 4). In the Topic Cloud, the size of each keyword reflects its number of occurrences in all topics. Since topics are ranked based on their similarities, the Topic Cloud view starts with a random topic at the top and places the topic with the highest similarity score next to the one above. As a result, similar topics are placed next to each other visually. To color the topics in Topic Cloud, we choose colors from the HSV (hue, saturation, value) space by dividing the space based on similarities between the list of topics. As a result, similar colors are assigned to topics with greater similarity. The same color scheme is also applied to the Topic Stream View (Figure 1 Left).

5.4 Time-sensitive Keyword extraction

Algorithm 2: extract time-sensitive terms
Input: topic-term distribution matrix ϕ; desired number of keywords per time frame N
Steps:
  1.
for each topic i **do**
    for each time frame t **do**
Identify a collection of documents $D_{i,t}$ focusing on topic i from entire text stream;
    end for
end for
  2.
for each term W in topic i from $D_{i,t}$ **do**
calculate term frequencies TF
end for
3. Re-rank the TF scores with topic-term probabilities: for each term W_m, weight(W_m) $= \lambda_1 \frac{TF_{i,t,m}}{\sum_t TF_{i,t,m}} + \lambda_2 \phi_{i,m} \log \frac{\phi_{i,m}}{(\prod_{k=1}^{K} \phi_{k,m})^{1/k}}$
4. Within each topic and time frame, select the top N terms as time-sensitive terms.

To accurately summarize what an individual event is about, we need to re-rank the keywords within a topic by considering which keywords are more prominent given the timespan of the event. In order to do so, within each topic, we first extract the most prominent keywords for each time frame. We adopted the keyword-ranking algorithm that Fu et al. proposed [35] to re-rank terms in each topic based on time factor. More specifically, we followed Algorithm 2

to extract time-sensitive terms given a topic-term distribution matrix, and the number of desired terms for each time frame N. The input for this algorithm is a text corpus divided into sub-collections using time frame and topics. Each sub-collection contains only the documents focusing on the specific topic and also published during the time frame. The algorithm follows a TFIDF heuristic to determine time-sensitive terms: (a) if a term occurs frequently in the sub-collection, it is important; (b) if the term also occurs in many other sub-collections, the importance is discounted.

The extracted time-sensitive terms could not only better aid the exploration of the text collections along the temporal axis, but also provide accurate labels for discovered events. The association of events with corresponding time-sensitive keywords is accomplished through aggregation of all time-sensitive keywords within a topical burst identified by our event detection algorithm.

5.4.1 Interaction and View Coordination

The Topic Cloud view and Topic Streams are coordinated through user interaction. Hovering the mouse over one topic stream in the Topic Streams would highlight the corresponding topic in the Topic Cloud. Hovering the mouse over a specific timespan within a topic stream would highlight the time-sensitive keywords. In addition, events from different topics can be highlighted by selecting named entities. Details of extracting named entities from events are introduced in section 6. In addition to view coordination, one important interaction LeadLine provides is choosing the "fineness" of the detected events. By simply moving the slider, a user can determine the scale of the examination.

In summary, section 5 presents algorithms that extract topical and time information that characterizes events. Visual representations of events are also designed based on the algorithmic results. The visualization not only allows interactive examination of the detected events, but also provides users the power to steer the event detection algorithm

6 ENTITY-BASED EVENT ANALYSIS AND VISUALIZATION

Complementing the topics and time attributes that are depicted in topic analysis, entity analysis is yet another crucial component. The named entities further reveal relationships between events by connecting them with information regarding people and location.

The primary goal for our name entity extraction is to extract people and locations from the textual streams and associate these entities with events. As illustrated in Figure 2, our named entity extraction identifies the named-entities once the data has been cleaned and prepared. The current entity extractor uses the LingPipe package [1] with Statistical Chunking and customized Dictionary-based Extraction. The three categories extracted are people, location, and organizations. This information is then inserted into the HBase storage platform, and attached to each corresponding news article. Note that location entities have also been enriched with geo-tags that are acquired alongside the contents.

While our current approach identified a sufficient amount of meaningful entities for news articles, the initial results from this approach on microblogs was fairly noisy. Similar findings have been identified in research in the AAAI community on the same issue, even with more state-of-the-art part of speech (POS) tagging [18]. We believe improvement on entity-based analytics for tweets is much needed, and do appreciate other representative research by MacEachren et al. on extracting entities from tweets [25]. Given the focus of this paper, we will focus on entities in news and demonstrate the utility of associating entities and their relationships in characterizing events and constructing narratives

6.1 Entity Graph and Geo-Mapping

Entities are primarily used in LeadLine to connect multiple events into a meaningful narrative structure. Similar to the term used in

Jigsaw [34], our distributional similarity between entities is computed based on the commonality of their contexts of occurrence in text. In particular, the association between entities is determined based on the co-occurrence of entities within the corresponding text. For example, two entities that co-occur in at least two news articles are considered connected. This gives us a contextual background and a baseline to depict the interplay between entities. Similar to the interactive event detection process, LeadLine enables users to interactively adjust the granularity of entity correlations based on their co-occurrence.

As shown in Figure 1 top right, we have created an interactive entity graph visualization to represent the connection and correlation between entities. A basic force-directed graph is utilized to allow users to dynamically explore the graph by showing and hiding links and entities. Instead of visualizing all the entities within a large graph, LeadLine constructs the graph content based on user's selections of topics and time factor as well as the types of entity. Different types of entities are color-coded. For example, people are shown in blue while organizations are shown in green.

Furthermore, location information is geo-referenced onto an interactive map to reveal the spatial distribution of entities contained in the text corpora (see Figure 1 bottom right). Both the geospatial view and entity graph are coordinated in a highly interactive manner. User interaction within one view is used to filter the associated entity in all views. Through such view coordination, different aspects of the entities can be examined simultaneously under different perspectives.

In summary, LeadLine is designed to utilize results from named entity extraction to provide people and location information for exploring and inferring relationships between events. Interactive visualizations are further used to enable users to navigate through implicit connections between entities.

7 UNCOVERING NARRATIVES BY ASSOCIATING EVENTS

As a tightly coupled topic and entity analysis and interactive visual interface, LeadLine enables the user to perform investigative analysis of events inside a text corpus with a guided exploratory environment. While this environment is useful in interactively examining the four attributes < ***Topic, Time, People and Location*** > that comprise an event, identifying narratives that describe a series of events is more difficult.

Inspired by Segel and Heer's narrative genres [32], we have developed a Partitioned Poster style interface designed to assist in associating events to uncover narratives (Figure 5). The interface is implemented using Bostock's D3.js library [3] to increase accessibility and to allow further exploration of the data by remote users.

The hierarchy of information in the interface is: 1) Entity, 2) Event, 3) Topic. The entity list can be very long for a large text corpus, so an autocompleted text entry field is used for entity selection. After selecting an entity, the user is presented with aggregate word-cloud and map views and a timeline view showing the events. The wordcloud view displays the frequency of topic keywords. The map view provides geolocations mentioned in the events. The timeline view gives a scope of events' durations and overlap.

The three views are coordinated with the timeline view being the main point of interaction. Hovering over an event in the timeline view causes the map and wordcloud views to display only topics associated with that event rather than the aggregate for the entire entity. The Partitioned Poster format is not just good for telling stories, but also for discovering existing stories in a data set. It reinforces the connection between the components of the visualization without suggesting an ordering to the elements. This allows a user to get both an overview, and details about an event, and to uncover the connections and triggers between different events. These discoveries help to reveal narrative structures in the event timeline.
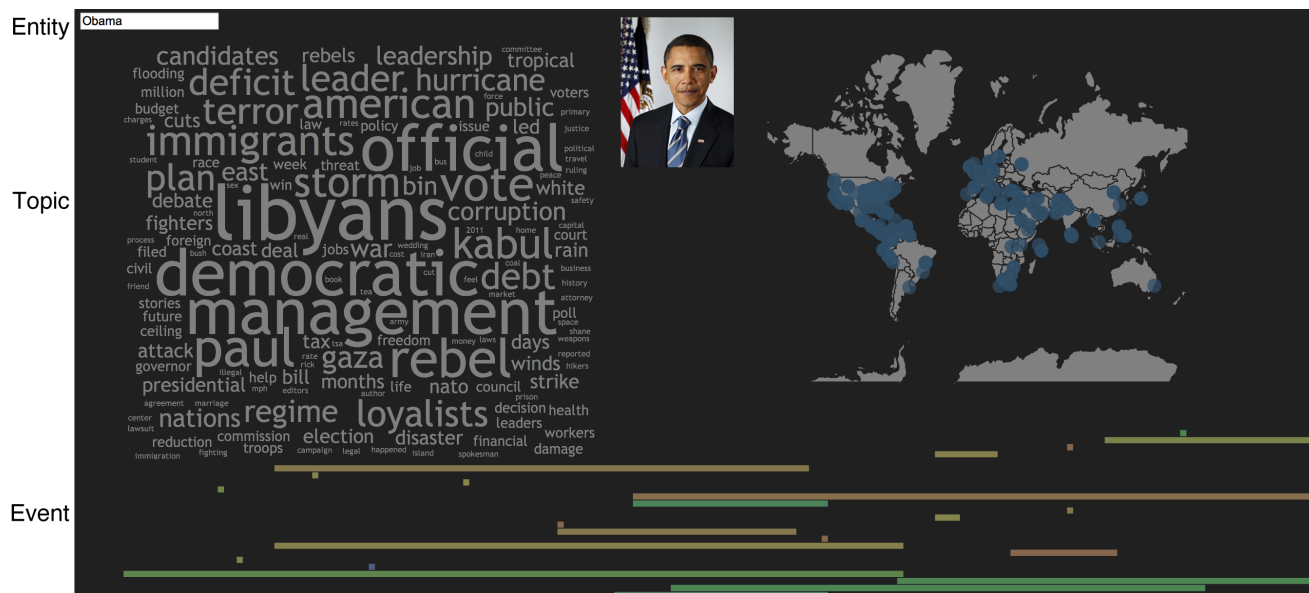
Figure 5: A narrative poster constructed from news events about President Obama.

8 CASE STUDIES

To evaluate the efficacy of LeadLine in automatically detecting events from text data and facilitating exploration of events based on topics, time and named entities, we conducted case studies using real-world textual data from CNN news [14] and Twitter [7]. For the case studies, we recruited 8 users to explore the events from the CNN and Twitter datasets. We provided training to our participants regarding the LeadLine system before the exploration. During the exploration, we asked the participants to think-aloud and the experimenter took notes of the findings. In the end, we conducted a post-study interview to collect feedback about the visualization.

8.1 Case Study 1: Exploring the Occupy Wall Street Movement

In this section, we first describe the dataset we use for the study. We also characterize challenges in analyzing such a large-scale social movement. We then present how LeadLine assists users in discovering major events and understanding the OWS movement through exploring these events from multiple aspects.

8.1.1 Data set and Background

The Occupy movement is an on-going series of demonstrations and is known for using social media for publicity and organization. The Occupy movement is long-lasting and wide-spread without central leadership. This creates challenges in understanding the direction of the movement. In addition, a wide range of goals were reported [5], including more jobs, more equal distribution of income, and bank reform. Given the prominent use of social media in the Occupy movement, it makes sense to analyze the voices from protestors and citizens.

To provide an overview of the OWS movement, we filtered our tweet collection described in Section 4 using the hashtag #occupy. The resulting dataset contains more than 100,000 tweets from Aug 19 to Nov 01. Given the length of the dataset, the time unit in the visualization is set to 6 hours, allowing users to see how an event evolves within a day. 15 topics were extracted using LDA.

8.1.2 Exploring the OWS Movement Through Major Events

Before the studies, we asked our participants if they were familiar with the OWS movement. All of them indicated that they had heard
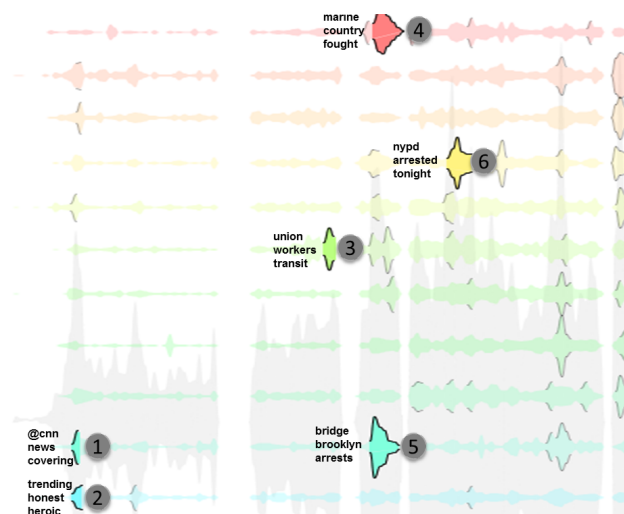
Figure 6: Major events discovered by the participants.

or read about the movement from news media, but none had a good understanding of the movement. In this task, we asked the participants to identify major events of the OWS movement and report what they have learned through exploring with LeadLine.

The participants usually started by quickly going through all topics in the Topic Cloud to gain an overview of the topics summarizing the tweets. Some participants then mainly focused on the Topic Stream View as they explored major events, while glancing at the Topic Cloud for information regarding a topic or a specific event. To make sense of what event triggered a certain topic burst, the participants clicked on the burst and accessed all tweets contributing to the burst. As they performed explorations in LeadLine, the experimenter took notes of their findings. For reporting purposes, we collected major findings from our participants, shown in Figure 6. Here we describe a few common discoveries from the participants:

Topic bursts #1 and #2 clearly mark the beginning of the Occupy movement on Sep 17, 2011, with tweets noting the coverage from

news media on the first day of OWS gathering (burst #1), and the trend of tweets mentioning the OWS movement picking up steam on Twitter (burst #2).

Topic bursts #3 and #4 mark the events of two forces joining the OWS movement to support the protestors. On Sep 28th, the board of the local union of the Transport Workers Union of America voted to support OWS, which splashed a lot of celebratory tweets (burst #3). Similarly, on Oct 1st, the burst of tweets (#4) was triggered by the marines joining the movement to protect protestors from the police.

Topic bursts #5 and #6 indicate marches by the occupy protesters. Topic burst #5 is triggered by protesters marching across the Brooklyn Bridge on Oct 1st. Some tweeted live updates of the march and of police arrests of the protestors, while others complained that there was not enough media coverage of the event. On Oct 5th, more than 10,000 protesters marched near Zuccotti Park. Tweets within initial time of the burst were about the march and the massive number of people involved. However, tweets in the later part of the burst were ranting about protesters being pepper sprayed, beaten and arrested by police. The event clearly triggered some outrage on Twitter (burst #6).

The events mentioned above are just a sample of all events explored and commented on by our participants. After the study, most of the participants mentioned that they now had a much better understanding of the OWS movement, including when the movement started, who was involved, and some major events within the large-scale movement.

8.1.3 Evaluating Detected Events

In addition to having participants identify and explore major events of the OWS movement using LeadLine, we compared our event results against timelines from Wikipedia [8] and the L.A. Times [4] regarding the same movement. Between Aug 19 and Nov 1, the Wikipedia timeline provided 24 events built by online communities, while The L.A. Times provided around 20 new articles describing the OWS movement. Depending on the level of details a user may want, LeadLine could provide events indictors ranging from 20 to more than 50.

Figure 7: Comparing events identified by LeadLine against timelines from Wikipedia and L.A. Times.
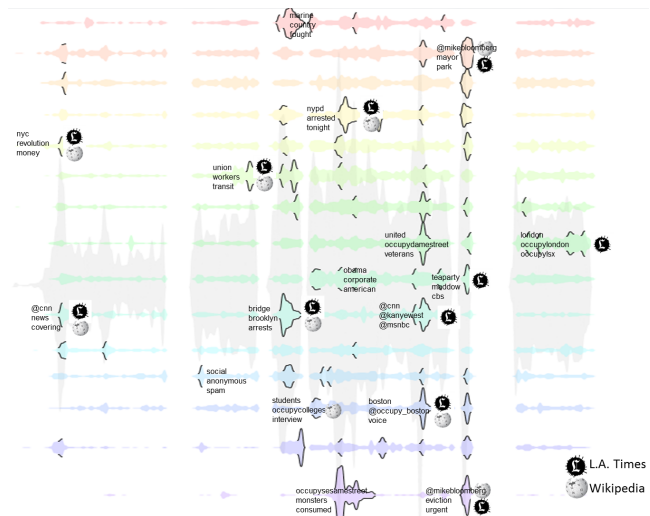
We also examined the event indicators to match them with events from Wikipedia and The L.A. Times. Figure 7 shows our matching results. The results indicate that most of the automatically detected events by LeadLine match the timelines from Wikipedia and

The L.A. Times, as annotated in Figure 7 with both logos after the majority detected bursts. Wikipedia also published events before September 17th that may have led to the OWS movement, which our automated detection did not identify due to insufficient tweets. However, LeadLine is able to show pre-September 17 activities that are worthy of further investigation. One interesting finding from our comparison is that The L.A. Times as a media organization reported more "photo-op worthy" events related to celebrities and politicians, while Wikipedia covered events centered around the OWS movement. As expected, LeadLine identified events that triggered more commentary on Twitter.

In summary, through interactively exploring the topical bursts LeadLine, the participants were able to identify interesting events and gain insights regarding the OWS movement. Through further comparing automatically detected events in LeadLine against timelines provided by Wikipedia and L.A. Times, we validated that the topical bursts accurately reflect true events.

8.2 Case Study 2: Constructing Narratives based on Events from News Stories

In the second study, we asked participants to first explore topical bursts and named entities related to events with LeadLine, and then focus on creating narratives based on their exploration.

8.2.1 Data preparation

In this study, the dataset was CNN news stories from Aug 15, 2011 to Nov 5, 2011. The news articles were harvested and organized using methods introduced in Section 4. 30 topics were modeled from a total number of 3,130 news articles. Named entities including people and location were extracted from the news corpus.

8.2.2 Uncovering Narratives based on Events

Figure 8: Events related to former Apple CEO Steve Jobs.

The exploration process of news events was similar to the first case study in that participants started by browsing all news topics to gain a quick overview of the corpus. Some participants then identified a topic of interest and examined the bursts indicating events within that topic. Others turned to the entity graph and the map to see the locations and people mentioned in the news corpus.

Here we report a sample of findings identified by our participants. Starting by exploring the entity graph, one participant was interested in gathering events related to former Apple CEO Steve Jobs. By clicking on the node of Steve Jobs in the entity graph, the participant filtered all events that mentioned Jobs' name in the Topic Stream View (Figure 8). Jobs was consistently mentioned in the events in the mobile-technology related topic (in blue), which made sense to the participant. He further browsed some of the events, such as Jobs' resignation from Apple in late August (blue burst in the middle), and the unveiling of iPhone 4S on the date

of Oct 4th, followed by the news of Steve Jobs' death (last burst in blue). The participant investigated further when he saw a topic burst in the health related topic in purple also highlighted by LeadLine to indicate relevance to Jobs. Upon inspection of the news articles (shown in Figure 8), the participant discovered that one CNN program was trying to piece together a detailed history Job's of health in light of his resignation from Apple." Now it all made sense to me, I just wouldn't have guessed it without seeing the events related to Steve Jobs in the visualization", the participant commented.

One of the most examined entities was President Obama. Participants discovered that he was a busy man during the three month period. Figure 1 shows topical bursts indicating events related to the president. He appeared in events such as immigration issues, political debates, natural disasters and international relations. After exploring the events in LeadLine and gaining insights about the news stories, multiple participants commented that they can now tell a story about the entities they focused on. They also expressed their desire to have a tool for them to gather and report their findings.

Inspired by the comments from the participants, we further developed a structure (described in Section 7) for building a narrative based on selected events. The narrative presents all events-related information regarding one person or location in a concise manner. The interface is web-based, so it is easily accessible from anywhere for reporting and revisiting.

9 DISCUSSION

In this section, we discuss the limitations of LeadLine, and outline future directions to extend the current research.

9.1 Limitations arise from the Use of Topic Modeling, Early Event Detection, and Named Entity Recognition

Our approach relies on automated algorithms to discover information regarding who, what, when, and where in order to characterize an event. Inevitably, the final results are affected by the performance of each algorithm. For instance, the interpretability of each "burst" in the topical streams depends on the topic modeling results. The accuracy of detected entities relies on the performance of the named entity recognition (NER) algorithm. In addition, the same NER algorithm performs differently on different datasets. Solving the issues in the short term is challenging, but we think it is useful to make users aware of these issues. In order to do so, we plan to borrow methods from uncertainty visualization to annotate different layers of uncertainty so that users can make more informed decisions during investigation and analysis.

9.2 Identifying Inter-topic Events

In the scope of this paper, we assume that each event is associated with one major topic. However, certain events may have an impact on multiple topics. In order to identify inter-topic events, entities and timespan can contribute to grouping bursts from different topics into one triggering event. In other words, if bursts in several topics occur at the same time and share the same set of entities including people and location, we can assume that they are triggered by the same event. If one event encompasses multiple topics, the number of topics may further be used to evaluate the impact of the events.

9.3 Future Improvements

There are several improvements we would like to pursue in the future. First, when visualizing events identified by our event detection algorithm, it is clear that the algorithm favors upward trends in the topic stream. Although the current results include the starting time for each event, we want to improve the event detection algorithm to include downward trends in the topic stream so that the event cycle is complete.

Second, an interesting challenge we face when cleaning the text data is the duplication issue in both the news and microblog data. For news corpus, the issue arises because news websites keeps the article content up-to-date by reusing the same URL with a different timestamp. For twitter, the issue lies in spam tweets and advertisements from different agencies. Our current implementation of addressing this duplication challenge is two-fold: first, we use the unique parameters for news and twitter contents (e.g. URL, posted timestamp, author, etc) to construct a unique identification (UUID). This unique identification is used as checksum in the first stage to filter duplicates with exact match. The second stage, our implementation utilizes the standard Longest Common Substring (LCS) algorithm to compare the content length between two articles that share similar contents. We believe this is of great use in reducing the skewing effects that are introduced by the duplicated in the text streams.

While this performs comparatively well on the news data, such two stage process suffers from a performance penalty due to the largely fragmented nature of tweets. This issue needs to be addressed since in our event characterization process, since both the named entity extraction algorithm as well as the topic modeling relies on word frequency calculations. Such duplication will undoubtedly skew the analysis results, producing inaccurate if not false event information.

10 CONCLUSION

In this paper, we present an interactive visual analytics system, LeadLine, that identifies meaningful events and allows users to examine the events that trigger changes in topical themes in news and social media. To discover events, LeadLine extracts information regarding who, what, when and where by integrating topic modeling, event detection, and named entity recognition methods. LeadLine also supports interactive exploration of events based on the 4Ws. Two case studies were conducted based on both news and social media data. The results indicate that LeadLine can not only accurately identify meaningful events given a text collection, but can also contribute to users' understanding of the events through interactive exploration.

REFERENCES

[1] Alias-i. 2008. lingpipe 4.1.0.
[2] Apache hadoop.
[3] D3.js api.
[4] La times.
[5] Occupy wallstreet report [online].
[6] Tutorial on event detection tutorial. In *KDD*.
[7] Twitter, inc.
[8] Wiki occupywallstreet.
[9] J. Allan, editor. *Topic detection and tracking: event-based information organization*. Kluwer Academic Publishers, Norwell, MA, USA, 2002.
[10] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi. Eddi: interactive topic-based browsing of social status streams. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 303–312, New York, NY, USA, 2010. ACM.
[11] D. Blei and J. Lafferty. *Text Mining: Theory and Applications*, chapter Topic Models. Taylor and Francis, 2009.
[12] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.
[13] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Crawling facebook for social network analysis purposes. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 52:1–52:8, New York, NY, USA, 2011. ACM.
[14] CNN. Cnn online. http://www.cnn.com/.
[15] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics

in text. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2412–2421, Dec. 2011.

[16] W. Dou, X. Wang, R. Chang, and W. Ribarsky. Paralleltopics: A probabilistic approach to exploring document collections. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 231 –240, oct. 2011.

[17] Event. Merriam Webster, http://en.wikipedia.org/wiki/Event.

[18] J. Foster, Ö. Çetinoglu, J. Wagner, J. L. Roux, S. Hogan, J. Nivre, D. Hogan, and J. van Genabith. hardtoparse: Pos tagging and parsing the twitterverse. In *Analyzing Microtext*, volume WS-11-05 of *AAAI Workshops*. AAAI, 2011.

[19] A. Goldenberg, G. Shmueli, R. A. Caruana, and S. E. Fienberg. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8):pp. 5237–5240, 2002.

[20] S. Havre, B. Hetzler, and L. Nowell. Themeriver: Visualizing theme changes over time. In *Proceedings of the IEEE Symposium on Information Vizualization 2000*, INFOVIS '00, pages 115–, Washington, DC, USA, 2000. IEEE Computer Society.

[21] S. A. Khan. Handbook of biosurveillance, m.m. wagner, a.w. moore, r.m. aryel (eds.). elsevier inc. isbn-13: 978-0-12-369378-5. *Journal of Biomedical Informatics*, 40(4):380–381, 2007.

[22] M. Krstajic, E. Bertini, and D. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2432 –2439, dec. 2011.

[23] C. A. Kurby and J. M. Zacks. Segmentation in the perception and memory of events. *Trends in cognitive sciences*, 12(2):72–79, Feb. 2008.

[24] D. Luo, J. Yang, M. Krstajic, W. Ribarsky, and D. Keim. Eventriver: Visually exploring text collections with temporal references. *Visualization and Computer Graphics, IEEE Transactions on*, 18(1):93 –105, jan. 2012.

[25] A. M. MacEachren, A. R. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford. Senseplace2: Geotwitter analytics support for situational awareness. In *IEEE VAST*, pages 181–190, 2011.

[26] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 227–236, New York, NY, USA, 2011. ACM.

[27] R. Mckee. *Story - Substance, Structure, Style, and the Principles of Screenwriting*. Methuen, 1999.

[28] D. C. Montgomery. *Statistical quality control*. Wiley Hoboken, N.J., 2009.

[29] D. Neill and G. Cooper. A multivariate bayesian scan statistic for early event detection and characterization. *Machine Learning*.

[30] A. Nutch. Apache nutch. http://nutch.apache.org.

[31] R. Sabhnani, D. Neill, and A. Moore. Detecting anomalous patterns in pharmacy retail data. *Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection*, Aug. 2005.

[32] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, Nov. 2010.

[33] L. Shi, F. Wei, S. Liu, L. Tan, X. Lian, and M. Zhou. Understanding text corpora with multiple facets. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 99 –106, 2010.

[34] J. Stasko, C. Gorg, Z. Liu, and K. Singhal. Jigsaw: Supporting investigative analysis through interactive visualization. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 131 –138, 30 2007-nov. 1 2007.

[35] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. Tiara: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 153–162, New York, NY, USA, 2010. ACM.

[36] J. M. Zacks and B. Tversky. Event structure in perception and conception. *Psychological Bulletin*, 127:3, 2001.

RiskVA: A Visual Analytics System for Consumer Credit Risks Analysis

Xiaoyu Wang, Dong Jeong, Remco Chang, Arun Pinto and William Ribarsky

ABSTRACT

Consumer credit risk analysis plays a significant role in stabilizing a bank's investments and in maximizing its profits. As a large financial institution, Bank of America relies on effective risk analysis to minimize the net credit loss resulting from its credit products (e.g. mortgage and credit card loans). Due to the size and complexity of the data involved in risk analysis, risk analysts are facing challenges in monitoring large amounts of data, comparing its geospatial and temporal patterns, and developing appropriate management strategies based on the correlation from multiple analysis perspectives. To address these challenges, we present RiskVA, an interactive visual analytics system that is tailored to support credit risk analysis. RiskVA provides risk analysts with interactive data exploration and information correlation, and visually assists them in depicting market fluctuations and temporal trends of the targeted credit product. When evaluated by analysts from Bank of America, RiskVA was appreciated for its effectiveness in performing in-depth risk analysis, and is considered useful in facilitating the bank's risk management operations.

KEYWORDS: Risk management, visual analytics.

INDEX TERMS: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 INTRODUCTION

Consumer credit risk analysis plays a significant role in stabilizing a bank's investments and in maximizing its profits. Credit risk management, in general, refers to the process in which the investors assess the risk of loss arising from a consumer who does not make payments as promised [3]. Credit risk typically occurs in investing and in the allocation of capital.

For most banks and financial institutions, loans and credit products are the largest and most obvious source of risk. In order for a bank to profit from a large consumer base, it must invest in credit products (e.g. credit cards, mortgages) that are reasonable to customers. However, the bank must strike a balance between the investments and the substantial amount of capital in its reserve; so that investments would be profitable yet sustain the bank's financial stability.

Therefore, the assessment of credit risk is, on the one hand, crucial for banks to position themselves to profit through balancing credit investments and returns; on the other hand, it is critical for the stability of an entire financial market. Inadequate risk management can result in severe consequences for companies as well as individuals.

As shown in their study of the correlations between recessions and banking crisis, Bloom et al. [4] suggested that the credit investment strategies were directly associated with the stability of the entire financial market; an unexpected credit crunch would lead to the complete disarray of the financial markets. For example, the loose credit risk management of financial firms was determined to be one of the factors that triggered the recession in 2008.

As a large institution, Bank of America (BOA) constantly faces the challenge of managing its credit risk. Their stability relies on the effective risk analyses to minimize the net credit loss resulted from its credit products, and to determine profitable market strategies. Essentially, BOA emphasizes the use of risk management to quantify the potential losses in an investment and to take the appropriate action given investment objectives and risk tolerance.

However, given BOA's wide range of credit investments, analyzing risk in such diversified portfolio has become an overwhelming process. This demands that analysts evaluate credit risks both temporally (i.e. identifying market turning points before and after the recession) and across credit markets (e.g. comparing product performances in major cities). Current analysis practices and analytical tools can't meet the challenges to comprehend the trends and patterns of markets from multiple perspectives.

Exacerbating this challenge is the increasing size and complexity of the collected credit data that each analyst needs to examine. This places an extra burden on each individual analysis process in terms of the efforts needed to acquire the most appropriate information. This is exacerbated further by the need to gather information from heterogeneous data sources and bring it into a common picture.

To help address these crucial financial challenges, we formed a research partnership with Bank of America to investigate novel analysis technologies. One of our first actions within this partnership was to observe and characterize the risk analysts' analytical workflow. This domain characterization granted us the opportunity to closely identify the key perspectives in risk analysis, and helped us gain insights on the design elements that are needed to facilitate such analysis.

In this paper, we present RiskVA, an interactive visual analytics system that is tailored to support credit risk analysis. RiskVA addresses the aforementioned challenges by supporting interactive data exploration and information correlation over a large corpus of credit data. It aids the domain analysts in depicting and comparing the performance of the credit products by visually revealing market fluctuations and temporal trends of the targeted credit products. To support individual analysis workflow, RiskVA further allows the analysts to choose different combinations of visualizations and to customize the visual interface based on their own preferences.

To evaluate the efficacy of our system, we conducted expert evaluations with risk analysts from BOA, and found that most analysts considered RiskVA to be useful and complimentary to their existing analysis needs. We further identified analysis scenarios for which our system could provide analysts with insights to develop appropriate risk management strategies.

Given these results, this work presents substantial qualitative advances over current practices in consumer risk analysis.

- It provides a visual exploratory environment to handle consumer risk data that scales to hundreds of thousands of credit data over any given length of time.
- It provides highly coordinated interactive visualizations to enhance both tactical and strategic risk analyses that are essential in identifying emerging risks.
- It provides customizable workspaces that support the individual analyst's analysis routines.

The remainder of this paper is structured as follows: Section 2 characterizes the analysis processes of risk analysts and describes our system's targeted users. Section 3 provides more detail about the limitations of current risk analysis practices. Section 4

presents our visual analytics system, RiskVA. Section 5 provides several scenarios in which our system can facilitate risk analysis. In Section 6, we present our evaluation with risk analysts and our discussion for advancing from the current stage. We conclude the paper in Section 7.

2 DOMAIN CHARACTERIZATION: CONSUMER CREDIT RISK ANALYSIS

In this section, we describe the nature of consumer credit and the current practice in conducing credit risk analysis. By discussing these in detail, we intend to shed light on the characteristic of these challenges and how an interactive visual analytics system can bring about a substantial improvement in risk management.

The importance of credit risk analysis, as part of financial risk analysis, comes from the New Basel Capital Accord (Basel II), published in 1999 and revised in 2004 by the Basel Committee on Banking Supervision (BCBS). In general, credit risk analysis is "*the process of establishing credit standards for investors and counterparties, assessing the portfolios of the existing credit products, and preventing deterioration in the credit standing of a bank's counterparties*" [3]. Financial companies typically use credit risk models to evaluate the insolvency risk caused by credits that enter into default,

According to BCBS, roughly speaking, the tradition in consumer credit has been to take performance data over a fixed time interval for a sample of consumers. Then each consumer is ranked by performance, where unsatisfactory performance is often equated with being 60 or 90 days overdue with repayments. This historical information is reviewed by risk analysts and is then used to model internal risk strategies based on the characteristics of each consumer. The modelled strategy is then applied to new consumer to determine whether they are above the cut-off level to
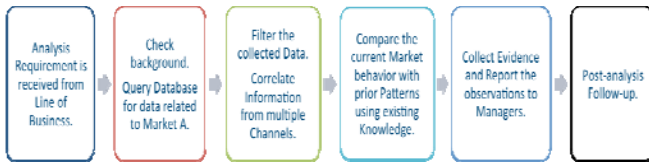

Figure 1. This is a typical risk analysis workflow observed and summarized from risk analysts in BOA.

be accepted to open a line of credit.

As shown in the analytical workflow (see Figure 1.), the risk analysts perform the credit analyses based on the following considerations:

- Requires for assessing the performance of credit products. For example, if the net loss of a credit product exceeds a certain threshold, the analysts need to evaluate the potential risk caused by that product.
- Shifts in general credit investment strategy. For instance, in the change of market focuses (i.e. focus shifts from high-end customer to mid-range consumers), most of the existing credit products would need to be re-evaluated in order to accommodate the changes. The analysts need to maintain an appropriate credit administration, measurement and monitoring process.
- Optimization of the portfolio of consumers or markets: e.g. analysts need to minimize credit loss and increase the product revenues. The analysts need to operate under a sound credit-granting process.

Thus, determining the risk in a consumer credit product relates to multiple factors that are intertwined with each other. While a set of queries can be easily used to retrieve delinquency patterns matching a limited set of existing hypotheses, risky market behaviors or hidden investment opportunities are more implicit and elusive. They are determined by a dynamic analysis context influenced by various factors, such as the markets' geospatial distributions, the prior investment strategies, and the competitors' behaviors. Hence, in the current practice, risk analysts often need to construct queries over multiple facets of a large body of credit data, searching for certain statistical values that may be indicative of high risk. Manually correlating these channels of information can be challenging and overwhelming: one type of analysis (e.g. only analyzing the known risks) would miss the hidden investment opportunity that could potentially lead to lost profits for the bank; whereas another type of analysis (e.g. only retaining portfolios with very low delinquency rates) could harm their relationships with their clients or regulatory agencies.

Exacerbating this challenge is the increasing size and complexity of the collected credit data that each analyst needs to examine. Millions of detailed consumer risk incidences are aggregated monthly to indicate the market behaviors, including information about the banks and their competitors, the state of the local economic environment, the consumer demographics, and the third party credit ratings. Thus, to comprehend the trends and patterns of a market and identify the investment opportunities, risk analysts are responsible for developing methods to utilize that large data corpus and predict the likelihood that a client who borrows money from a financial institution will default or fall behind on a loan payment. This places an extra burden on the individual analysis process in terms of the efforts needed to acquire the most appropriate information. At present, risk analysts do not have the capability to investigate all the patterns and activities, not to mention conducting thorough analyses of data over time.

In summary, credit risk analysis is an important yet challenging analysis process, due to the increasing size and complexity of consumer risk data. Effective analysis tools are needed to address these challenges. On the one hand, such tools must be tailored to specific credit risk management practices depending upon the nature and complexity of the business's credit analysis activities. On the other hand, they must enable comprehensive analysis of a large body of credit data to help verify known hypotheses, as well as allow discovery of hidden features in the dataset.

3 RELATED WORK

In recent years the idea of using visualization to support financial analysis has gained a lot of interest. The forefront of visual analysis within the financial market can be categorized in several main sectors, namely temporal analysis, market analysis, and investment analysis. For temporal analysis, several efforts have previously investigated representing and analyzing changes in financial data [1, 2, 8]. Particularly, recent work by Ziegler et al. [20] presented useful clustering techniques to help visualize the temporal financial changes.

To support market analysis, Keim et al.[9] have presented pixel-based visualization techniques showing the performance of individual stocks in high detail. In addition, the tree-map-based "Map of the Market" by Wattenberg[17] had introduced an effective way to examine and compare different market and stock performances online.

Finally, many researchers have emphasized the integration of visualizations and mathematics in maximizing investment profits [11, 16]. A representative work in this sector is the work by Maciejewski et al. [13], which combined the Winner's and Loser's Curse to reach optimal investment decisions.

Given the complexity of risk analysis described in the previous section, these individual visualizations are not particularly suited for depicting multidimensional data on their own. However, these visualizations can be integrated into systems

divisions and grounded our investigation on a field study conducted within the bank.

During the design phase of this project, we communicated with risk analysts from this division on their current practices as
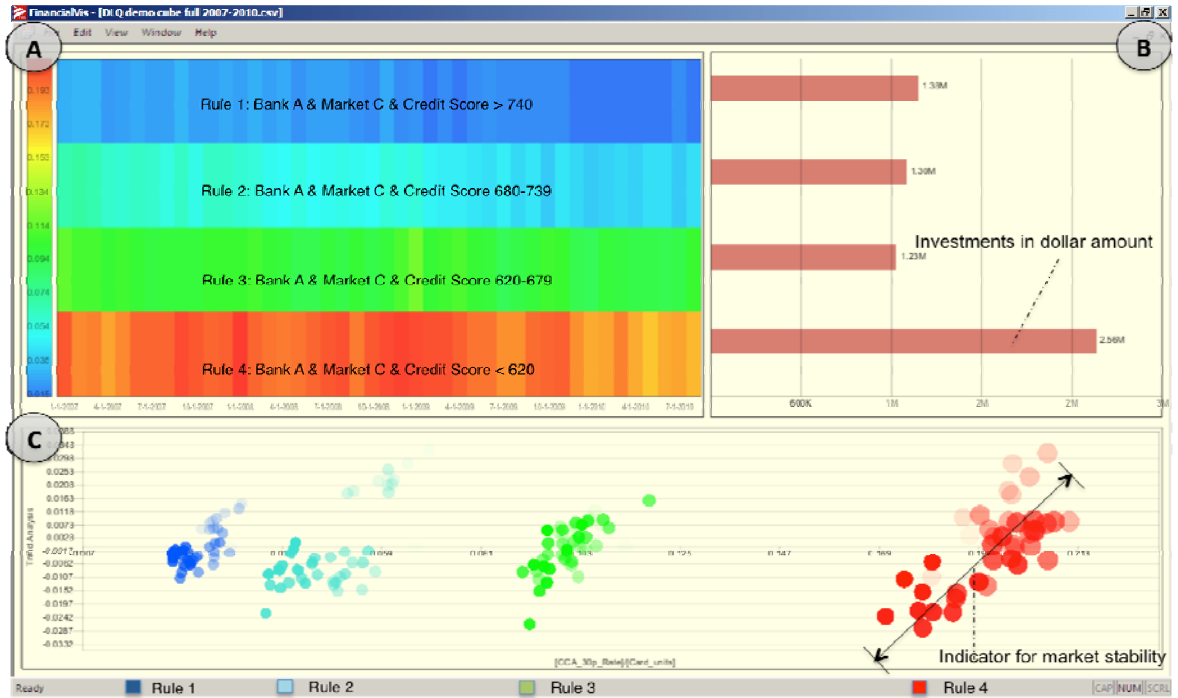

Figure 2.This is an overview of RiskVA, including the *Entity Heatmap* view (A), the *Product Comparison* view (B) and the *Trend Analysis* view (C). Here, analysts examine the 30-day delinquency rate (late payment) in Market C across a wide range of consumers.

handling multidimensional data by using multiple coordinated views [12] in conjunction with highly interactive exploratory techniques. Systems for integrating large data source, such as JigSaw[15], PatentVis[10] and IRSV[19], enhance analysis tasks by providing an integrated visual environments for direct data manipulation.

While those systems provided feasible approaches for flexible data explorations, the situation in consumer risk analysis differs from such scenarios due to its time-critical nature, which requires a concise yet dynamic structural visual interface to effectively performance both tactical and strategic analyses on emerging risks. Examples for systems that intend to address such analysis are GTDVis [18] and WireVis[7]. Compared to GTDVis, which reveals global terrorist attacks patterns based on individual attributes, our system emphasizes revealing the combined impacts to the credit markets based on associating multi-channel risk information. While both RiskVA and WireVis present concise visualizations for large financial datasets, RiskVA also supports a customization mechanism to support the analysts to dynamically generate rules and to create analysis workspaces.

In the following sections, we detail the challenges of the current risk analysis process, and describe our visual analytics system, which is designed to address these challenges by encoding the essential analyses into a cohesive visual analysis environment.

4   IDENTIFYING CHALLENGES AND ANALYTICAL REQUIREMENTS IN CURRENT RISK ANALYSIS PROCESS

To design a visual analytics system that is tailored to the risk analytical workflow in Bank of America, we established a long-term collaboration with its Consumer Credit Risk Solution

well as their needs for good analysis of consumer credit risks. We carried out multiple interviews and discussions to observe the day-to-day operations performed by the risk analysis team. The interviewees held a broad range of positions, including risk analysts who focused on analyzing the consumer credit products, managers who were in charge of business planning and crafting risk management strategies, and risk management architects who emphasized the identification of novel technologies that could be useful for risk analysis.

The interview data collected was used to characterize these analysts' task activities, and further used to develop the design requirements for a visual analytics system. Interviews with representatives from this team revealed the analytical needs of the risk analysts, including fusing multiple streams of data, retrieving information for context-dependent tasks, and analyzing their findings.

In general, the risk analysis team constantly needs to respond to market changes and conduct analyses involving the assessment of asset quality, the adequacy of provisions and reserves, and the balance of delinquency and investment. In addition, they are required to generate shared results effectively (e.g., a report of delinquency analysis or a summary of a market performance).

As specified in Figure 1**,** the analytical tasks of conducting risk analysis often include requirement specification, data aggregation, information organization and correlation, and result sharing. To analyze risk of a credit product, an analyst often starts by gathering relevant content from multiple data sources for a comprehensive view of that product. This aggregated dataset not only includes BOA's own data, such as the delinquency rating for that product (e.g. 30-day payments), but also data from credit rating agencies, such as credit scores. To improve their own

assessments of risk, the analysts then filter this large collection of data and attempt to organize it in a clear and consistent manner to support the awareness and sense-making process.

Tools, in this context, are considered as a means to transform their hypotheses into desired task actions. Currently, the risk analysts primarily use tools, such as SQL databases, Excel and emails, to produce and communicate analysis related contents. In the process, the analysts' prior experiences (i.e. knowledge of a potentially deteriorating credit product) are used, and further task actions are taken to be used in their analytical process. Although these analysts currently use a number of different tools, we found that they were lacking tools actually designed to support their analysis workflows and provide the detailed information they need. This finding demonstrates the need for a tool that supports the users' analytical workflows and helps them effectively perform necessary analysis actions.

Therefore, the primary goal of our system is to address these challenges in accordance with the analytical requirements of the risk analysis team. A detailed characterization of these three challenges, as well as how they are addressed in our system are described in the following sections.

4.1 Support the Identification of Emerging Risks

Consumer risk analysis requires analysts to "know you customers". This includes knowing the performance of credit products for the customers individually, in commercial markets, and statistically. An important part of knowing the customers is the assessment of their overall activities in terms of risk. Certain credit products (e.g., credit loans, because they are unsecured debt without collateral) are inherently riskier than others. The corollary to "know your customers" is "know your investments." A financial institution must know where and how its credit investment is being spent in order to accurately assess the emerging credit risk. With the limited analytical tools available today, performing this comprehensive risk analysis is not easy.

A synthesis of credit information and investment knowledge from all relevant sources is therefore needed for risk analysis. As described earlier, risk analysts are demanding tools to support analysis of multi-channels of credit information. These tools must support not only data integration from multiple information channels, but also information correlation that brings together end products such as knowledge of delinquency ratings that represent the down shifting of a credit market.

Especially, the analysts require tools that focuses on the following aspects:

- **Market Analysis:** The changes and trends of a credit market have significant influences on the investment strategies for a bank. Typical markets are metropolitan areas where the bank focuses its credit products. The analyses of these markets concentrate more on performance than physical locations. According to risk analysts, examining the health of a credit market is a primary task for risk analysis.
- **Temporal Analysis:** By thoroughly analyzing the temporal changes of a credit product, the analysts can compute the deterioration rate of that product and its related net loss. In addition, a risk analyst can adjust their future product investments by assessing the outcomes under changing credit conditions or from previous behavior. Therefore, the ability to capture temporal information is of great value to risk analysis when assessing the risk of a credit product. However, temporal analysis in existing tools is limited to a per product basis. Having a complete picture of the fluctuation of products in all markets that could help risk

analysts spot abnormal investment behaviors could be very beneficial.

- **Product Comparison Analysis:** Typically, the bank invests in a wide range of credit products to maximize its investment profit. While current query-based analysis could help risk analysts follow the changes of a credit product and typically on known relations, it only allows risk analyst to focus on a limited set of credit products. Given the diversity of the products, tools that could assist risk analysts comparing and comprehending the behaviors of these products while uncovering hidden relations would be helpful.

In practice, the risk analysts often examine a mixture of these analyses, such as depicting the trends in the markets through both market and temporal analysis, or examining a product's impacts to the markets using both market and product comparison analysis.

4.2 Support More Strategic than Tactical Analysis

Due to the burdensome necessity of analyzing multi-channel credit information, analysts often think narrowly about their investigative tasks. In particular, they think in terms of known patterns or in terms of activities that have been identified externally (e.g., from line of existing business or client) rather than in terms of what the patterns in the data are revealing to them.

In addition, consumer credit analysis requires the analysts to know the patterns of market changes for the customers individually, in different markets, and statistically. However, the analysis team is often hindered by the analytical tools, the time, or the sufficiency of evidence in thinking more broadly about the meaning of the credit products in terms of larger strategies to determine risks (especially previously unknown risks) and uncover hidden investment opportunities or their benefits to the overall banking strategy.

Therefore, careful thought must be given to gathering information to assist the analyst in providing objective, fully reasoned assessments backed by evidence and avoiding inevitable pitfalls and biases. Hence, analytical tools must be able to help the analyst effectively cut through noise and irrelevant data, explore the large body of credit data, and combine information from multi-channel information sources into a strategic risk analysis.

4.3 Support Individual Analysis Routine

Risk analysis is currently an art that is learned through long apprenticeship and then practiced. It is embedded in the creative reasoning processes of the practitioners. While Figure 1 illustrates some general analytical workflows in risk management **[3]**, specific credit risk management practices may differ among analysts depending upon the nature and complexity of their analysis goals and expertise. Sometimes, even the same analysts need to take alternative analytical practices just to accommodate the changes in focuses and priorities. Therefore, it is important for tools to support such diversified analytical needs and provide risk analysts with the flexibility to combine and sequence the analytical components to fit their own workflow.

5 RISKVA: A VISUAL ANALYTICS SYSTEM FOR ANALYZING CONSUMER CREDIT RISKS

In response to these identified analytical requirements for an integrated, efficient, analytics tool tuned to the consumer credit risk analysis environment, we designed RiskVA, an interactive visual analytics system that helps domain analysts in depicting and comparing the performance of the credit products by visually revealing market fluctuations and temporal trends of the targeted

credit products. Throughout the system design and implementation phase, we maintained close communication with the consumer credit risk solution group and routinely showed our progress and received feedback for our prototypes.

In the following sections, we first explain the data integration process that enables RiskVA to effectively combine multiple information channels, and then describe how each of the analytical requirements is depicted in our system. For privacy and proprietary reasons, details about the market information, bank associations and consumer information have all been anonymized in the following sections and in the figures. Of course, when the bank analysts use RiskVA, all this information is depicted.

5.1 Data Integration

At the heart of RiskVA is a data cube structure [14] that is customized to handle the large-size and complex credit data. Shown in Figure 3, the design of a three-dimensional cube structure was determined collectively with inputs from the risk team and provided them with a means to correlate multiple-channels of credit risks.

On the conceptual level, this data cube structure is specified to accommodate the risk analysts' tasks, presenting the rules of the desired credit information. In particular, such rules are constructed around the key elements in risk analysis, including the *entity* (e.g. FICO scores, wealth level, market ID, and etc.), *the temporal information*, and the *credit product variables* (e.g. credit loans, mortgage, and etc.).
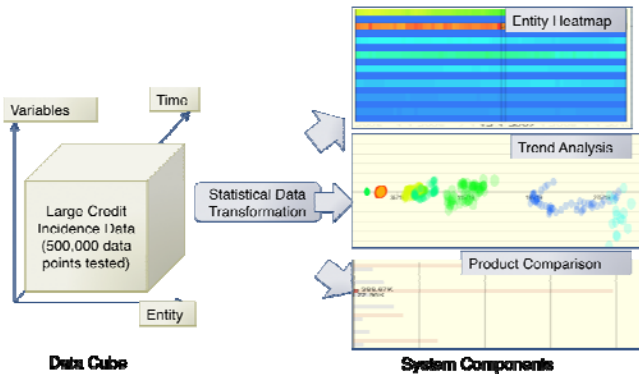

Figure 3. This is the system pipeline for RiskVA

The use of rule (e.g. market A & bank B & Time C) in the cube structure can effectively help risk analysis to navigate through the large body of high-dimensional credit data, and locate the desired information. For example, the rules can help to answer specific questions like, "What's the behavior of consumers with FICO score below 690 in market M?" In addition, the rules can be recorded and reused in different analysis scenarios, making it possible for RiskVA to trace individual's analytical process and share the analytical evidences between groups of analysts (section 5.3). The ease of creating rules provides the risk analysts with the flexibility to customize the credit information to fit with their own analytical workflow. RiskVA utilizes the rules to support the risk analysts in customizing their individual analytical environments (section 5.3).

Finally, the rules are an essential part in the constructions of the visual representation. They inform the views to filter unnecessary data elements and to present the analyst with most relevant information. RiskVA further uses these rules to coordinate and update the visualizations (section 5.4).

On the implementation level, this data cube follows the previous work [14] on creating an effective structure for slice-

and-dice data from multiple aspects. As illustrated in Figure 3, to optimize the memory usage, the cube structure is disseminated into three parts: the meta-cubes that stores the rules, the virtual-cubes, which enable the comparison of credit products over a large dataset, and the physical data cube that points to the actual credit data. Due to page limits, the details of the implementation are beyond the scope of this paper.

5.2 Design Interactive Visualizations to Support the Identification of Emerging Risks

Since one cannot usually depict diverse rules, we need to visualize the activities of the corresponding markets in order to reveal the behaviors of the targeted credit products. Given the complexity of the related information, no single view could fulfil all the analytical requirements and show all the necessary data. Therefore, RiskVA is designed as system of coordinated views that would allow the analysts to see different data, while being able to understand the connections between the views easily. In particular, RiskVA encodes the three essential analyses as described in the following with a set of visualizations, each of which correspond to facilitate market analysis, temporal analysis and comparison analysis, respectively.
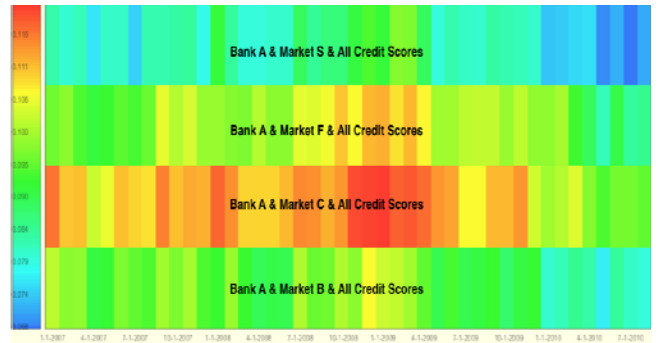
Figure 4. This is the overview of the *Entity Heatmap* view. Each Horizontal bar corresponds to a user-created entity rule.

5.2.1 Depict market behaviors with *Entity Heatmap* view

RiskVA utilizes entity heatmap to display the statistical measurements associating investment markets (e.g. population in the US West with credit scores less than 690) and credit products (e.g. mortgage or credit loans), as the former influences the behaviors of the latter. As shown in Figure 4, the heatmap is based on a grid where columns are the timestamps, and whose rows are rules that indicate the market performance of different credit products. This design aims to provide the analysts are direct sense of how the markets' performances over time.

At the intersection of a particular column and row, the cell is color-coded with a value derived from the combination of market/credit card in that time-period. Such values are associated with the market performance of a particular credit performance indicator, such as the 30-day delinquency rates or credit loans. Depending on the users' focus on the measurement displayed in the grid (e.g., difference value range or granularity of increments), RiskVA enables the user to interactively apply various color schemes to the visualization. At the beginning of each market analysis, the heatmap presents the overview of the targeted credit performance indicator, using a simple scheme where the color is computed based on the min/max values. In doing so, RiskVA aims to provide analysts a common picture that helps them

quickly recognize some significant market trends and changers, informing the overall strategies about the targeted markets.

Using the entity heatmap view, the risk analyst can then spot at a glance the markets that perform worse than the other ones, relating to a given credit product. As shown in Figure 4, the heatmap view is enhanced with a user-configurable rule capability that makes it possible for analysts to visually compare patterns of credit product behavior across different markets. For example, Figure 7 (right) shows the comparison of overall markets fluctuation between two financial institutions in the period of the 2007-2010. The color-coded heatmap (Figure 7 (right)) clearly indicates the performance differences between two institutions, where the lower delinquency rate (blue) suggests one has more stable credit products than the other.

In addition, once general understandings about a particular credit product are established, RiskVA would further facilitate the analyst in deepening his understanding of the impact of that credit product to the market; it enables the analysts to interactively parameterize the heatmap view with more focused rules, and helps them depict the market impacts with finer analytical context. In many cases, this helps analysts to disseminate the general market trends into multiple populations, and to adjust their investment strategies based on the risks associated with each group. As shown in Figure 7 (left), a further analysis of the lesser-performing institution indicates that the population with lower than 620 credit score and larger debt tends to pose higher risks than other populations.

We designed the heatmap view to be highly interactive. It enables the analysts to interactively select, highlight, and sort credit information; each of which is accompanied by detail tooltips. It also allows the analysts to apply statistical analysis over a particular time period or investment market for vertical and horizontal analysis (e.g. Compound Annual Growth Rate (CAGR) and standard deviation).

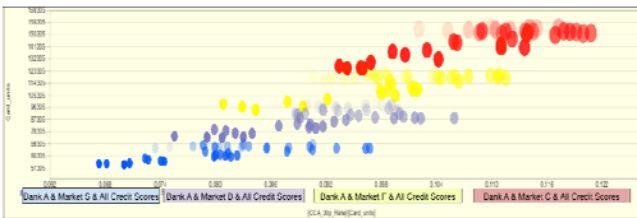### 5.2.2   Reveal temporal patterns by *Trend Analysis* view

Figure 5. The overview of Trend analysis.

The ability to examine the stability of a credit market over a period of time is of great value to the analysts. Especially, when a market is turbulent, the analysts are required to identify the most vulnerable consumer credit products, and to report the possible cause of the market fluctuation. Through manually examining the changes of all the credit products prior to that time frame, the analysts may eventually be able to identify the weak line of the products line and report a plausible cause. However, current practices and tools still limit such crucial temporal analysis to analysis on a per product basis. Our coordinated visual interface enables the analysts to effectively identify these temporal patterns.

RiskVA utilizes the trend analysis view (Figure 5) to support the needed visualization of the market behaviors over time. It shows the overall temporal performances of each credit product, and allows the analysts to compare the stability of that product in different markets. The x-axis of this view shows the progression of time, and the y-axis shows the performance of each credit

product in a particular market. Such performance can be actual investment amount of the product (e.g. total/average investment of mortgage loan), or it can be the trend that is calculated to indicate the stability of that product (i.e. the numerical differences between the current investment cycle and the last one).

As illustrated in Figure 2 (C), each dot represents the 30-day credit delinquency rate, and its transparency shows the temporal trail for that market. The more opaque, the closer that market is to the current time. If a market shares less drastic changes, such as the green dots to the left, it would reside in a more clustered group. On the other hand, if the market is like the one shown on the right in Figure 2 (C), where the red dots a less grouped, this pattern suggest the delinquency ratings (i.e. the dots) in that market have changed drastically over the years, indicating a larger fluctuation over the previous investment cycle.

To facilitate efficient interactions with the temporal analysis, the analysts can quickly select a specific market or product by hovering the labels. The analysts can further examine the details of a specific time period or certain product/market, through interactively filtering to the desired analysis items.

5.2.3   Compare products in *Product Comparison* view

Comparing performance between credit products plays an essential role in determining the bank's investment strategies. If a credit product continues to pose net loss in a particular market, this should be quickly identified and further inspected by risk analysts.

Figure 6. This is the *Product Comparison* View. Here, the analysts is comparing the investments distributions in two groups of consumers.

To present the comparisons between credit products, we use an interactive bar chart view as shown in Figure 6. In this view, the credit products involved in a particular market are grouped and represented in horizontal bars, which are directly associated with the rule in the heatmap view (see section 5.2.1). Different products are distinguished by the assigned colors (e.g. blue for mortgage and red for credit loans). The length of each bar corresponds to the actual investment values of that product, and can be further customized based on aggregation methods such as summation or average.

By placing the different credit products side-by-side, the analyst can directly compare the performances of these investments for different consumer populations. As shown in Figure 6, the analysts can clearly verify the well being of the targeted market, given the fact that the credit investments in that market are distributed proportionally to the consumers' credit scores (e.g. the bank invests more on consumers with higher credit scores and less on consumers with lower scores). To further examine the details of these credit products, the analysts can mouse over each bar and correlate it with the information in the heatmap.

5.3 Create Customizable Workspaces to Support individual analysis routines

To support the diversified analysis goals and user preferences, RiskVA presents the analysts with customizable analysis workspaces. Workspace is the main analysis environment of the RiskVA system. We define workspace as a user-configurable combination of the above three visualizations that provides a

can use RiskVA to revisit their previous workspaces and to continue their analyses.

5.4 Coordinate Views to Support Strategic Analysis

The utilization of entity rules provides a foundation for RiskVA to coordinate between the three views and workspaces. Since the underlying structure for data cubes (e.g. meta-cubes and virtual-cubes) is the same, the information passing between these view
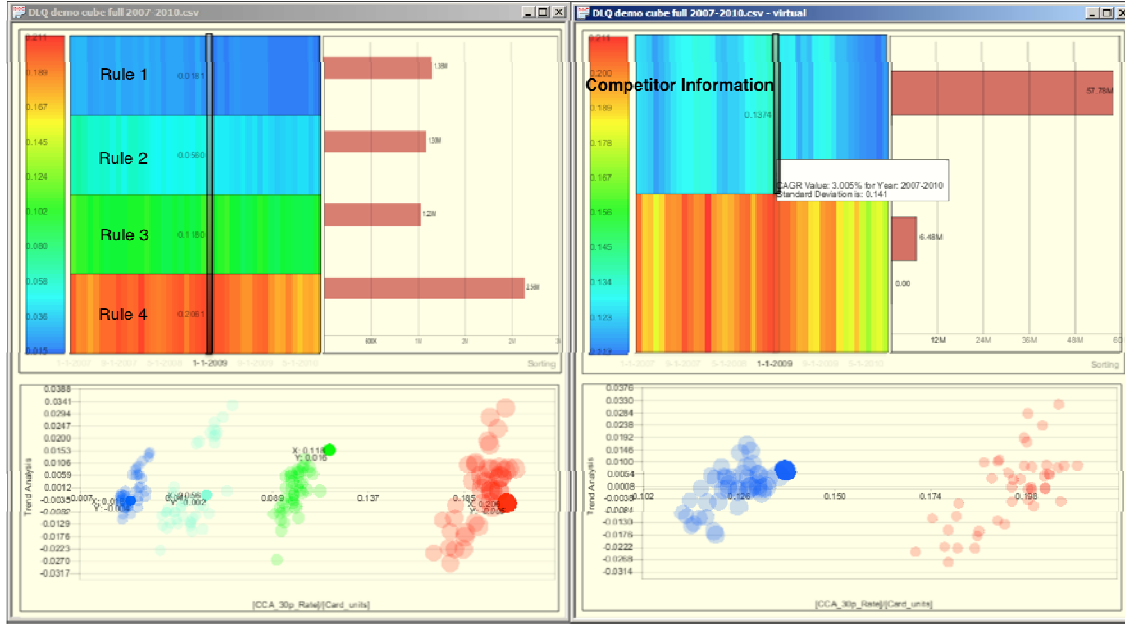

Figure 7. Two workspaces placed side-by-side. Here, analysts examine both workspaces together, looking for correlations between banks.

particular analysis context. Much like a probe system [5], each workspace builds upon the data cube structure and empowers the risk analysts to depict multiple-channels risk information at the same time. For example, the analysts can now examine and compare two different credit datasets for market performances in different analysis contexts. See Figure 7.) Moreover, RiskVA also handles workspace coordination. As detailed in the next section, it compares individual entities in different analysis context (e.g. credit load performance in US West v.s. US East). If there are matching entities (e.g. date, market, or rules), RiskVA can then link the multiple workspaces using this information and provide the analysts a coherent understanding about the risks that are embedded in different contexts.

During the analysis process, RiskVA allows the analysts to have full control of the creation of a workspace. Given their analytical needs, the analysts can directly branch off their analysis by duplicating the existing workspaces or construct a completely new workspace with a different set of data. Such a branching mechanism enables risk analysts to easily shift their focus while still maintaining their train of thought embodied in the branching structure. In addition, RiskVA also helps the analysts to customize their workspaces by providing them the flexibility to combine and sequence the analytical components to fit their own workflow (i.e. remove or hide certain visualizations or apply specific statistical analysis equations).

To help analysts maintain their train of thought and resume or repeat their previous analysis processes, RiskVA automatically records the history of the users' workspace usages. It logs the time when a workspace is created, branched off, and closed. It also records the above system customizations (user's view preferences during a particular analysis). Thus, at any given time, the analysts

representations becomes straightforward. Two levels of coordination mechanism are implemented in RiskVA currently, namely, within workspace coordination and between workspace coordination.

For within workspace coordination, RiskVA tightly coordinates the three views together so that updates in one view are immediately reflected in others. The analysts can now simultaneously interact with credit products, markets, consumers and etc., and monitor the correlation between all the information channels in a cohesive way.

For example, when an analyst mouses over a particular market in the trend analysis view, all related cells in the heatmap are highlighted to illustrate the consumer base for that market. At the same time, all the credit products that are associated with this market are also highlighted in the product comparison view, showing their performances and correlations.

For between workspace coordination, the analyst can hover over the dates (or another entity) in the heatmap view in a workspace, highlighting all the consumer information of that particular month (Figure 7.). All the views in the other workspace then highlight items involving the same entity in their own analysis context: the heatmap will highlight the related markets; the trend analysis will present the appearance dates; and the comparison view will show the credit product that may associate with these items.

This coordination is especially useful validating investment strategies in multiple market environments and verifying results in different analysis contexts. This coordination is made possible largely because of the shared properties that generally exist in many risk datasets at banks.

Both levels of coordination bring the risk analysts a cohesive depiction of the consumer risk market with coordination among

important views and correlation between different significant contexts.

6 ANALYSIS SCENARIOS

Identifying and understanding the cause of market deterioration is a key step for risk analysts to develop corresponding investment strategies. To conduct an unbiased risk analysis, it is necessary for analysts to monitor and compare credit products for both the competitors and their own. Based on our discussions with the risk analysis team at BOA (2 risk managers and 5 analysts), we have observed that the analysis of a market typically follows three analysis stages, namely identifying risky markets, comparing product performance between banks, and identifying potential causes for the fluctuation. The following scenario was identified with the risk analysis team at their regular strategy meeting. The agenda for this meeting was to discuss the bank's credit product performances in several major investment markets using RiskVA. As observers, we documented their analysis processes and helped them become familiar with the system. We also provided explanations about certain features in our system during their exploration.

To pursue this scenario, as shown in Figure 2, RiskVA was initialized with consumer credit data from year 2007 to 2010. To depict the performances of individual markets, the risk analysis team utilized the entity heatmap view to check if any interesting fluctuation pattern could be identified. The team found a set of markets with high 30-day delinquency rates (warmer colors in Figure 2 (A)) over the entire time span, which indicated a large body of late payments and the potential net losses for the bank. To get a clearer picture of the behavior of those markets, the analyst turned to the trend analysis view to examine the development of these markets. As illustrated by the cluster of red dots in Figure 2 (C), the analysts noticed that one investment market (red dots) was particularly vulnerable (i.e. the cluster scattered over the time), suggesting an unstable product performance since the beginning of the credit crunch in late 2007 until recently. Given that the general investment market is recovering since 2009, the analysis team decided to first take a closer look at unusual market behavior.

Instead of drilling down to that market, the team utilized the built-in workspaces to branch off their current analysis to keep track of their analysis processes. As shown in Figure 7, the team created a new workspace to compare the performance of the bank's own credit products with other competitors in that market. A quick glance at Figure 7 indicates that the competing banks on average invested more in that market and maintained a quite healthy performance. This finding immediately raised several questions: could the client base affect the market performance? Or was it caused by the unbalanced or sudden increase of investments in that market (e.g. mortgage v.s. credit loans)? Although these were all possible causes of the market deterioration, the risk analysts had no definitive answers or evidence to confirm their hypotheses by looking at the product comparison view alone.

Trying to verify these hypotheses, the team started to search for clues from the investment history of that market. By using fine-grain rules, they found that the investment in that market had always been a steady amount, and a reasonable proportion between secured and unsecured credit products. This therefore rules out the possibility if investments patterns being the cause of this fluctuation. However, a closer examination of client bases in the heatmap view suggested a different story. Figure 2 (B) showed that, compared to the typical strategy of pursuing consumers with higher credit scores, these markets invested on a fair amount of population with lower credit scores but with long credit histories.

A quick check on the trend analysis view (Figure 5) further indicated that, the trend of this consumer group has peaked since 2008. A quick reference to the recent financial news around that market confirmed that there was an increasing amount of unemployment in market, which gave the risk analysts reasons to conclude that the changing in client bases may be a key factor in causing the fluctuation of this market.

Given the unsatisfactory market performance and concerns for losing more investment, the risk analyst team indicated that this market needed more attentions to bring its performance back on track. They also decided they needed some strategies to alleviate the pressure imposed on the customers and to help revive the market. After this exercise, the risk analyst commented on the effectiveness of RiskVA in helping them to explore the credit markets, as well as in identifying possible cause of market fluctuations. Although simple, the scenario has demonstrated the usefulness of RiskVA in support of strategic risk analysis.

7 EXPERT EVALUATION

To assess the efficacy of our system, we conducted expert evaluations with consumer risk analyst and risk managers from Bank of America. The goal for this process is to perform summative evaluation to measure how well RiskVA could facilitate the actual risk analyses.

During several on-site visits, we demonstrated design of the system and the utilities of the visualization to a total number of 8 risk managers and analysts from or related to the consumer risk analysis team. We invited risk analysts to perform in-depth analyses using the system, in a think-aloud manner. We observed and documented the details about their analysis processes. Finally, we concluded the evaluation by gathering their feedback and comments about our system. Since RiskVA has been deployed to this team, we also conducted email follow-ups to see if there were additional comments they would like to share with us.

7.1 Visual Facilitation on Tactical and Strategic Risk Analyses

One of the benefits of RiskVA that was noted by all risk analysts was its visual exploration environment that enables them to perform more strategic rather than tactical analysis. All the risk analysts consider being able to interactively perform all three analyses (i.e. market, temporal, and comparison analyses) at the same time to be powerful in portraying the detailed, dynamic nature of the emerging risks.

To support tactical analysis, RiskVA allowed the analysts to utilize their prior knowledge about the market to efficiently verify known risk patterns, and helped them to pursue their tactical goals using resources at hand. In RiskVA, rules were used to facilitate the analysts to interactively filter and analyze the credit information at different granularities.

When this capability was presented to the risk analysts, they spontaneously formulated a variety of rules to find credit information in the current market. All analysts were generally satisfied with the efficiency of using RiskVA and appreciated the flexibility to perform customized analysis. Specifically, one of the seasoned analysts pointed out that the ability to do such interactive analysis served two roles in supporting his tactical analysis. On the one hand, he considered RiskVA as an efficient method to "slice and dice" information to monitor market conditions and to test hypotheses. On the other hand, he thought the current implementation of RiskVA addressed another important aspect in risk analysis: the ability to verify and validate the accuracy of this new technology. Being able to interactively construct rule helped him match and confirm his prior

expectations and the visual representations, and gain confidence in using the system in his daily analysis.

In assessing its efficacy in handle strategic risk analysis, many analysts considered RiskVA to have the advantage to let the data tell the story about where risks are emerging. In particular, it reduced the amount of noise they have to sift through in order to see the broader picture and home in on suspicious outliers, enabling the analysts to explore possible risk patterns that were previously unidentified. They agreed that our tool assisted this analytical process by visually providing a global pattern as well as details on demand. As demonstrated in the scenario (see section 6), RiskVA helped risk analysts to effectively analyze their data across multiple dimensions and assisted them in determining the cause of market deterioration. All the analysts found the system practical, and believed that the system would be useful in helping them to perform more strategic risk analysis. As summarized in one of the analysts' comments, "[RiskVA] first provides me the general idea of what's going on with the market. And quickly and interactively let me navigate into a specific interested analysis segments. It allows me to get more hands-on analysis, and to check what I might have missed in analyzing the data".

One suggestion was to provide additional geospatial analysis, and display more regional risk entities. Analyst would like to see incorporated information like distances, densities, and areas in analysis as possible explanatory variables. One participant suggested that "some higher resolution geospatial view can be used to drill down below the census Bureau's Core Based Statistical Area (CBSA) level, or create a geographical segmentation of the US that's independent of CBSA, ZIP, or districts. Then we can compare our performance in these new geographic entities."

7.2 Customizable Workspaces and History Tracking

Using the workspace metaphor, RiskVA enabled the analysts to perform their tasks in a flexible and customizable environment. It provided the analysts with the flexibility to interactively combine and sequence different visualizations, customizing the workspace to fit their individual analysis routines. RiskVA utilized the workspace structures to sustain a dynamic analysis environment; it enabled the analysts to branch off their analysis at any time by duplicating any existing analysis workspaces. All the created workspaces were coordinated through the identification of similar market entities.

All participants appreciated the flexibility of the interface, finding it useful for customizing the system to only utilize the necessary visualizations in their analyses. In particular, they liked the ability to construct different workspaces to simultaneously analyze different markets groups. They thought this would be quite helpful in understanding the relationships between these markets. One of the managers commented that, "[RiskVA] brings the analyst at the center of the analysis, with their subjective attitudes, to interact with the data. This is where I see visual analytics can improve our risk management process."

Furthermore, RiskVA logged the analysis workspace history to enable the analysts to capture and revisit their previous analysis states. Most analysts found the idea of tracking analytical trails intriguing. While this was still a preliminary feature, the analysts had already noted its effectiveness in managing their diverse analysis practices.

8 DISCUSSION AND FUTURE WORK

We undertook this research to design a visual analytics system that facilitates the risk analysts' tactical and strategic consumer credit risk analyses. To this end, we presented RiskVA, an

interactive visual analytics system that demonstrated unique and effective capabilities for a class of problems that involve complex consumer risk analyses.

The design of RiskVA is grounded in the task analysis results of a group of risk analysts from Bank of America. These results provide us clear identification of general domain analysis process, including fine-grain task activities, task flows, and overall analysis objectives. We further disseminated this general analysis process into individual analytical requirements (section 4), and transformed them into the specific system implementations through iterative prototyping with the risk analysts (section 5). Given the positive feedback from risk analysts, we found such design process to be particularly effective in designing an analysis-rich visual analytics system.

There are limitations to our research that should be addressed. Specifically, our research characterized the domain analytical workflow through interviews, which generally are self-reported by participants. Our research could also be limited, in that it modelled the analytical workflow from a retrospective viewpoint, whereas Brows et al. [21] demonstrated that problem spaces and solutions are established and changed dynamically in interactions with people and the environment. Therefore, our understanding of domain task flow may be constrained to the risk analyst' general way of performing tasks. One effective way we used to alleviate this constrain is enable the risk analysts to customize RiskVA to fit their own analytical workflows (section 5.3).

In addition, our research is limited by its evaluations with domain experts. Given the privacy and proprietary considerations in BOA, we only evaluated RiskVA through expert evaluations. While the results are positive, we believe much can be learnt if alternative methods were available. In particular, we would like to evaluate the risk analysts' knowledge gain from using our system. However, developing evaluation strategies to accurately assess the effectiveness of a visual analytics system is challenging yet beyond the scope of this paper. At this point we do not have a clear outline on the best evaluation approach; the design of guidelines for systematically evaluating a visual analytic system would be one interesting future direction for our research.

9 CONCLUSION

In this paper, we presented RiskVA, an interactive visual analytics system that demonstrated unique and effective capabilities for a class of problems that includes certain complex consumer risk analyses. RiskVA supports a thorough analysis of a financial institution's own data with data from other sources, including competitors, for a comprehensive view will permit these institutions to better make their own assessments of risk, independent of, and more focused than, assessments they get from ratings agencies that may not be suitable for the their situations.

By placing risk analysts in the center of their analytical processes, RiskVA provides analysts with customizable analysis workspaces, interactive data exploration, and the capability to correlate information over a large corpus of credit data. In our expert evaluations, risk analysts confirmed the novelty and utility of RiskVA to facilitate them in performing in-depth risk assessments, and further expressed interests in using it in their daily tasks. With such encouraging feedback, we are current deploying RiskVA to the consumer credit risk division in BOA.

These results indicate the efficacy of the cognitive task analysis process we undertook at the beginning. It is essential that this task analysis be carried out and also essential that the task analysis be made flexible to support exploration and unforeseen analyses. RiskVA and other tools we have developed indicate that this process is general. It is certainly clear that RiskVA is a tool

for emerging risk analysis that is applicable across financial institutions.

REFERENCES

[1] J. Allebach. Binary display of images when spot size exceeds step size. *Applied Optics*, 15:2513–2519, August 1980.

[2] M. Ankerst, D.A. Keim, H.-P. Kriegel, Circle Segments: A Technique for Visually Exploring Large Multidimensional Data Sets, in Proc. of Visualization 96, Hot Topics Session, 1996

[3] Basel Committee for Banking Supervision. www.bis.org/bcbs

[4] Nick Bloom (2007),'The Impact of Uncertainty Shocks',National Bureau of Economic Research Working Paper No. W13385,also available as CEP Discussion

[5] T. Butkiewicz ; W. Dou; Z. Wartell; W. Ribarsky; R. Chang, "Multi-Focused Geospatial Analysis Using Probes," Visualization and Computer Graphics, IEEE Transactions on , vol.14, no.6, pp.1165-1172, Nov.-Dec. 2008

[6] E. Catmull. A tutorial on compensation tables. In *Computer Graphics*, volume 13, pages 1–7. ACM SIGGRAPH, 1979.

[7] R. Chang.; M. Ghoniem.; R. Kosara; W. Ribarsky,; Y. Jing ; E. Suma, C. Ziemkiewicz,; D. Kern; A. Sudjianto, "WireVis: Visualization of Categorical, Time-Varying Data From Financial Transactions," Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on, vol., no., pp.155-162, Oct. 30 2007-Nov. 1 2007

[8] D.A. Keim, Pixel-Oriented Visualization Techniques for Exploring Very Large Databases, Journal of Computational and Graphical Statistics, Vol 5, pp. 58 -77, 1996

[9] D.A. Keim, T. Nietzschmann, N. Schelwies, J. Schneidewind, T. Schreck, and H. Ziegler. A Spectral Visualization System for Analyzing Financial Time Series Data. in Proc. of Eurographics/IEEE-VGTC Symposium on Visualization (EUROVIS'06), pp. 195-200, 2006

[10] S. Koch, H. Bosch; M. Giereth, T. Ertl; , "Iterative integration of visual insights during patent search and analysis," Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on , vol., no., pp.203-210, 12-13 Oct. 2009

[11] J. Lin, M. Vlachos, E. Keogh, and D. Gunopulos, Iterative Incremental Clustering of Time Series, Advances in Database Technology - EDBT 2004, 9th Intl. Conference on Extending Database Technology, 2004

[12] J. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV '07, pages 61–71, July 2007.

[13] A. Savikhin, R. Maciejewski and David S. Ebert. Applied Visual Analytics for Economic Decision-Making. IEEE Symposium on Visual Analytics Science and Technology (VAST), pp. 107-114, 2008.

[14] C. Stolte, D. Tang, and P. Hanrahan. 2003. Multiscale Visualization Using Data Cubes. IEEE Transactions on Visualization and Computer Graphics 9, 2 (April 2003), 176-187.

[15] John Stasko, Carsten Gorg, and Zhicheng Liu, "Jigsaw: Supporting Investigative Analysis through Interactive Visualization", Information Visualization, Vol. 7, No. 2, Summer 2008, pp. 118-132

[16] S. Rudolpn, A. Savikhin, and D.S. Ebert,; FinVis: Applied visual analytics for personal financial planning, Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on , vol., no., pp.195-202, 12-13 Oct. 2009

[17] M. Wattenberg. Visualizing the stock market. CHI Extended Abstracts on Human Factors in Computing Systems, pp. 188 189, 1999

[18] X. Wang, E. Miller., K. Smarick, W. Ribarsky, and R. Chang (2008), Investigative Visual Analytsis of Global Terrorism. Computer Graphics Forum, 27:919-926

[19] X. Wang, W. Dou, S. Chen, W. Ribarsky, and R. Chang, "An Interactive Visual Analytics System for Bridge Management", presented at Computer. Graph. Forum, 2010, pp.1033-1042.

[20] H. Ziegler; M. Jenny; T. Gruse; D.A. Keim.; Visual market sector analysis for financial time series data, Visual Analytics Science and Technology (VAST), 2010

[21] J. S. Brown and P. Duguid. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working Learning, and Innovation. Organization Science, 2(1):40–57, 1991.

Towards a Visual Analytics Framework for Handling Complex Business Processes

William Ribarsky   Derek Xiaoyu Wang   Wenwen Dou   William J. Tolone

Charlotte Visualization Center
College of Computing Informatics
University of North Carolina at Charlotte

Abstract

Organizing data that can come from anywhere in the complex business process in a variety of types is a challenging task. To tackle the challenge, we introduce the concepts of virtual sensors and process events. In addition, a visual interface is presented in this paper to aid deploying the virtual sensors and analyzing process events information. The virtual sensors permit collection from the streams of data at any point in the process and transmission of the data in a form ready to be analyzed by the central analytics engine. Process events provide a uniform expression of data of different types in a form that can be automatically prioritized and that is readily meaningful to the users. Through the visual interface, the user can place the virtual sensors, interact with and group the process events, and delve into the details of the process at any point. The visual interface provides a multiview investigative environment for sensemaking and decisive action by the user.

1. Introduction

The idea of "just in time (JIT) manufacturing" has been around for some time and has been implemented in large scale production environments. Lately there has been the need for *dynamic* just in time manufacturing where the distribution and even types of products produced may change fairly quickly, often in response to a previously unforeseen need. This is the case for some large government entities that must supply materiel in support of changing missions and for some large manufacturers.

In this paper we describe a framework we have developed in cooperation with a partner engaged in dynamic manufacturing. A major issue is that the business process (in this case manufacture and delivery of a variety of systems, which could be complex themselves or components of even larger systems) is
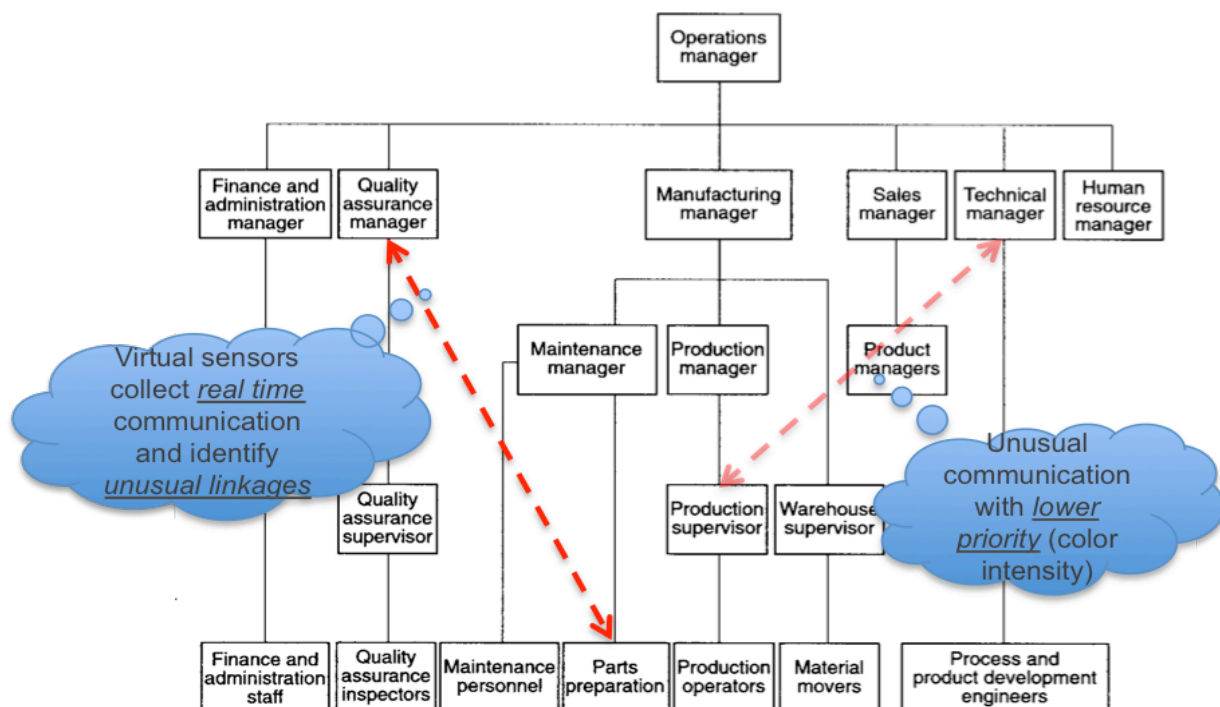

Figure 1. An organization structure for a production operation with unusual communication paths

complex and changing. It can be across several parts of a large manufacturing and delivery organization and across many different suppliers. A disruption at any point in the organization or among the suppliers could result in a failure to deliver full capabilities on time, which would be a serious problem.

A main issue that can arise is shown in the following example, illustrated in Figure 1. Suppose the assembly of a system that must be deployed by a certain date is dependent on the availability of a component that is provided by an external supplier. Further the component is not installed till late in the assembly process. Since this is a dynamic manufacturing process where different systems with different externally supplied components are required to be produced at different times, the organizational structure cannot be closely aligned with the manufacturing process. In this case the procurement and assembly components are in different part of the organizational structure with different reporting paths. In the illustration, a clue to the problem is given by unusual communication being opened up between the parts acquisition/ preparation department and the quality assurance manager. (This isn't the only possible path; the communication between the production supervisor and the technical manager could possibly be due to the same or a related problem.) There may also be unusual activity within a

department even without external communication. In Figure 1, this is picked up by a "virtual sensor" (discussed further in Sec.5.) Unless there is appropriate and timely communication across the structure and in particular to the operations manager (and above), the manufacturing process will break down. This is to be contrasted with a non-dynamic JIT process. In this case, the steps of the process are more fixed, permitting better organizational alignment from the start and better communication. Furthermore, non-dynamic JIT permits predictive planning where a forecast model based on past production and other factors is imposed to predict inventory needs, among other needs [1]. Such predictive methods will be much harder to develop in dynamic JIT manufacturing. Finally, most JIT planning approaches have been applied to much more uniform manufacturing processes (e.g., the production of a certain car model) than are the case here, where quite different systems are produced at different times.

The framework we develop here attacks this problem by applying visual analytics for exploration, discovery, hypothesis building and testing, and decision-making.
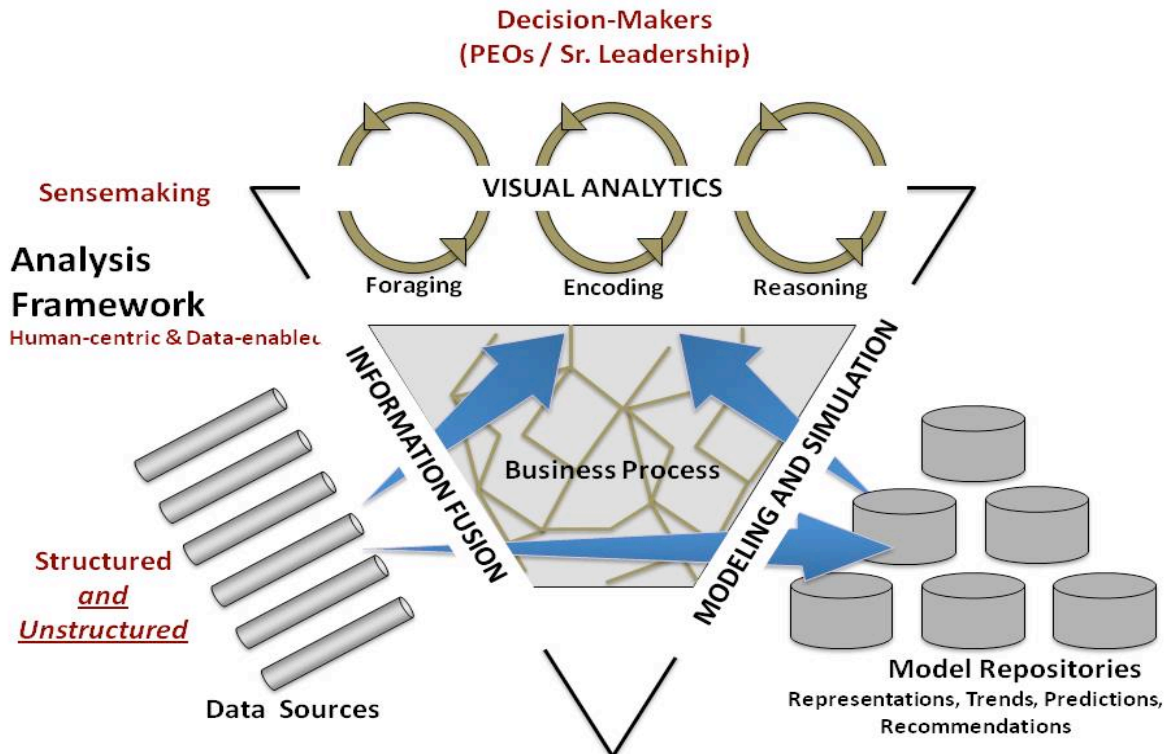
Figure 2. Overall visual analytics framework for complex business processes

2. Relevant Work

There is a history of business process visualization in the business industry [2,3,4,5]. In the area of customer relations management, Azvine et al. presented a configurable business process analytics tool that constantly monitors the performance of a decision model on both overall and individual levels with the goals of customer satisfaction and operational excellence [6]. Noting that existing business analytic applications are usually closed-loop decision making systems which only present output to operational managers, Azvine et al. incorporated visualization components into an intelligent business analytics system in order to put users in the loop and adjust business operations based on users' analyses in real time.

In the domain of Process-Aware Information System (PAIS), which pertains to both administrative processes and cross-organizational processes, visualization techniques have been used to aid in the understanding of process schemas and their run-time behavior through simulation [7] and process mining [8]. For example, to improve business process models, Alast et al. combined process mining with visual analytics to incorporate human judgment into otherwise static business process models [8]. On the one hand, process mining supports automatic discovery of a business model and checking how a process model conforms to actual process executions. On the other hand, visual analytics combines automated analysis with interactive visualizations to allow decision-makers to apply their flexibility, creativity, and domain knowledge to come to an effective understanding of situations in the context of large data sets. The authors noted that insights obtained from visualizations could be used to improve processes by removing inefficiencies and addressing non-compliance. In addition to process mining, visualization techniques have also been applied to visualizing work items for an overall business process [9]. Leoni et al. incorporated visualizations to support work assignment in process-aware information systems. The visualizations enable users to better select work items to ensure the performance of the overall business process [9].

3. Probing Dynamic Business Processes

In collaboration with our partner, we have developed the following set of questions that must be effectively addressed in order to achieve a successful dynamic business process:

- What are overall trends? Is the overall process on schedule?
- What are the detailed trends for individual programs within the manufacturing and production process? (A program is the production of a particular system, usually distinct from other systems that may also be produced during the production process.) What's the cost-schedule performance over time?
- What is the capacity of each part of the operation, including excess capacity?
- Is something going wrong and where (including things that have not been fully recognized yet)?
- Is resource re-allocation needed, and where?
- If there is a problem, what expertise should be deployed to most quickly solve it?

To have any chance of answering these questions, the production executives must have continuously available the latest information on the manufacturing process at the program and component levels. To this end, our partner has developed a "dashboard" that accepts data from all stages of the process for all programs. The dashboard system does a summary analysis to determine where the process, at any stage, stands with respect to a few simple benchmarks. Importantly, production managers can provide annotations and comments at any point in the process. They can also communicate with each other or the executives via the dashboard. The dashboard displays comparative information on the production process for all components and programs. It shows some information on trends and when a component is falling behind schedule according to a simple milestone analysis. It has some drill-down capability so that the executive can get more information about why a trend is occurring.

Although this sort of analysis is necessary, it is not sufficient for enabling a dynamic business process without interruptions or missed deadlines. For one thing, disruptions in different parts of the organization may affect downstream production in unexpected ways. Further, there may be no direct line of communication between the parts of the organization that will be affected (although there could be informal communication). Finally, since unexpected disruptions will arise, it is only possible to know of them after the fact and often with incomplete information as to cause. Thus it is difficult to put into place predictive or even prospective monitoring and response, nor to pursue some of the questions above such as those having to do with re-allocating resources and expertise as an event is unfolding. In the next sections, we discuss how visual analytics can be applied to address these issues

and produce more successful outcomes for dynamic business processes.

4. Visual Analytics Framework

Visual analytics is the science of analytic reasoning facilitated by interactive visual interfaces [10]. It is meant to support exploratory analysis leading to discoveries since it is frequently applied to complex real world problems with large amounts of data. These problems are often open-ended with no clear path to solutions [11], and, as a result, it is often unclear what pertinent knowledge the data may contain [11]. Thus the analysis needs to be exploratory to support discovery of hidden relations, patterns, and trends. Once discovered, these aspects should be investigated in detail, by referring back to original data and gathering additional evidence to confirm or refute hypotheses that are formed. A comprehensive visual analytics approach would support all these aspects. In this paper we apply such a comprehensive approach, employing visual analytics methods for both automated analyses and user-led exploration through the interactive visual interface.

Our first step is to embed the dashboard system in a visual analytics framework. This is depicted schematically in Figure 2. The repository of data sources (lower left) will inevitably contain both structured and unstructured data. Since a range of sources will be involved, the repository will be heterogeneous. Since useful information could be found anywhere (including perhaps outside the business) and, once found, should be used, some of these sources could be incomplete or fragmentary. Analytic tools must thus be flexible, able to handle different kinds of data, including incomplete data, yet producing results that contribute to a "common picture" that will be meaningful to the user.

Inevitably some of the data in a business process will be textual, containing comments, annotations, descriptions, reports, human communications, and so on. These texts will tend to be partly or mostly unstructured. Therefore we apply methods we have developed for extracting topics and topic-based events over time from unstructured texts [12]. These results are combined with named entity extraction for people's names, dates including future dates, and locations (if desired). The analysis can be applied, with approximations, either to streaming data [13] with results in a couple of minutes or less, or more slowly and accurately to histories over a period of time. These methods are quite flexible and have been applied to many different types of texts including reports, research papers, patent data, twitter streams, online news, and customer messages [1,12,14]. To these textual analyses, we add methods for exploring categorical data and numerical data, including finding trends over time. The categorical methods, for example, reveal relations between multiple categories at selected times or time ranges, which are related to specific events in the topical event analysis. For all these methods plus the embedded dashboard, we develop a new interactive visual interface, as described in section 6.

The analytic methods apply not only to the data repository but also to the business processes embedded in the middle of Figure 2 and, indeed, even to the model repository at the lower right. The business processes produce new information themselves and thus must be included to produce a most useful and comprehensive common picture. (This is indicated by thickening red lines that cross the middle of the figure.) In fact, as we have seen above, it is quite important to analyze information created during the business process since that will produce direct knowledge of the capabilities and capacities of different parts of the process as well as signal when things are going wrong and what parts of the process are affected.

A general definition of a physical event is "a meaningful occurrence in space and time" [15]. For the purposes of this framework, we modify this definition to focus on "process events", which we define as "meaningful occurrences in time with respect to the process". In both cases, time is central. For this framework, bursts of activity over a relatively short time scale, unexpected trends or relationships (that appear in a short time span), or outliers could all qualify as process events. We have shown that events can be automatically found and organized (including putting them in a hierarchical structure) for social media such as Twitter [16]. This work, involving topic modeling, also demonstrates how users can quickly attach meanings to the events. The hierarchical structure becomes significant when one deals with complex data for which one can have many events. As data grows in size and comprehensiveness, it will tend to get complex in this way. Certainly the complex manufacturing and production processes considered here will benefit from hierarchical structuring, which will give them high level meaning and make the overall processes easier to understand.

5. Virtual Sensors

We now have the mechanisms we need to build the effective framework illustrated in Figure 2. The main

idea is to instrument the whole business process (acquisition, production, deployment, etc.) with "virtual sensors". These sensors are flexible so that they can access any data stream (e.g., unstructured text, categorical data, numerical data, model outputs) anywhere, such as in the data repository, in the business processes themselves, or in the models. Typically sensors are placed at a set of general points that are effective for monitoring many different types of production. Then they are placed at additional points that are useful for a specific type of business process. Finally they can quickly be placed at other points by the manager, often to understand some abnormality in the process. This set of networked sensors then report to the visual analytics layer at the top of Figure 2. This layer provides a set of "sensemaking" capabilities to explore, understand, reason with, and test out hypotheses w.r.t. the sensor results. Since these results are expressed in a common language of process events, they can be displayed together and manipulated in an interactive visualization, though some details with respect to the events will be different depending on the underlying data or process. At any point, the manager can select one or more events and get at the underlying data.
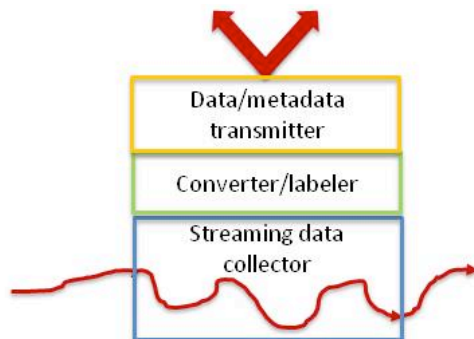

Figure 3. Virtual sensor structure

The structure of the virtual sensor is depicted in Figure 3. The bottom layer grabs data directly from the business process stream. In the case of the business partner mentioned above, the dashboard already collection points that can be used to insert the sensors at several places. The middle layer converts these data to a standard internal format recognized by the analysis framework in Figure 2. It also attaches whatever metadata or labels are available (time range and time stamps, data descriptors, data units, etc.). The top layer transmits the data and metadata to the analysis framework.

Full analysis is performed by an analytics engine in the analysis framework, depending on the type of data. For instance, the analytics engine applies topic modeling and topic-based event extraction to unstructured text messages [12]. It applies temporal signal processing and anomaly detection to streams of numerical data to pull out both events and anomalies. It performs categorical and multidimensional analysis to categorical or attribute data. It performs relational and other statistical analyses to connect events across different types of data and to find similar signatures. These analyses are automated and all results are in a common event-based format regardless of the original input format (using the process event definition in Sec. 4). Thus, for example, alignments of events from the text messages with events from categorical or numerical data streams would be readily spotted in the embedded visual interface described in the next section and then delved for deeper relationships. The common event-based format has the capacity for depicting additional "decorations" of the events stemming from details of the different types of original streaming data.
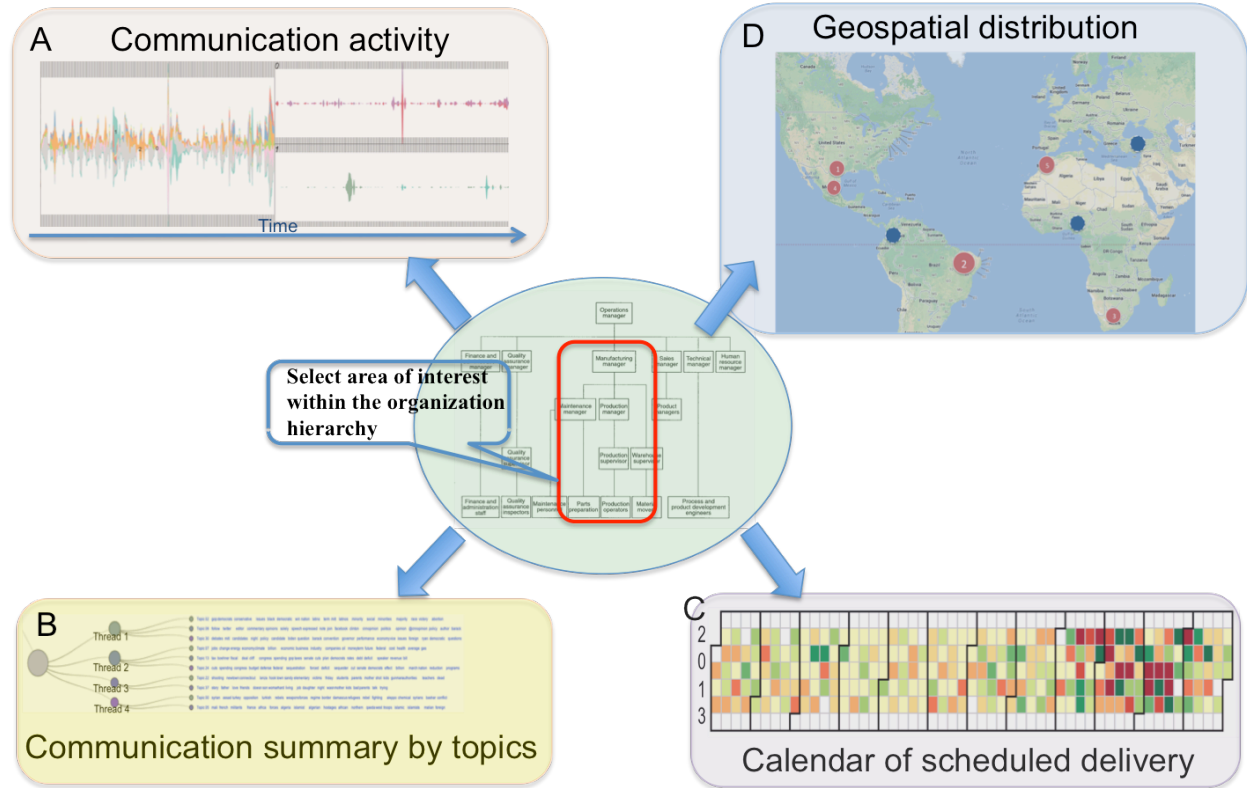
6. The Embedded Visual Interface

The visual interface is designed to support sensemaking in the business processes. More specifically, the visual interface incorporates 4 different views that highlight temporal, geospatial, and topical summaries using both the information available a priori (such as knowledge of the organization hierarchy) and information collected in real time by the virtual sensors.

At the heart of the visual interface is the automatic and real-time sensory data capturer. Our framework is set up to run on a single user's computer, and captures information flows to and from that user around office documents, calendars, emails, organizational charts, and Web pages, etc. Depending on the user analysis needs, the visual interface utilizes various meta-information that comes along with the virtual sensor results. In particular, it collects information about which documents, emails, and web pages were read for that user, and how long they were open. It also collects document metadata, such as information about the senders and recipients of email messages.

All this heterogeneous data is stored in a unified NoSQL sensory data repository for future analysis. On the fact collection level, tools like UpLib [17], can be utilized to extract information from and about each data input, including its title and authors, its text, the people and other entities that it mentions, its paragraphs and its images. All of the captured information is indexed and grouped with its related documents. On the event level, our signal processing then performs event analysis, time series analysis, clustering, and narrative reconstruction based on the

Figure 4. Embedded visual interface with labels to indicate functions.

collected organizational facts. All this is then interactively presented to the user through the embedded visualization interface.

6.1 Visual Components

Instead of presenting the diverse analyzed event structures through a keyword search interface, our framework embeds the investigative retrieval cues (e.g., who, when, where, what) into a coordinated multilevel visualization system.

At a high level, our visual interface encodes the four cues with a set of four visualizations, each of which presents a particular aspect of the organizational and business process activity. To provide a lower level detailed view, our interface also presents a visualization that integrates related activity information for a single worker or a sensor. Using this multi-level structure, the interface helps users to cohesively find the specific details they need.

6.1.1 Organizational Hierarchy View
To give an overview of the complex operational environment, we designed an Organizational Hierarchy view that gives a cohesive overview while encapsulating organizational knowledge. Therefore, a

user can start the analysis process by recalling event activities beginning from any retrieval cue that they remember (e.g., the department that mostly likely to initiate the process), or they can focus on events or cross-organization communication links that the automated analysis has identified as important. (See the application example in Sec. 7 for an illustration of the latter.) Interaction in any view cause updates in coordinated views so that the maximum associated information is provided permitting the user to converge on the desired conclusion quickly.

6.1.2 The Where: Geospatial view.
In the geospatial view, locations involved in the overall business process are highlighted in an interactive map. The importance of the location is represented by the size of the circles and could be determined by user-selected criteria. The numbering on the circle indicates the sequence of the manufacturing or delivering process. The color of the circles is used to denote main/alternative supply or manufacturing locations. The map supports standard user interactions such as zooming and panning. The map is coordinated with other views in that selections within the map will filter to events related to the chosen locations.

6.1.3 The When: Activity Heatmap View

To facilitate multi-scale temporal analysis, we developed an Activity Heatmap view that allows analysts to monitor events in both retrospective and future time frames. This view shows how a business process unfolds over time. It presents both the number of organizational activities (e.g. message exchange) that have interacted with business process, and the types of that business and its key personnel. This view is created as an interactive calendar for ease of interaction and it shows the temporal trends and patterns of organizational activities. Each cell in Figure 4C represents aggregated mentions of that date in all the existing business processes; dates that have been mentioned more frequently appear in a darker shade of blue. For example, we can quickly see that there are multiple delivery events that happen on the July 30th of this year.

Using this view, manager can highlight a time range to select a subset of data (e.g., an hour to a day or months) to be analyzed retrospectively. Besides showing general trends and patterns, the temporal view also allows the user to drill down into time periods. When the user selects a time period on the horizontal axis in the center of the view, our visual interface will zooms into that period of time and present the most relevant business information.

6.1.4 The Who: Communication Activities between participating parties

To help corporate managers and decision makers efficiently retrieve specific events of interest, we designed the Communication activities view to aggregate both the documents and the people that a user has interacted with during a particular period of time. Like Lee *et al* [18], our activities view allows the user to filter and sort information based on automatically-extracted data facets, including different communication types and format. The user can visually depict the relationships between the extracted organizational facts with selected business projects. For example, the user can choose to see or hide activities with email, with office documents, with Web pages, or with people. Each of these facets can be turned on or off by pressing an associated button.

In order to fit the activity information into a reasonable guided exploration, our visual interface sorts events by importance, and displays the most important documents at the top and with the most salient presentation by computing the importance value centered around the relevance to its core business process. As described in Sec. 7, the organizational hierarchy view is also important in highlighting which events should be followed. In addition, to enable fast

exploration, a summarized information panel (see Figure 4(D) right) is shown when the mouse hovers over a visual element representing an activity. Like the Document Card [19], this panel includes a readable thumbnail and aggregated information about that visual element. If the user needs more details, the user can double click on the visual element to bring up a specific detail view.

6.1.1.4 The What: Communication Summary by Topics

To support the analysis of communications among different teams and parties during the overall business process, we designed the Communication Summary view, which provides an overview and permits analysis based on streaming or accumulated communication data. The analysis enables users to discover pressing issues that need to be addressed in order to keep the entire business process on track.

We assume most of the communication data are unstructured texts sent between communicating parties or attached as annotation to the stages of the business process. The virtual sensors collect all communications and put all the received communications into a database. The attribute in the database contains sender, receiver, time stamp, communication content, and other available information. All collected texts are then summarized into semantically meaningful topics using state-of-the-art topic models [20] as shown in Figure 4B. The topic summary presents individual topics to summarize what the communications are about. Coupling with the temporal trend of the topics shown in figure 4A, one can quickly discover when a burst of communications has occurred and what issues were discussed. Other forms of data that are collected and analyzed are discussed in Sec. 7.

We now have complete support for the sensemaking process depicted at the top of Figure. 2. The foraging stage is supported by the distribution of virtual sensors and their outputs. The encoding is supported by the conversion to a common event-based format in the analytics engine. The reasoning (including hypothesis-building and testing) is supported by the exploratory visual interface, which includes new placement of virtual sensors, and review of original data. All these stages are iterative and interconnected through the visual interface. In the next section we will discuss how this sensemaking can be applied.

7. An Application Example

In this section, we will build on the example given in the Introduction and illustrated in Figure 1. A real

world manufacturing operation of any size will have a bigger organizational structure than depicted in Figure 1 and will have a layer of senior management ("C-level": CEO, CIO, CTO, CFO, etc.).above the production manager. Because of complexity, the changing nature of dynamic JIT manufacturing and production, and the fact that signals of interest can come from unexpected places, our approach must be exploratory, bringing information from disparate sources into a common picture for evaluation and action. However, it is also data-driven, and since the initial data analytics especially are automated, there will be some tension between automated and user-directed processes. The visual analytics framework and the interactive visual interface are designed to minimize this tension and to make these processes work together in support of effective user actions.

A main aspect of the data-driven analysis is to identify process events (defined in Sec. 4) and to arrange them so that they tell a meaningful story when interpreted by the user. In the case of the dashboard used by our business partner, process events can be derived from numerical and categorical data collected at each stage of the dynamic manufacturing process, status reports indicating whether each stage or component of the process is on, behind, or ahead of schedule at that point, and individual annotation including comments by managers or engineers for that stage. There is also a mechanism for communication among managers, engineers, operators, and supervisors across the organization. The dashboard thus provides comprehensive data that we use in the above implementation of the visual analytics framework. However, though it provides comprehensive data, we have heard from the executive in charge that the dashboard is not sufficient to fulfill the need for on-time dynamic processing with minimum disruption—too many real or potentials points for breakdown can

Figure 5. Illustration of virtual sensors deployed in manufacturing process. Four virtual sensors are labeled V.S. in the diagram.

be overlooked; likewise too many opportunities for improvement and efficiency can be missed.

Thus we instrument the dashboard data and annotation collection points throughout the manufacturing process with virtual sensors, as illustrated in Figure 5. Process events are then identified in multiple ways. One way of doing so is by identifying bursts of activity over relatively short time scales in the topics derived from the textual annotations and communications. There could also be a trend in the numerical data that is unexpected compared to the normal manufacturing process. In the example here of not having a particular part when needed farther along in the manufacturing process, the burst of activity centers around one or more topics derived from both the annotations and the communications among workers indicating the part, the component system it goes into, and the supplier. An appropriate scale for this burst of activity is a few days. (In other words, text messages on the same topic accumulate in a burst over a few days.) Named entity extraction is also applied to the texts to indicate what sections of the organization (and individuals) are communicating, the names of parts, systems, and suppliers, and dates mentioned in the texts (either past or future). In addition, analysis of the inventory signal can indicate a drop in the supply of the crucial part over time that becomes significant enough to generate another event. The importance of the process events from the topic analysis is further raised because communication occurs across organizational boundaries between the parts manager, the quality assurance manager, and the technical manager (and their teams). The organizational knowledge of what is and is not unusual communication across the organizational hierarchy is encoded in the visual analytics framework. We have shown that such temporal bursts of topical activity can be automatically identified and that they can be connected to real events with high probability. We have successfully done this event analysis for streaming Twitter data (e.g., events during the development of the Occupy Wall Street movement over a year's time) [16] and collections of research proposal abstracts and full-text research papers [21] where the events are often connected with new programs started by funding agencies. These quite different text collections and quite different time scales demonstrate that the topical event-identification methods are general.

Up to this point the analysis is all automatic, but the process event results based on both statistical techniques (e.g., signal processing) and process and organization knowledge permits highlighting of events, anywhere in the process or organizational structure,
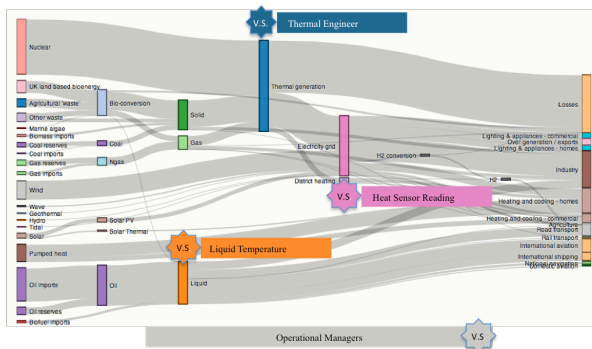
that are likely to be of interest to the high-level manager. The manager starts to interact with these results via the visual interface in Figure 4. Through the central organization hierarchy view, he sees highlighted links between the parts group and the quality manager and technical management groups that he knows are unusual. A check of the communication activity view shows when these bursts of communication occurred. The problem has not yet shown up in the calendar view (i.e., the affected process components are not yet falling behind schedule). He can then dig down to the actual texts of topical messages and can filter for geographical distributions of both suppliers and the finished systems, which will be displayed in the geographical view. He can set a new virtual sensor to watch the inventory stream (and to collect historical data). If there is a more complex problem, there may be several relevant topics. In this case, the hierarchical topic view (communication summary view) is useful. The view shows which topics are related and can be grouped and also permits re-grouping by the manager. The data associated with the topics, including text messages and other data, are reorganized as well. In summary, the manager can quickly find issues that should concern him and then effectively analyze them, moving towards making key decisions.

The visual analytics framework has other useful attributes. The process of investigation by the manager that leads to a conclusion and a plan of action is itself captured in terms of selected process events, topics, the manager's own annotations, and interactions with the views in the visual interface. We have found in our studies of bank fraud analysts that an interface such as this can reveal high level strategies and its own meaningful story of how an expert reaches a decision [22]. This story can be shared with other managers, giving a rich argument for a course of action. In addition, like the fraud analysts, the managers in this dynamic manufacturing process have expertise developed through experience that is not easy to share. However in the case of our partner, the problem is exacerbated because, due to the nature of the business, there is some flow of managers into and out of their positions and thus a knowledge and experience gap. Use of the visual interface and underlying analytics by

experts for a variety of situations can be captured and made available as training modules for these new managers, giving them insights that would be hard to obtain without long experience.

8. Conclusion and Next Steps

We have presented a visual analytics framework for handling complex business processes. It is applied to dynamic just in time manufacturing, but it is applicable to a range of agile business processes.

In order to organize data that can come from anywhere in the complex business process and can be in a variety of types, we have introduced the concepts of virtual sensors and process events. The virtual sensors permit collection from the streams of data at any point in the process and transmission of the data in a form ready to be analyzed by the central analytics engine. Process events provide a uniform expression of data of different types in a form that can be automatically prioritized and that is readily meaningful to the users. Through the visual interface, the user can place the virtual sensors, interact with and group the process events, and delve into the details of the process at any point. The visual interface provides a multiview investigative environment for sensemaking and decisive action by the user.

We have shown that with this visual analytics environment, the user can answer the major questions that must be addressed in order to achieve a successful dynamic production environment. These include finding overall trends and then detailed specific trends in the production process; determining, as early as possible, what may be going wrong and where in the process; determining what the capacities of various parts of the process are and what resource re-allocation is needed; and determining what expertise should be deployed to most quickly solve a problem.

We are now working on detailed set up and evaluation of this framework. This will lead to improvements and to the reporting of concrete case studies. However, the response to the initial design and implementation is positive.

9. Bibliography

[1] Stephen A. Ruffa, (2008). Going Lean: How the Best Companies Apply Lean Manufacturing Principles to Shatter Uncertainty, Drive Innovation, and Maximize Profits, AMACOM (American Management Association)

[2] Bobrik, R., Reichert, M., Bauer, T.: View-based process visualization. In: Proceed- ings of the 5th International Conference on Business Process Management BPM 2007. Volume 4714 of LNCS., Springer (2007) 88–95.

[3] Luttighuis, P., Lankhorst, M., Wetering, R., Bal, R., Berg, H.: Visualising business processes. Computer Languages 27 (2001) 39–59.

[4] Streit, A., Pham, B., Brown, R.: Visualization support for managing large business process specifications. In:

Proceedings of the 3rd International Conference on Business Process Management BPM 2005. Volume 3649 of LNCS., Springer (2005) 205–219.

[5] Wright, W.: Business Visualization Adds Value. IEEE Computer Graphics and Applications 18 (1998) 39.

[6] B. Azvine, D. D. Nauck, C. Ho, K. Broszat, and J. Lim. 2006. Intelligent process analytics for CRM. *BT Technology Journal* 24, 1 (January 2006), 60-69.

[7] Hansen, G.: Automated Business Process Reengineering: Using the Power of Visual Simulation Strategies to Improve Performance and Profit. Prentice-Hall, Engle-wood Cliffs (1997).

[8] W.M.P. van der Aalst, M. de Leoni, A.H.M. ter Hofstede. Process Mining and Visual Analytics: Breathing Life into Business Process Models. *BPM Center Report BPM-11-16, BPMcenter.org*, 2011.

[9] Massimiliano Leoni, W. M. Aalst, and A. H. Hofstede. 2008. Visual Support for Work Assignment in Process-Aware Information Systems. In *Proceedings of the 6th International Conference on Business Process Management* (BPM '08), Marlon Dumas, Manfred Reichert, and Ming-Chien Shan (Eds.). Springer-Verlag, Berlin, Heidelberg, 67-83.

[10] Thomas, J.J. and Cook, K.A. (Eds.), Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE Press, 2005.

[11] Daniel A. Keim: Solving Problems with Visual Analytics: The Role of Visualization and Analytics in Exploring Big Data. BTW 2013: 17-18

[12] Dou, W., Wang, X; Skau, D., Ribarsky, W., Zhou, M.X., "LeadLine: Interactive visual analysis of text data through event identification and exploration," *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* , vol., no., pp.93,102, 14-19 Oct. 2012.

[13] Kraft, T., Wang, X., Delawder, J.,Dou. W., Yu, Li.,Ribarsky, W., "Less After the Fact: Investigative Visual Analysis of Events from Streaming Twitter", in Submission to LDAV, 2013.

[14] Wang, X., Dou, W., Ma, Z., Villalobos, J., Chen, Y., Kraft, T. and Ribarsky, W. I-SI: Scalable architecture of analyzing latent topical-level information from social media data. *EuroVis 2012 Computer Graphics Forum*.

[15] William Ribarsky, Eric Sauda, Jeffrey Balmer, and Zachary Wartell. The Whole Story: Building the Computer History of a Place. Hawaii International Conference on Systems Science (HICSS 2012).

[16] Dou, W., Wang, X., Ma, Z. and Ribarsky, W. Discover Diamonds-in-the-Rough using Interactive Visual Analytics System: Tweets as a Collective Diary of the Occupy Movement. The 7th International AAAI Conference on Weblogs and Social Media (ICWSM), Workshop on Social Media Visualization (SOCMVIS), 2013.

[17] W. Janssen. The uplib personal digital library system. In JCDL '05: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, pages 410–410, New York, NY, USA, 2005. ACM.

[18] B. Lee, G. Smith, G. G. Robertson, M. Czerwinski, and D. S. Tan. Facetlens: exposing trends and relationships to support sensemaking within faceted datasets. In CHI '09: Proceedings of the 27th International Conference On Human Factors In Computing Systems, pages 1293–1302, New York, NY, USA, 2009. ACM.

[19] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In CHI '04: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 415–422, New York, NY, USA, 2004. ACM.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan. "*Latent dirichlet allocation*". J. Mach. Learn. Res., 3:993–1022, March 2003

[21] Dou, W., Wang, X., Chang, R. and Ribarsky, W.. ParallelTopics: A Probabilistic Approach to Exploring Document Collections, IEEE VAST 2011.

[22] Chang, R.; Ghoniem, M.; Kosara, R.; Ribarsky, W.; Jing Yang; Suma, E.; Ziemkiewicz, C.; Kern, D.; Sudjianto, A., "WireVis: Visualization of Categorical, Time-Varying Data From Financial Transactions," *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* , vol., no., pp.155,162, Oct. 30 2007-Nov. 1 2007 doi: 10.1109/VAST.2007.4389009

Visual Analysis of Situationally Aware Building Evacuations

Jack Guest and Todd Eaglin and Kalpathi Subramanian and William Ribarsky

Charlotte Visualization Center, Department of Computer Science
The University of North Carolina at Charlotte, Charlotte, NC 28269, USA

ABSTRACT

Rapid evacuation of large urban structures (campus buildings, arenas, stadiums, etc.) is a complex operation and of prime interest to emergency responders and planners. Although there is a considerable body of work in evacuation algorithms and methods, most of these are impractical to use in real-world scenarios (non real-time, for instance) or have difficulty handling scenarios with dynamically changing conditions. Our goal in this work is towards developing computer visualizations and real-time visual analytic tools for building evacuations, in order to provide situational awareness and decision support to first responders and emergency planners. We have augmented traditional evacuation algorithms in the following important ways, (1) facilitate real-time complex user interaction with first responder teams, as information is received during an emergency situation, (2) visual reporting tools for spatial occupancy, temporal cues, and procedural recommendations are provided automatically and at adjustable levels, and (3) multi-scale building models, heuristic evacuation models, and unique graph manipulation techniques for producing near real-time situational awareness. We describe our system, methods and their application using campus buildings as an example. We also report the results of evaluating our system in collaboration with our campus police and safety personnel, via a table-top exercise consisting of 3 different scenarios, and their resulting assessment of the system.

Keywords: Evacuation, visual analysis, situational awareness, emergency response

1. INTRODUCTION

In any emergency or incident involving large urban structures (arenas, stadiums, college campuses), the safety of the occupants is of paramount importance. Thus every building has a set of passive safety features (sprinkler systems, fire extinguishers) and evacuation plans or routing maps posted at various points within the building. However, in recent years, more active approaches to studying evacuations from large urban structures or street networks have gained importance, and have been based on mathematical and algorithmic approaches. These methods study the problem of evacuating the occupants from the structure in the shortest possible time, also known as the egress time. What is lacking in these methods is the ability to handle real-world scenarios involving *large and complex structures that may involve multiple buildings,* and more important, the *ability to react in real-time to dynamic changes in the scenario,* such as blocked stairwells or hallways, and provide useful recommendations to responders who can mitigate damage, injury, or loss of life. For a commander overseeing the evacuation, the ability to clearly and unambiguously understand the dynamically changing situation is very important, being useful for optimal allocation of limited resources and personnel, and for making other timely decisions.

In this work, we address these challenges with the primary goal of responding to dynamic events in real-time during an emergency. We begin with an existing heuristic based route planning algorithm,[1] make certain modifications, and embed this as part of a visual analytic system that permits complex real-time interactions during the event. This then permits dynamic changes in the building accessibility to be incorporated, and alternative routing assessed in real-time. To accomplish this, we employ an LOD graph representation of the underlying urban structures that permits real-time recommendations to be presented to the emergency planners

---

Further author information: (Send correspondence to Kalpathi Subramanian)
Kalpathi Subramanian: E-mail: krs@uncc.edu, Telephone: 1 704 687 8579
Jack Guest: E-mail: jguest@uncc.edu
Todd Eaglin: E-mail: teaglin@uncc.edu
William Ribarsky: E-mail: ribarsky@uncc.edu

and responders for informed decision-making. This is combined with realistic models for building occupancy and traffic flow. Interactive visual analytic tools permit quick exploration of possible impacts of the current situation, and the increased situational awareness for emergency commanders and responders permit more optimal use of scarce resources, for instance, dispatching responders to areas of need in congested sections of a building, handle casualties, etc.

We begin with a review of current work on evacuation algorithms, followed by a description of the techniques we use to create our 3D models and building networks. This is followed by a description of our modified capacity based route planner, followed by a detailed description of the major features of our interactive visual analytic system and its reporting and recommender functions. We describe the use of this system on typical scenarios using our campus buildings as a test case. We evaluated our system and its performance via a table-top exercise, consisting of 3 scenarios: gas leak in building, active shooter and an explosion in an adjacent structure. Performed in collaboration with our campus police and safety personnel, we report on their assessment and recommendations.

2. PREVIOUS WORK

The study of computer based evacuation modeling has evolved with both mathematical and algorithmic approaches. These approaches are categorized as *macroscopic* and *microscopic*.

2.1 Macroscopic Models

Macroscopic approaches focus primarily on minimizing egress time. The evacuees are treated as a unit or group of units and moved from source to destination. Interaction between these units are defined by capacity/congestion rules. Linear programming methods based on Network Flow are one of the earliest approaches, yielding optimal solutions but at high algorithmic cost, and often impractical to use in real-world scenarios. The Maximum Flow Problem is one network solution that is implemented with costs as high as $O(n^3)$[1] yielding a best known cost of $O(nm \lg n^2/m)$. Naoyuki Kamiyama et al.[2] apply two initial conditions to the network flow problem. For each vertex the sum of transit times of arcs on any path takes the same value, and for each vertex the minimum cut is determined by the arcs incident to it whose tails are reachable. These assumptions resulted in a 2d grid network and they solved the transhipment problem in $O(n \lg n)$ time. Shekhar and Yoo[3] compare models relevant to the study of nearest neighbor paths. Also Kim, et.al[4] discuss contraflow in reconfigured networks for emergency route planning. This work gives insight into reconfiguring a network that is damaged and therefore relevant to our work.

Our model is based on a heuristic approach, the Capacity Constrained Route Planner algorithm proposed by Shekhar et al.[1] This approach attempts to find lower cost algorithmic solutions at the expense of the detail of each evacuee's egress. These approaches are interesting because they can be evaluated quickly from a user perspective as the network-flow problem is reduced to a generalized shortest path problem. The inputs are a graph structure and evacuee populations. The output is a route plan with start times, and a location matrix for each evacuee for each defined time segment. We have adapted this algorithm to meet the more challenging requirements for real-time decision making in large urban environments, as well as the ability to inject situational changes during an emergency.

2.2 Microscopic models

Microscopic approaches use agent based modeling, where each evacuee is governed by unique rules of behavior. Interaction between individuals and their environment are defined based on spatial and social parameters. These approaches do not exclude the deterministic path or goal that is the foundation of the route planner, but they rely on behavior rules applied to each evacuee to overcome the "lack of detail" inherent in network flow or heuristic planners. These methods are also referred to as Agent Based Models(ABM). The goal remains to minimize the time to evacuate individuals to safe zones.

Because of the adaptive parameterization of ABM, neural networks and fuzzy-logic[5] approaches can be adapted to building evacuation simulation. Discrete Particle Swarm Optimization,[6] use of velocity and spatially based rules of interaction[7] are other approaches. A general description of these approaches can be found in.[8]

The most important aspect of ABMs are the rules applied to each evacuee. Castle et. al.[8] describe a detailed list of rules and attributes. We believe our system captures sufficient detail using a congestion model. However, because agents can be trained, we are evaluating Q-Learning/SARSA techniques to add human factors to future work.

2.3 Visual Analytics

Visual analytics involves effectively combining interactive visual displays with computational transformation, processing and filtering of large data.[9] One focus of visual analytics is real-world problems involving situationally-aware decision support. Andrienko et al.[10] focused on automatic generation of transportation schedules for evacuation from a disaster zone; visual analytic tools were used for verification by human experts. Campbell and Weaver[11] used interactive visualization tools for hospital evacuation scenarios that involved training first responders. The work of Kim et al.[12,13] focused on use of mobile devices for situationally aware emergency response and training, and thus their approach is similar to our work. They demonstrated their system with an evacuation simulation of the Rhode Island club fire of 2003. Our system is considerably more general and is scalable to large urban buildings and provides the means to interrupt the simulation based on new situational information or dynamic changes.

The use of linked views is an important technique to connect different representations of information within a single visualization, with applications specific to urban structures.[14,15] Sensor networks are used within buildings to help create interactive visual analytics. The work of Ivanov et. al[16] is not specific to building evacuations, however their use of data graphs to interact with maps provided inspiration for our cross platform bar chart displays that interact with our simulation. Visual analytics tools such as Jigsaw[17] are available for integrating process output data from an application.

3. METHODS
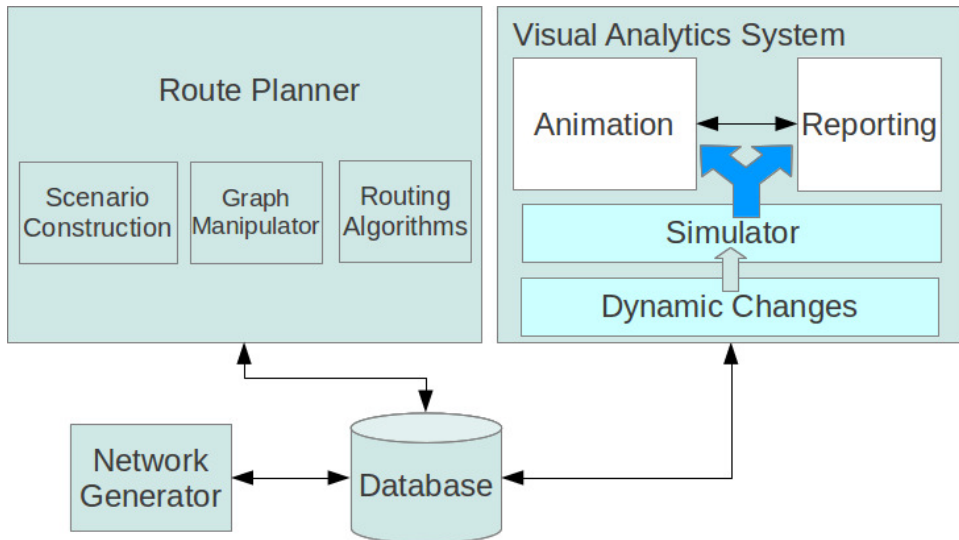
Figure 1. Visual Analytic System Architecture. Route planner involves preprocessing and routing calculations that serve to initialize the system, with processed results stored in a database. The visual analytic system uses visual abstraction and scalable representations that permit real-time interaction, injection of dynamic situational changes and visual analysis.

Fig. 1 illustrates the major components of our visual analytic system for situationally aware evacuations. There are two major components to the system. Functions in the Route Planner involve a significant amount of apriori processing and initialization. The visual analytic system is a highly interactive system that is user driven and can inject and respond to dynamic changes during evacuations. It also consists of reporting and visual analysis functions that can assist an emergency planner on exploring different scenarios, as to the use and deployment of resources, dispatch responders, effect of rerouting occupants, etc.

In our earlier work,[18] we described a semi-automatic system that constructed a *building graph*, incorporating key elements of a georeferenced urban structure critical to evacuations, such as hallways, stairways, elevators and entrances/exits. This building graph generator is used to process urban structures and is stored in a PostgreSQL database. We have successfully processed over 70 buildings of our campus using the graph generator.

We next describe the main components of our visual analytic system.

3.1 Route Planner

Our route planner provides facilities for loading evacuation objects, that are combinations of urban structures, pathways, streets etc. Evacuations are built as combinations of structural objects and route planner objects and saved in the database. The Route Planner consists of the following components:

3.1.1 Scenario Construction

Scenario construction includes loading single or multi-building evacuation objects, selecting a route planning algorithm (currently limited to our modified capacity constrained route planner,[1] and setting visualization modes and evacuation parameters, such as building capacity, egress width, evacuee speed and density, and stairway up resistance.

(a)   (b)

Figure 2. Building graph Simplification. Removing redundant nodes. (a) a section of a building floor with labeled building elements, (b) simplified graph.

(a)   (b)

Figure 3. Building graph simplification to a zone graph. (a) building graph of a campus building, (b) transformation to zone graph representation. Yellow spheres are nodes and cyan tubes represent edges. The red tubes represent paths to exits

3.1.2 LOD Graph Construction/Manipulation

All computed egress paths are saved in the database as node to node connections. Nodes are defined based on their function, and can be rooms, hallways, stairways, elevators, and exits. The evacuation application also creates muster points, that represent locations where evacuees are ordered to congregate during an emergency. During the process of extracting centerline points the geometry of hallways is sampled at approximately two foot intervals.

This process is necessary to insure accuracy, particularly at corridor elbows. However, this raises computational issues with extremely large buildings or multi-building graphs, and hinders real-time performance(expensive routing calculations) when dynamic changes need to be accommodated during an event. We address this problem by simplifying the graph with two different levels of detail.

Level 1. Remove Redundant Nodes. In this step, we simplify the original building graph by removing nodes that do not impact routing, for instance, shortest path calculations. The larger sampling rate used for accurate centerline calculations results in nodes that can be removed for use in building routes. Nodes and edges are collapsed in the process. Beginning with any node with three or more edges (or any arbitrary node), edges are followed until a node with 3 or more edges is encountered. This becomes a node of the simplified graph and a new edge is created connecting to the previous node, as can be seen in Fig. 2. The process is continued until all nodes have been visited. Our route planner is executed on these simplified graphs for scenarios in which all original egress paths are available. In our experiments, we see a factor of 5-8 reduction in the number of nodes in the simplified graph.

Level 2. Zone Graphs. For rapid computation of paths when dynamic changes are injected during an event, the simplified graphs can still be large, especially in multi-building evacuations. In these circumstances, we further simplify the building graphs into *zone graphs*, by segmenting the building into evacuation sensitive zones: for instance, stairwells, elevators and exits form the critical elements of any egress path. An example zone graph is illustrated in Fig. 3, involving stairwells, hallways and exits(in red).

To compute the zone graph, we use the precomputed evacuee paths to first associate each node with a zone that is closest to it, using an iterative procedure. In the second pass, the graph connectivity is established by keeping track of the zone of related objects that are encountered in these paths (adjacency lists are maintained). Intermediate nodes(the yellow spheres in Fig. 3(b)) are also identified by paths that cross multiple zones and are further used to complete the graph construction. Zone objects span floors in multi-storey structures, with appropriate floor identification for proper path determination during routing calculations. The number of nodes in the resulting zone graph depends on the number of zones and the number of floors in the building. In our experiments, a further factor of 5-7 reduction in the number of graph nodes was seen.

3.1.3 Routing Algorithms

We have implemented a modified version of the Capacity Constrained Route Planner(CCRP),[1] which is illustrated in Algorithm 1 (See Appendix A). Given a directed graph with node and edge capacities, the algorithm repeatedly computes the shortest path for each evacuee with available capacity. If a path is found, then it assigns as many evacuees as possible through that path, i.e., until the capacity of any node or edge along the path is exceeded. This is followed by moving all the evacuees at that time step. The process repeats until all evacuees have found paths to exit the structure. The final step is to evacuate the remaining evacuees in the building(who already have paths, but not exited the building).

We have augmented the CCRP algorithm by specifying the movement of the evacuees(in addition to finding the paths) at each iteration. Secondly, the algorithm is modified to work with our simplified graphs; in the simplified graphs, weights of the collapsed edges are accumulated and assigned to the new simplified edges. Running the routing algorithm on the simplified graphs makes it more scalable to larger urban structures as well as facilitating dynamic changes to the graph that will require rerouting occupants around blockages or other hazards caused by the emergency event. Finally, although each evacuee has a set average speed (3 ft. per sec.), evacuees cannot exceed the set density threshold. Thus, as congestion builds up, evacuee movements are naturally slowed down. Additional data structures are maintained to make these computations efficient.

3.2 Visual Analytics System

The visual analytic system (Fig. 1) consists of a simulator that accepts user input during an emergency, a 3D interactive animated display of the ongoing evacuation, and reporting and analytic modules. All these views accept direct input and the views are linked to update automatically, resulting in presenting the user with the most current information.
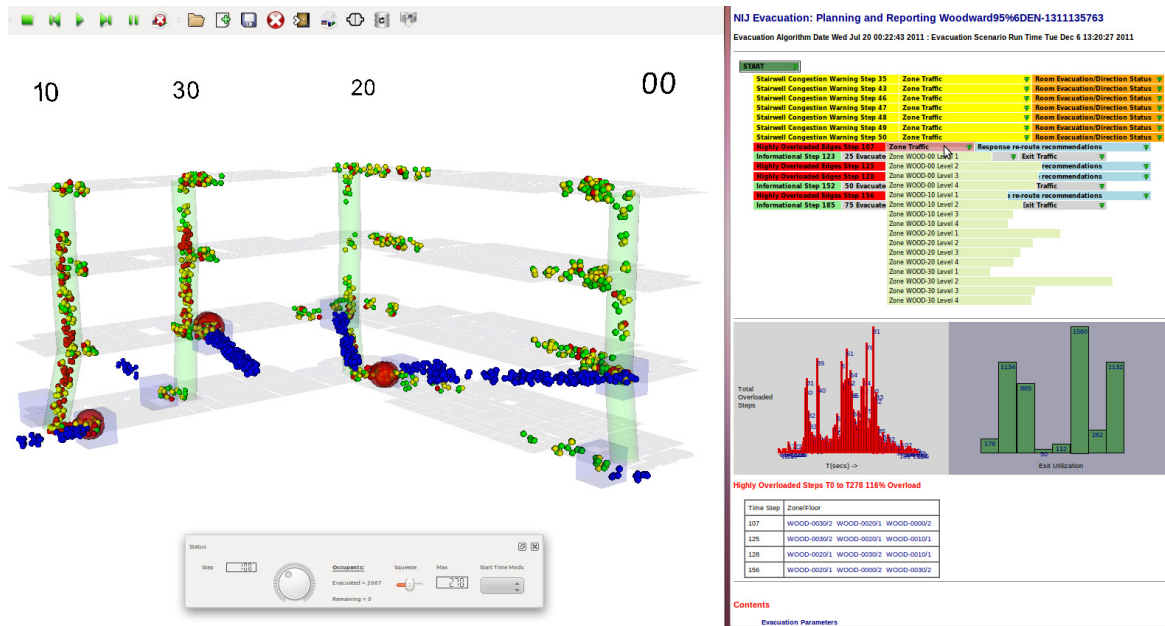
Figure 4. Visualization Design. The upper left panel is the 3D view of the building undergoing an evacuation. Spheres encapsulate evacuee population densities, permitting easy identification of congestion points in the building. Vertical tubes (light green) represent stairways/elevators. Cubes on the first and second floor indicate exits. Lower left indicates the status bar for animation control. The upper right panel is for displaying reports of significant events, that can be drilled down. Lower right panels(bar graphs) show aggregate information on exit occupancy as well as events arranged on a timeline. The report views and the 3D views are linked for immediate updates.

3.2.1 Simulator

The simulator loads evacuation plans that were saved for scenarios under normal static conditions. Since dynamic changes affect only a small part of the structure, a *large amount of preprocessed data can be reused*, contributing to our real-time performance. The simulator accepts user defined dynamic changes (specification of blockages or casualty reports through the 3D display or the report modules) and modifies computed paths, and reroutes occupants. In addition it updates the generated reports and responder recommendations or action plans.

3.2.2 Visualization Design

Fig. 4 presents our interactive visualization system. Almost all of the interaction are via *direct manipulation*. There are three components that make up the design. On the left is a 3D animated view of the urban structure, where the user can load and play evacuation simulations. Evacuees are represented as spheres and colored green, yellow, or red based on low to high congestion. Partially transparent light green tubes represent stairwells, while purple cubes are exits. Blue polygons represent areas that can be occupied. Large red spheres of varying opacity represent edge (capacity) congestion. On the bottom left is the *status tool widget*, that allows moving around in the animation with a slider, indicating current step, evacuee counts and total simulation time. The top right panel is the *significant event* window. The rectangular bars are menus with varying levels of detail of the simulation report. The bottom right panel is a scrolling widget with interactive charts and graphs for interacting with the simulation and visual analysis.

- **Congestion Representation.** Congestion is the primary concern of each evacuee and predictions of future congestion and mitigation is of concern to the emergency response teams. In our system, congestion levels range from green to red (low to high).

- **Temporal Cues.** The color and size of spheres is modified at significant event times. For example sphere size is enlarged when the simulator starts moving evacuees from a source location. The resultant pulsing

in the animation yields important spatial and temporal information to the user about where a new source of traffic will originate and likely areas of future congestion.

- **Details on Demand.** The *Significant Event* window in the reporting tool uses a colorized layered menu so that the visualization of significant information is presented as needed by the user. This allows a maximum amount of reporting while allowing for quick event scanning in the event report. A user sees a limited top level distribution of data unless there is a reason to drill down deeper into the event for details. In the top right of Fig. 4 the user has selected a *Heavy Zone Congestion* item (in red) with its time step. The user can explore further via a mouseover operation. to reveal a bar graph that shows the relative congestion of each zone in the building.

- **Interaction On Linked Views.** The simulation is manipulated by direct interaction over the 3D animation and report views. For instance, a blockage can be introduced via the 3D view, and simulation rerun to generate new (rerouted) paths for impacted evacuees; the report view is updated to reflect the situational change. Similarly the interaction with the reports menus, charts, etc. temporally updates the 3D animation view. All such operations are performed in near real-time as all computation are performed on simplified graphs, promoting. interactive visual analysis.

4. IMPLEMENTATION

Our system is built utilizing open source toolkits on a Linux system, using Python 2.7.[19] 3D rendering is done with the Visualization Toolkit(VTK)[20] and the GUI is produced with QT4.[21] We use a PostgreSQL database with the PostGIS extensions. The database is accessed running on the local machine with direct package calls. The reports section is an HTML/Javascript window inside a QT4 widget. This allows automatic porting to mobile device browsers.

5. EXAMPLE SCENARIO

Here we describe three experimental scenarios to illustrate the use of our system. In this scenario, there are 3500 evacuees. The maximum egress capacity is set to 1 evacuee per cubic foot and navigation speed is set at 3ft/sec (congestion can slow down or halt evacuees during a simulation). The building is loaded to 95 percent capacity.

5.1 Situation 1: No Blockages

Figure 4 is a screen shot of our visual analytic system, loaded with a campus building with no inaccessible areas. Evacuees are represented by spheres, clustered visually using an algorithm which arranges them based on egress width. A small red sphere indicates an evacuee cluster on a congested step. Large red spheres of varying opacity indicate congestion greater than 116 percent of rated capacity. In the timestep shown in Fig. 4, the user has clicked on the reporting panel(upper right), representing a highly loaded event at timestep 108 sec. The user has also rolled over the zone congestion bar to reveal the detailed zone congestion graph. As indicated by the bar at Zone 30 Level 2 this is the most traveled and congested route. The application suggests that a responder be dispatched to this area. As the user rolls over the associated 'Response Reroute Recommendations' menu bar in the list the recommended action can be made visible.

Access for responders can be found by looking at the green exit utilization bars(lower right bar chart of Fig. 4), and choosing a low utilization exit. The zone exits in this scenario that are not utilized are on the 1st level. This type of information can be a powerful dispatch tool for the emergency commander to make an informed decision.

Each bullet in the significant event list (upper right panel in Fig. 4) serves as a visual clue to the overall execution of the scenario. The list covers the entire evacuation. The yellow stairway warnings can be drilled deeper to see which areas are becoming congested. These cues are important for responders to quickly react during the beginning of an event or if a campus lock down has been released. The room evacuation, direction status options can be rolled over to indicate the direction from which the traffic is proceeding, which in turn could result in congestion at a later point.
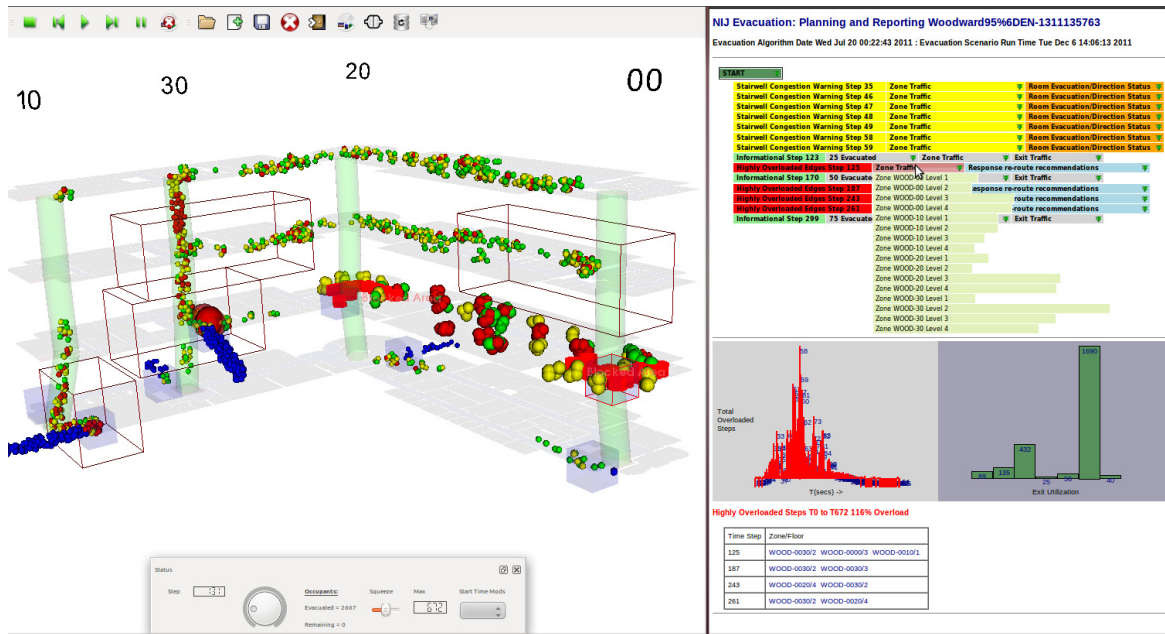
Figure 5. Two blockages have been placed in the building at 45 seconds into an emergency evacuation. (1) Floor 2 at zone 20 stairwell and (2) floor 2 at zone 00 stairwell. Individuals are shown trapped between them by enlarged spheres. Wireframe cubes indicate traffic flow areas, obtained by rolling over the bar chart at the bottom of the report window.
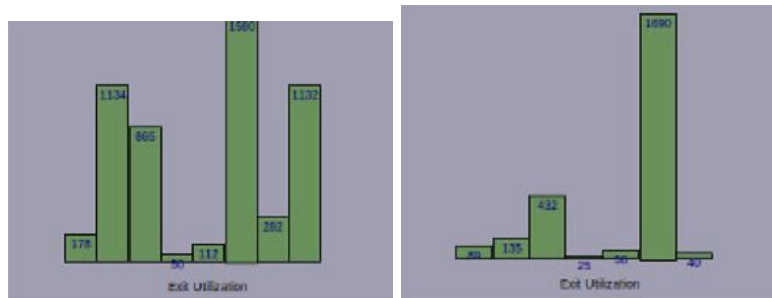

Figure 6. Before and after congestion: Exit utilization.

5.2 Situation 2: Induced Blockages

As shown in Fig. 5, two blockages have been introduced into the scenario of Fig. 4. The blocked areas represented by red squares spread out over several square feet on the second floor at zone 00 and zone 20. In this example, a total of 3100 evacuees were rerouted and all reporting recalculated in 2.8sec. Note that in the control bar at the bottom of the 3D animation view the maximum evacuation time has increased to nearly 7 minutes from less than 3 minutes.

Fig. 6 shows the drastic shifts in the movement of people from a standard evacuation of the building. This example serves to show that evacuation modeling of normal (non-blocked) scenarios is considerably different than when blockages are introduced. In particular, notice the difference in the utilization of the exits. In the blocked case, the congestion occurs earlier and is steeper, resulting in longer times for all evacuees to exit the building.

Mouse rollover on the significant event list view in Fig. 5 shows that the third and fourth floors are getting backed up above the exit at zone 20 floor 2 which is loaded heavily even during a normal scenario. Because the event occurred early there were a number of evacuees occupying the upper floors.

Fig. 7 further contrasts the blocked and non-blocked cases. Here (a,b) illustrates the unblocked case 108 sec. into the evacuation, and (c,d) for the blocked case at 131 sec. We compare the zone traffic via the light green bar charts. The bar charts show total zone traffic from the beginning to the end of the evacuation scenario.
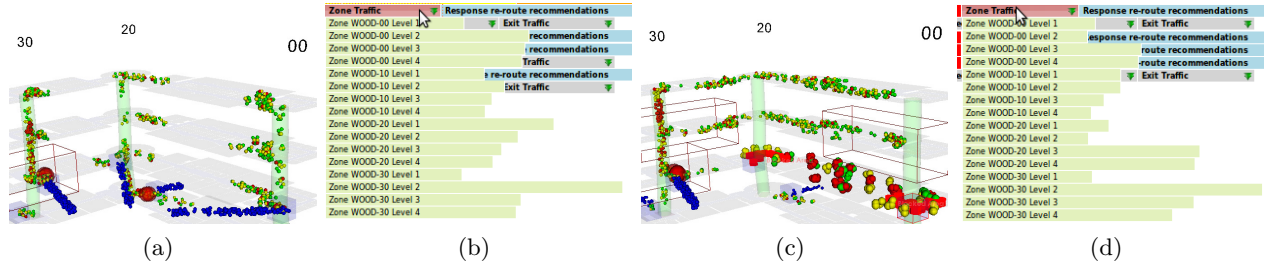
Figure 7. Contrast between normal(unblocked) vs. blocked scenarios. Two blockages have been introduced. (a,b) 3D view at 108 seconds into the evacuation, with top floors mostly evacuated, (c,d) 3D view at 139 sec. Floors 3 and 4 are still heavily occupied. Zone traffic (right panels) confirm and illustrate the aggregate picture of the traffic across the entire evacuation.
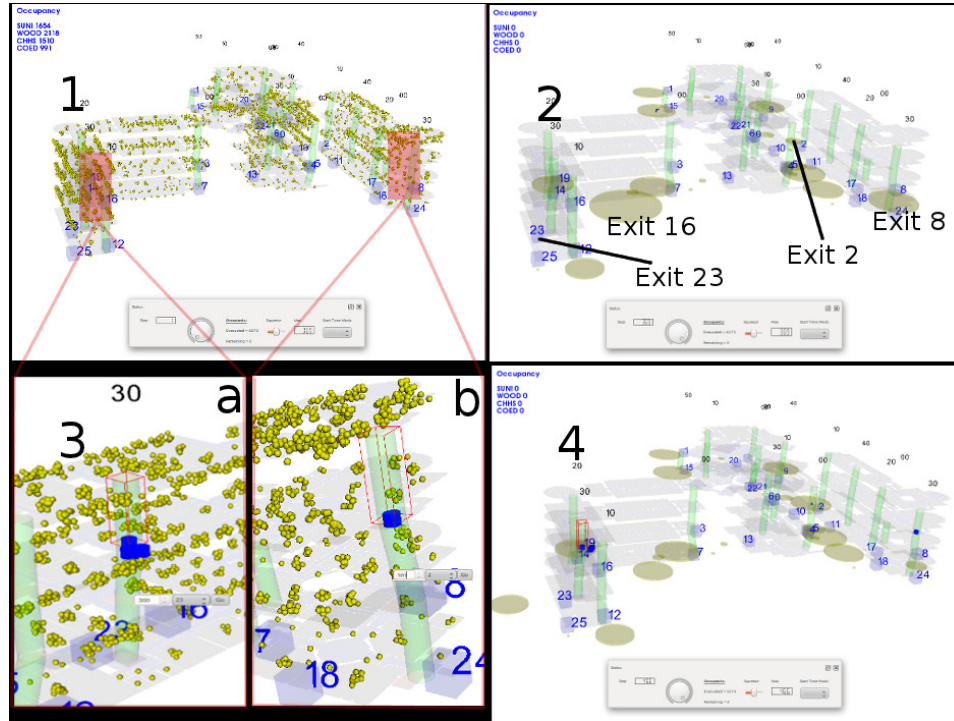

Figure 8. Rerouting evacuees from congested areas in a 4 building evacuation simulation. **1.** Four building cluster. **2.** Resulting evacuee density circles after simulation. **3a and 3b.** User added rerouting flags as indicated by the blue discs and associated with the opaque red rectangles in 1. **4.** Resulting evacuee density circles after modified simulation, changing the routes of evacuees to exit 2 and exit 23 in their respective buildings.

Even though the bar chart for the blocked case is the total picture it is still clear from both the building and zone traffic charts that the traffic on the top two floors is heavier in the blocked case, and in particular, is shown by the size of the bars labeled Zone WOOD-20 Level 3, Zone WOOD-20 Level 4, Zone WOOD-30 Level 3, and Zone WOOD-30 Level 4. The total picture of the bar chart and the single step picture of the building reinforce each other.

5.3 Situation 3: Rerouting Evacuees

When certain parts of a building are blocked, we can reroute evacuees in that area to other nearby less utilized exits. Also, our system permits a selected number of evacuees to be rerouted to reduce congestion at a stairwell or exit. This operation is performed in real-time and the simulation played through to evaluate the traffic or congestion patterns resulting from such an intervention, and thereby result in dispatching a responder to the affected area.

Figure 8 illustrates an example evacuation from a cluster of 4 academic buildings. The original evacuee densities are illustrated in panel 2. Exits 8 and 16 are heavily used, as indicated by the area of their exit circles. The red cubes in panel 1 have been interactively selected(panel 3 shows a zoomed-in view of these areas) for rerouting occupants within those areas. This is followed by specifying the number of evacuees to be rerouted to specific exits (here exits 2 and 23 were chosen). Panel 4 shows the results of these actions, leading to reduced densities at exits 8 and 23.

This function has value for planning and training. Building lockdown and release operations can benefit from such 'what-if' style scenarios that brings together rich spatio-temporal information into the hands of first responders. In this example, running the entire scenario from initiation to results and analysis took approximately 2 minutes. When large collections of buildings are involved with traffic routed to the adjacent street networks, such tools can be invaluable for effective and timely evacuation as well as optimal asset deployment.

6. EVALUATION: TABLETOP EXERCISE

The development of our application has included regular feedback and demonstrations with campus emergency and safety personnel, including the chief of police, other safety officers, and campus business continuity staff. As part of evaluating the system, we conducted a table top exercise with our campus police. We ran the application through three different scenarios to determine our system's usability, effectiveness, and need for improvements. A business continuity office staff member designed the scenarios. The campus police chief, a senior police officer, and the software team participated in the exercise.

All three scenarios involved a cluster of four campus buildings and a base scenario for the evacuation of approximately 5000 evacuees. The preprocessing step was timed at approximately 8 minutes. All simulations used this base evacuation object. Video of each of the 3 exercises (excerpts in attached video) were recorded for analysis, followed by feedback from the emergency personnel. The system was operated by a member of the software team while commands were received from the police chief.

6.1 Scenario 1. Gas Leak in Building.

Figure 9 shows time sequenced snapshots of a simulated gas leak somewhere in the exercise area. Initially the gas leak was reported as "near Woodward Hall". The police chief requested a simulation start. As seen in Fig. 9(a) the buildings are being evacuated as expected with all exits being utilized. Several seconds into the simulation (sim time: 7:26:38) a report is received that the leak is in the "courtyard", as shown in the red ellipse in the figure. The simulation is halted and reset. The police chief instructed first responders to be dispatched to the building exits facing the courtyard. Also, entrance/exits into the courtyard were to be blocked from further use.

The simulation was restarted based on the new situation. We interacted with the software by placing blockages at the requested areas from 7:27:27 until 7:28:19 (Figs. 9(a), 9(b)). At this point, the software began to recalculate the 5000 evacuee paths. At 7:29:08 calculations were completed and the reporting process rebuilt, including the scenario timeline and the temporal congestion and exit utilization charts. The police chief requested to see the simulation based on the new situation.

Evacuees are confirmed to be exiting the buildings away from the hazard, as seen in Figs. 9(c), 9(d). The simulated time to exit all buildings increased from 318 seconds to 687 seconds. There were large evacuee populations in the areas of Woodward hall opposite the hazard and it was noted that due to the blockages in the second floor, some of the evacuees were trapped.

6.2 Scenario 2. Active Shooter in Building.

Figure 10 shows time sequenced snapshots of a simulated active shooter exercise in the Woodward hall. First, the police chief ordered a campus lock down and the building to be evacuated. At this point we switched from the base evacuation scenario of the lower quad (building cluster) to a base scenario of Woodward hall. The reason for this is that the scenarios are built as objects and run apriori. We could have placed blockages in the locked down buildings but chose to open a single building scenario for the purposes of the exercise.

Some highlights in this exercise include: building rerouting and reporting occurs in 49 seconds (3 seconds for evacuee rerouting and 46 seconds for report generation). The total time here is similar to the multi-building
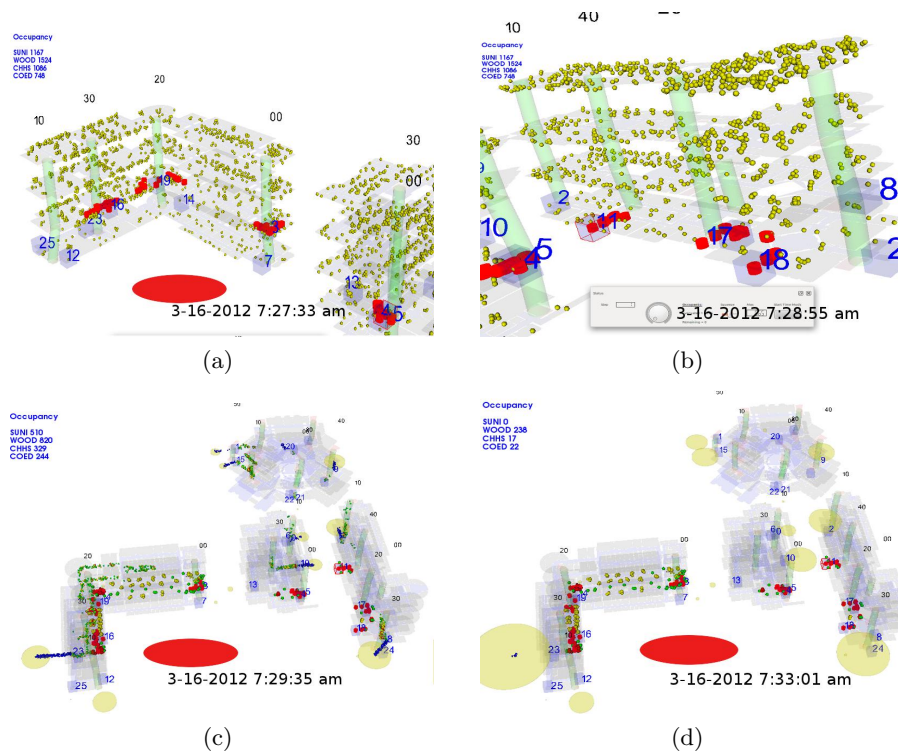
Figure 9. Tabletop Exercise: Gas Leak in Building. At approximately 7:20 a gas leak is reported. Lower quad campus possibly affecting four buildings. Evacuation simulation begins. At **Sim. Time: 7:26:38.**, leak confirmed near red ellipse. Simulation suspended, assets are deployed to prevent evacuation into the hazard. Views modified to simulate asset activities at building exits. (a)**Sim. Time; 7:27:33.** Blocking building exits into quad (b)**Sim. Time; 7:28:55.** Blocking building exits from adjacent buildings into quad complete. Signal sent for application to perform situationally aware rerouting, (c) **Sim Time; 7:29:35.** Visualization of new simulation, evacuation in progress, simulating responders interaction at exits to affected areas, (d) **Sim. Time; 7:33:01.** Evacuees are avoiding hazard and and exiting to safe zones, evacuee densities are indicated by areas of yellow circles.

evacuation because our base scenario included 3600 evacuees. This simulates a highly overloaded building to exercise the software for testing.

6.3 Scenario 3. Explosion in Utility Plant

An explosion in the RUP(regional utility plant) building created a scenario where the four building evacuation simulation of Figure 9 was also used. This scenario also found evacuees blocked in the upper floors and the explosion created a hazard in the building courtyard. As reports were received the building floors were blocked and the simulation was started. As more reports were received it became obvious that the personnel would exit toward the hazard in the courtyard. The exit density circles alerted the police chief to this problem and emergency personnel were dispatched to redirect these evacuees. At this point the police chief requested the exits facing the courtyard to be blocked. The simulation was restarted and evacuation times and exit results were evaluated as in previous scenarios.

6.4 Analysis and System Assessment

We detail below both the observations from first responders as well as the important features and current limitations of our evacuation system, as noted from the table top exercise. Overall, the feedback from the chief of police (who played the role of incident commander) and his officers was positive and consisted of the following observations:
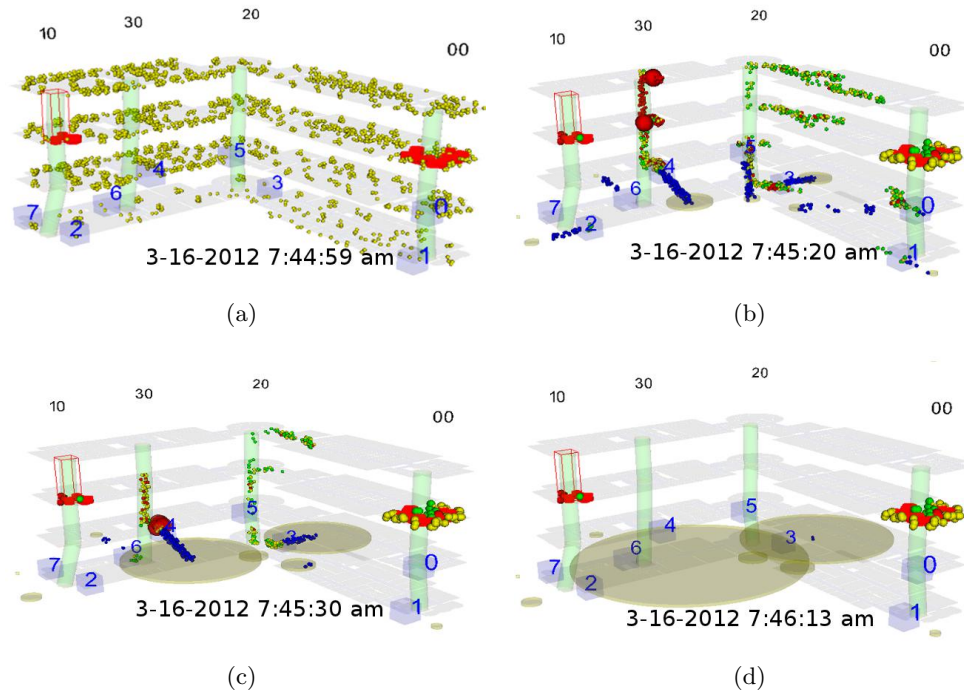
Figure 10. Tabletop Exercise:Active Shooter. At approximately 7:40 a shooter is reported at Woodward hall. Campus is locked down. Reports are received that 3rd floor stairwells are blocked at each end of the building. At 7:44:10 blockages are placed in Woodward by the operator and the simulation is recalculated. (a)**Sim. Time; 7:44:59.** Processing for new evacuation simulation is complete, commander orders visualization of new simulation, (b) **Sim. Time; 7:45:20.** Trapped evacuees noted at 3rd floor zone 00 and zone 10, (c)**Sim. Time; 7:45:30.** Extreme congestion noted at stairwells in floors 2, 3, and 4 at zone 30. (d)**Sim. Time; 7:46:13.** Simulated evacuation complete. Evacuee populations are indicated by approximate area occupied circles.

- The ability to see the 3D layout of the buildings and surrounding areas and get a sense of the current situation was considered the most valuable. The ability to see the evacuation unfold, the buildup at congestion points and the ability to direct evacuees away from a hazard were considered critically important.

- The near real-time responsiveness of the system and the ability to see the evacuation under blockages was valuable for assessment and taking appropriate action, such as dispatching first responders.

- In the gas leak exercise, a review of the evacuation helped the commander quickly size up the situation (number of evacuees, exit routes, etc) and order a building evacuation. Once the hazard was located, evacuees were routed away from it by injecting suitable blockages at key points in the building.

- In the active shooter scenario, the police chief noted that the exit utilization and congestion reports would be an invaluable tool for first responders to analyze the condition of a building and dispatch personnel.

- Additional work on the user interface will be needed to ensure minimize delays during a dynamically changing situation; for instance, blockages are specified one at a time; a 'lasso' style interface to to specify multiple blockages was considered more intuitive and efficient.

- A limitation of the current system is its inability to localize blockages to the exits, trapping evacuees in the vicinity.

- A visualization issue that is common to visual analytic systems is visual clutter and the ability to un-ambiguously visualize critical information. As we extend our system to incorporate tens of buildings in evacuation scenario, these issues will require careful design and representation choices, with input from responders.

7. CONCLUSIONS

In this work we have presented a visual analytic system for situationally aware evacuations of large urban structures. The goals of this work are to provide visual analytic tools that can be used in real-world scenarios (large urban structures, and dense collections of buildings) and more importantly, be able to run evacuation scenarios in the context of dynamically changing conditions. We have developed and used an LOD representation of building graphs that can be used as part of a visual analytic system for near real-time response. This in turn permits situational changes to be incorporated into the underlying models and evacuees rerouted. Finally, our visual analytic system provides recommendations through the reporting functions that can be used for effective use of scarce resources in dispatching responders to areas of need during the emergency. We evaluated our system with first responders, including the campus police chief, a senior police officer and public safety and business continuity/planning personnel. Input from these experienced personnel is invaluable. A tabletop exercise was performed with three different scenarios(gas leak in building, active shooter, and explosion) overseen by the police chief, acting as the situation commander. Overall, the system performed well, as evidenced by direct feedback from the first responders, with valuable suggestions to improve the system.

8. ACKNOWLEDGMENTS

APPENDIX A. MODIFIED CAPACITY BASED ROUTE PLANNER

Algorithm 1: Modified Capacity Based Route Planner

Input: ;
(1) G(N,E): Directed Graph, N nodes, E edges;
(2) Node Properties: capacity, occupancy;
(3) Edge Properties: capacity, travel time;
(4) Set of Source Nodes;
(5) Set of Destination Nodes;
(6) Set of evacuee objects;
Result: Evacuation Plan : Routes with schedules of evacuees on each route
foreach *evacuee i at each source node s* **do**
    path_found = find shortest path p from s to all destinations   with available capacity;
    if *path_found* **then**
        while $p < max\_capacity$ **do**
            /* can route evacuee via p */ evacuees[i].path = p;
        end
    end
    move all evacuees;
end
while *evacuees not at destination nodes* **do**
    move all evacuees;
end

REFERENCES

[1] Lu, Q., George, B., and Shekhar., S., "Capacity constrained routing algorithms for evacuation planning: A summary of results.," *Springer-Verlag Berlin Heidelberg 2005* , 291–307 (Sept. 2005).
[2] Kamiyama, N., Katoh, N., and Takizawa, A., "An efficient algorithm for the evacuation problem in a certain class of networks with uniform path-lengths.," *Discrete Applied Mathematics* , 3665–3677 (2009).

[3] Shekhar, S. and Yoo., J. S., "Processing in-route nearest neighbor queries: a comparison of alternative approaches," in [*Proceedings of the 11th ACM international symposium on Advances in geographic information systems*], (2003).

[4] Kim, S., Shekhar, S., and Min, M., "Contraflow transportation network reconfiguration for evacuation route planning," *IEEE Trans. on Knowl. and Data Eng.* **20**, 1115–1129 (August 2008).

[5] Lo, S. M., Liu, M., and Yuen, R. K. K., "An artificial neural-network based predictive model for pre-evacuation human response in domestic building fire," *Fire Technology* , 431–449 (Sept. 2009).

[6] Fang, G., "Swarm interaction-based simulation of ocucupant evacuation.," *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application* (2008).

[7] Guy, S. J., Chhugani, J., Kim, C., Satish, N., Lin, M., Manocha, D., and Dubey, P., "Clearpath: Highly parallel collision avoidance for multi-agent simulation," in [*Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2009)*], Grinspun, E. and Hodgins, J., eds. (2009).

[8] Castle, C. J. E. and Crooks., A. T., "Principles and concepts of agent-based modelling for developing geospatial simulations," in [*Centre for Advanced Spatial Analysis. University College London*], **110**, 1–52 (2007).

[9] Thomas, J. and Cook, K., [*Illuminating the Path: The Research and Development Agenda for Visual Analytics*], IEEE Press (2005).

[10] Andrienko, G., Andrienko, N., and Bartling, U., "Interactive visual interfaces for evacuation planning," in [*Working Conference on Advanced Visual Interfaces(AVI) 2008 Proceedings*], 472–473, ACM Press (2008).

[11] Campbell, B. and Weaver, C., "Rimsim response hospital evacuation: Improving situation awareness and insight through serious games play and analysis," *Journal of Information Systems for Crisis Response and Management* **3**, 1–15 (Jul-Sept 2011).

[12] Kim, S., Maciejewski, R., Ostmo, K., Delp, E., Collins, T., and Ebert, D., "Mobile analytics for emergency response and training," *Information Visualization* **7**(1), 77–88 (2008).

[13] Kim, S., Yang, Y., Mellama, A., Ebert, D., and Collins, T., "Visual analytics on mobile devices for emergency response," in [*IEEE Symposium on Visual Analytics Science and Technology (VAST)*], 35–42 (2007).

[14] Meiguins, B. and Meiguins, A., "Multiple coordinated views supporting visual analytics," in [*Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration*], 40–45, ACM, New York, NY, USA (2009).

[15] Ivanov, Y., Wren, C., Sorokin, A., and Kaur, I., "Visualizing the history of living spaces.," *IEEE Transactions on Visualization and Computer Graphics* **110**, 1153–1159 (November 2007).

[16] Andrienko, N., Andrienko, G., and Gatalsky., P., "Towards exploratory visualization of spatio-temporal data.," in [*3rd AGILE Conference on Geo-graphic Information Science*], (2000).

[17] Stasko, J., Gorg, C., Liu, Z., and Singhal, K., "Jigsaw: Supporting investigative analysis through interactive visualization," in [*Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology*], 131–138, IEEE Computer Society (2007).

[18] J.Liu, K.Lyons, Subramanian, K., and Ribarsky, W., "Semi-automated processing and routing within indoor structures for emergency response applications," in [*Proceedings of SPIE Conference on Defense, Security, and Sensing, April 2010*],

[19] van Rossum, G. and Drake, F., [*An Introduction to Python*], Network Theory Ltd. (2003). WWW: www.python.org.

[20] Schroeder, W., Martin, K., and Lorensen, B., [*The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*], Prentice Hall Inc., 4th ed. (2006). www.vtk.org.

[21] Qt. http://qt.nokia.com.