Integrating Visual Analytics Curriculum Into Your Own Classroom

Ross Maciejewski

Ross Maciejewski
E-mail: rmacieje@asu.edu
http://rmaciejewski.faculty.asu.edu/
http://lab.vader.asu.edu
The Ongoing Data Deluge

- Since 2003, digital information has accounted for 90% of all information produced\(^1\)
- In 2009, drones from Iraq and Afghanistan recorded 24 years of video footage
- In 2010, the amount of information added annually to the digital universe was estimated to be nearly 1 ZB
- Wal-mart process > 1 million transactions per hour
- By 2013 Cisco estimates the annual internet traffic will be 667 EBs

---

The Ongoing Data Deluge

- FedEx’s ships more than 8.5 million packages per day (http://www.fedex.com/us/about/today/companies/corporation/facts.html)
- Consumers carry more than 1 billion Visa cards worldwide. More than 450 million of those cards are in the United States (http://www.creditcards.com/credit-card-news/credit-card-industry-facts-personal-debt-statistics-1276.php)
Data Overload

- Opportunity: Huge amounts of data available in digital form and ready for analysis!
  - But, how can we make sense of this data?
  - How can we harness this data in the decision making process?
  - How do we avoid being overwhelmed by all of this data?
Having Fun In the Data Deluge

- More people are injured on toilets than by skiing or snowboarding.
- More households have dogs as pets than cats, but cat lovers are more likely to have multiple pets.
- Americans are eating more peanuts and drinking less coffee.
- More Americans belong to a fantasy sports league (10.6 million) than to book clubs (5.7 million).
- Book club members are outnumbered by avid bird-watchers (5.8 million).

Facts from the Statistical Abstract of the United States, 2013
Why Is This Important to Educators?

- In 2011, a report by the McKinsey Global Institute projected that the US needs 140,000-190,000 works with “deep analytical” expertise.
- These are not just computer science jobs!
- Harvard Business Review calls data science “the sexiest job in the 21st century”.
- We need to train the next generation of scientists for data analytics!

1- [http://www.mckinsey.com/Features/Big_Data](http://www.mckinsey.com/Features/Big_Data)
What Goes Into an Degree in “Data Science”?

• Typically Data Science is a Masters Program
• Basic programming
• Data cleaning and scraping
• Predictive statistics (regression modeling)
• Ethics classes on data privacy implications
• What about communicating findings, what other skills do we need?
The Power of Visualization

- The goal is to take all of this *data* and transform it into *information*
- How many terabytes of data we have collected doesn’t matter, it’s how many petaflops of *insights* we can generate from this data
- We need to make the data understandable to people and a key way of doing this is through *visualization*
What Skills Are Students Missing?
Non-Data Components of Graphs

- Axes and legends can often be as important as the data themselves.
- Poor axis choices and label choices can lead to confusing visualizations.
- Axis tick labels provide cognitive context for most basic plot types.
- They support estimation and contribute to the overall appearance of the graphic.
- Cleveland suggests choosing the scales so that the data rectangle fills up as much of the scale-line rectangle as possible.

Extension of Wilkinson’s Algorithm

- Utilizes four components for optimizing tick placement
  - Simplicity $= 1 - \frac{i - 1}{|Q| - 1} - j + v$
  - Density $= 2 - \max\left(\frac{\rho}{\rho_t}, \frac{\rho_t}{\rho}\right)$

<table>
<thead>
<tr>
<th>Input</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$</td>
<td>preference-ordered list of nice step sizes</td>
<td></td>
</tr>
<tr>
<td>$(d_{min}, d_{max})$</td>
<td>data range</td>
<td></td>
</tr>
<tr>
<td>$\rho_t$</td>
<td>target label density</td>
<td></td>
</tr>
<tr>
<td>$f_s_t$</td>
<td>target font size</td>
<td></td>
</tr>
<tr>
<td>Formats</td>
<td>set of label formats</td>
<td></td>
</tr>
<tr>
<td>(defaults shown in Table 3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Other
  - $q$ element of $Q$
  - $j$ skip amount
  - $i$ index of $q$ in $Q$
  - $(l_{min}, l_{max})$ labeling sequence range
  - $v$ equals 1 if labeling includes 0, 0 otherwise

Graph Aspect Ratios

• Our ability to perceive trends and patterns in a given display is heavily influenced by the aspect ratio
• Aspect ratio affects densities, relative distances and orientations
• Several methods have been proposed for automatically selecting the aspect ratio
• Aspect ratio: $a = \frac{\text{width}}{\text{height}}$

How Can We Integrate Data Analytics Into Our Classes?
National Health and Nutrition Examination Survey (NHANES)

*Designed to assess health and nutritional status of adults and children in the US*

- Combines interviews and physical examinations
- Exams a nationally representative sample of about 5,000 persons per year
- Includes demographic, socioeconomic, dietary and health-related questions
Healthy Eating Index (HEI)

Measure of diet quality that assesses conformance to federal dietary guidance

USDA uses HEI to monitor diet quality of the US population

HEI-2005 uses recommendations of the MyPyramid guidance as the baseline


### Healthy Eating Index—2005 components and standards for scoring

<table>
<thead>
<tr>
<th>Component</th>
<th>Maximum points</th>
<th>Standard for maximum score</th>
<th>Standard for minimum score of zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Fruit (includes 100% juice) (HEI-1)</td>
<td>5</td>
<td>≥0.8 cup equiv. per 1,000 kcal</td>
<td>No Fruit</td>
</tr>
<tr>
<td>Whole Fruit (not juice) (HEI-2)</td>
<td>5</td>
<td>≥0.4 cup equiv. per 1,000 kcal</td>
<td>No Whole Fruit</td>
</tr>
<tr>
<td>Total Vegetables (HEI-3)</td>
<td>5</td>
<td>≥1.1 cup equiv. per 1,000 kcal</td>
<td>No Vegetables</td>
</tr>
<tr>
<td>Dark Green and Orange Vegetables and Legumes²  (HEI-4)</td>
<td>5</td>
<td>≥0.4 cup equiv. per 1,000 kcal</td>
<td>No Dark Green or Orange Vegetables or Legumes</td>
</tr>
<tr>
<td>Total Grains (HEI-5)</td>
<td>5</td>
<td>≥3.0 oz equiv. per 1,000 kcal</td>
<td>No Grains</td>
</tr>
<tr>
<td>Whole Grains (HEI-6)</td>
<td>5</td>
<td>≥1.5 oz equiv. per 1,000 kcal</td>
<td>No Whole Grains</td>
</tr>
<tr>
<td>Milk³ (HEI-7)</td>
<td>10</td>
<td>≥1.3 cup equiv. per 1,000 kcal</td>
<td>No Milk</td>
</tr>
<tr>
<td>Meat and Beans (HEI-8)</td>
<td>10</td>
<td>≥2.5 oz equiv. per 1,000 kcal</td>
<td>No Meat or Beans</td>
</tr>
<tr>
<td>Oils⁴ (HEI-9)</td>
<td>10</td>
<td>≥12 grams per 1,000 kcal</td>
<td>No Oil</td>
</tr>
<tr>
<td>Saturated Fat (HEI-10)</td>
<td>10</td>
<td>≤7% of energy⁵</td>
<td>≥15% of energy</td>
</tr>
<tr>
<td>Sodium (HEI-11)</td>
<td>10</td>
<td>≤0.7 gram per 1,000 kcal⁵</td>
<td>≥2.0 grams per 1,000 kcal</td>
</tr>
<tr>
<td>Calories from Solid Fats, Alcoholic beverages, and Added Sugars (SoFAAS) (HEI-12)</td>
<td>20</td>
<td>≤20% of energy</td>
<td>≥50% of energy</td>
</tr>
</tbody>
</table>

¹Intakes between the minimum and maximum levels are scored proportionately, except for Saturated Fat and Sodium (see note 5).
²Legumes counted as vegetables only after Meat and Beans standard is met.
³Includes all milk products, such as fluid milk, yogurt, and cheese, and soy beverages.
⁴Includes nonhydrogenated vegetable oils and oils in fish, nuts, and seeds.
⁵Saturated Fat and Sodium get a score of 8 for the intake levels that reflect the 2005 Dietary Guidelines, <10% of calories from saturated fat and 1.1 grams of sodium/1,000 kcal, respectively.
Playing Data Detective

So, we’ve collected all of these surveys, now what?

Well, how about Exploratory Data Analysis where we visually inspect the data and ask questions about it to help form hypotheses.

What are some easy questions to ask?

• What is the average … (total fruit consumption)?
• What is the average … (milk consumption)?
Utilizing Survey Cycle 2-3

- **Cycle 2:** 2001-2002
- **Cycle 3:** 2003-2004
- **Sample size of ~15000**

**Other variables:**
- Gender
- Age
- Race
- Total energy for the day (kcal)
- Total saturated fatty acid (gm)
- BMI

So with more data, maybe we can think of more questions!
What is the average ... ?
What is the average ... ? Meat Consumption
A Century of Meat

American consumption of chicken and beef rose substantially after World War II, aided by the development of intensive farming methods, the proliferation of fast-food restaurant chains and supermarkets and the adoption of reliable home refrigeration.

Beef consumption peaked in 1976 but then declined, in part because of the publication of new dietary guidelines and studies that associated saturated fats and cholesterol with heart disease.

*Note: per capita availability of boneless meat is a proxy for human consumption, and is lower than retail weight or carcass weight. Bones, offal and game are excluded.

Sources: U.S. Department of Agriculture (data); news and company reports; “Putting Meat on the American Table,” by Roger Horowitz

What is the average … ? Total Fruit Consumption
What is the average ... ? Dark Green and Orange Vegetable Consumption
How Do These Compare?
So, What Can I say About the Population as a Whole?

- *We consume lots of SOFA*s
- *We are not good at eating our vegetables*
- *We eat lots of grains …*
- *But not lots of whole grains*

- *But what does this say about someone like you?*
What if I further segment the data? Filter by Age, Gender, Etc.?

Let’s look at Females, Age 19-30

18.5 < BMI < 25.0

Total population in our sample = 511
NHANES in Females Age 19-30, Average BMI
Back to the Entire Population!
Forming Hypotheses

What do we think we see when comparing this group of females to the entire population?

- Their SOFA consumption is more evenly distributed
- Their meat and bean consumption is more skewed
What if I further segment the data? Filter by Age, Gender, Etc.?

Let’s look at Males, Age 14-18, BMI < 18.5

Total population in our sample = 145
NHANES in Adolescent Underweight Males
Back to the Entire Population!

- Total Fruit
- Whole Fruit
- Total Vegetables
- Dark Greens
- Total Grains
- Whole Grains
- Milk
- Meat and Beans
- Oils
- Saturated Fat
- Sodium
- SOFAS
Forming Hypotheses

What do we think we see when comparing this group of females to the entire population?

- Their vegetable consumption is more evenly distributed
- Their meat and bean consumption is more skewed
- Their SOFAs are more evenly distributed
What if I want to ask: How is … related to … ?

From our histograms, we came up with interesting insights into our data

But what if I want to know about relationships between the data?

More like:
Do people that consume more milk consume more vegetables?

We can try a scatterplot!
I don’t see any obvious patterns in this 😞
Now what do I do?
What if I further question the data? All relationships by Age, Gender?

How many of these did you really expect me to do?

These histograms were made in Excel

One can also use SAS, JMP, MatLab

In all of those programs we have to write scripts, modify variables, etc.

We can’t directly manipulate our variables

So, being a data detective can be very tedious

And these were just simple comparisons!
So what do we do now?

We can utilize interactive graphics combined with underlying analytical processes to help us be better detectives!
The Curse of Dimensionality

A term coined by Bellman in 1961

Refers to the problems associated with multivariate data analysis as the dimensionality increases the available data becomes sparse

Sparsity is a problem for any method that requires statistical significance

Sometimes data dimensions are redundant and can be reduced with minimal information loss

In visualization we are also limited with screen space and the number of available visual variables, so choosing the most appropriate dimensions is key
The Curse of Dimensionality

So, what can we do?

- We can incorporate prior knowledge of the data
- We can smooth the target function
- We can reduce the dimensionality
The Practicality of the Curse of Dimensionality

For a given sample size, there is a maximum number of features above which the performance of classifying samples will degrade rather than improve.

In most cases, the additional information that is lost when discarding some features is compensated by a more accurate mapping in the lower-dimensional space.

So, how do we know what features we can throw away?

For visualization this implies that there are some features of a dataset that will be better to visualize (contain more information) than others!
Curses! I’m Still Overwhelmed!

Probably so, we are creating a lot of data in the world

Overall, NHANES is actually a relatively modest set of data

As our datasets get more variables we need ways to question the data

What I really want to know is are there relationships/groups/clusters in the data, how can I see these when we have so many dimensions?

One way is to just cluster the data and draw these clusters in an arbitrary space (we saw this in the demo)

Another is to try dimensional reduction!
Dimensional Reduction

Two approaches are available to reduce dimensionality

- **Feature extraction**: creating a subset of new features by combinations of the existing features
- **Feature Selection**: choosing a subset of all the features

1. Given a feature space $x_i \in \mathbb{R}^N$ find a mapping $y = f(x): \mathbb{R}^N \rightarrow \mathbb{R}^M$ with $M < N$ such that the transformed feature vector $y \in \mathbb{R}^M$ preserves (most of) the information or structure in $\mathbb{R}^N$

2. An optimal mapping is one that does not increase error
What Does Dimensional Reduction Mean For You?

**Feature Selection**: choosing a subset of all the features

- Well, that’s what we were already doing
- We use our own expert knowledge to select features of interest
- Note we could also use machine learning and statistics here too if we were so inclined
What Does Dimensional Reduction Mean For You?

Feature extraction: creating a subset of new features by combinations of the existing features

- You can think of this in terms of the NHANES data
- Let’s say I do an analysis and I find a correlation between SOFA consumption and sodium consumption (the more SOFAs, the more sodium)
- I could combine SOFAs and sodium into a new variable – SSOFAS
- I make up some formula to do this:
  - \( SSOFAS = \alpha \times SOFAs + \beta \times sodium \)
- So basically we just combine variables that are correlated
Principle Components Analysis

One of the most commonly applied dimension reduction techniques

PCA is a deterministic analytical procedure that utilizes an orthogonal transformation to reduce a set of sample observations with potentially correlated variables into a set of uncorrelated variables called principal components

The number of principal components will always be less than or equal to the original number of variables in the sample set

Principle Component Analysis

The main limitation of PCA is that it does not consider class separability since it does not take into account the class label of the feature vector

- PCA simple performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance
- There is no guarantee that the directions of maximum variance will contain the most interesting/important features
That sounded really hard!

- Yes, but we have software to help us!
- In JMP Principal components can be accessed through the Principal Components command on the Analyze > Multivariate Methods menu.

- But, how does principal component analysis help me?
- What can principal component analysis tell me?
Principal Component Analysis

- Let’s go back to NHANES for an example and focus on the HEI values.
- Each person in my data set has 12 HEI values, so you can think of each person as a point in 12-dimensional space.
- It is hard to understand all the relationships in this space, PCA tries to figure out what dimensions to combine for you automatically.
- After you run PCA you will get a set of what you can think of as new variables.
- Since there were 12 variables in our NHANES data, we will get at most 12 principal components (but hopefully less!)
Ok, you’ve convinced me to run PCA on NHANES and I get 8 components, but now I’m confused!

The output is such that your first component accounts for as much of the variability in the data set as possible.

The second component accounts for the second most variability, etc.

Typically we will plot the data using the first two principal components as axes.

We can also write out each component to analyze it.
Underweight Adolescent Males

FV1 ≈ 0.14 HEI1 + 0.07 HEI2 + 0.02 HEI5 + 0.02 HEI6 - 0.09 HEI7 + 0.19 HEI8 + 0.05 HEI3 + 0.03 HEI4 + 0.20 HEI9 + 0.20 HEI10 - 0.15 HEI11 + 0.91 HEI12

FV2 ≈ 0.13 HEI1 + 0.06 HEI2 + 0.05 HEI5 + 0.01 HEI6 + 0.70 HEI7 - 0.27 HEI8 - 0.05 HEI3 + 0.02 HEI4 - 0.48 HEI9 - 0.30 HEI10 - 0.18 HEI11 + 0.25 HEI12
Females Age 19-30, Average BMI

FV1 ≈ 0.13 HEI1 + 0.11 HEI2 + 0.04 HEI5 + 0.03 HEI6 - 0.00 HEI7 + 0.11 HEI8 + 0.08 HEI3 + 0.07 HEI4 + 0.17 HEI9 + 0.10 HEI10 - 0.12 HEI11 + 0.94 HEI12

FV2 ≈ -0.01 HEI1 + 0.01 HEI2 + 0.02 HEI5 + 0.03 HEI6 + 0.72 HEI7 - 0.07 HEI8 + 0.00 HEI3 + 0.00 HEI4 - 0.18 HEI9 - 0.62 HEI10 - 0.22 HEI11 + 0.08 HEI12
Putting it all together

• *If I know things about you …*
  • Age, gender, BMI

• **Can I guess your typical daily diet?**
  • Probably not very accurately

• *But, if I ask you a few questions, maybe I can get closer!*
Putting it all together

This is the NHANES data
I know everyone’s diet

Split by demographics
(Learned or *a priori*)

Total Population

I can first split the data by demographics

This is you.
Maybe you tell me your age, gender, BMI?
(Inform me maybe?)

Now I know your diet!
(Well sort of)

This is still you.
I know you’re in this large group
But I want to know your subgroup.

If I know how many sodas you drink, maybe I can narrow it down from three groups to two

Maybe if you ate a salad today tells me more!

Now my demographics groups having groups based on diet!
Summary

• *Data analysis is hard!*

• *We need better tools to be data detectives*

• *By being data detectives, we can ask questions and maybe create predictions about unknown data based on available data (we make a model of the data)*

• *Visual analytics helps us in the data detective work!*

• *But we still need to understand what the analyses do for us so that we can put them into practice.*
Questions?
(http://rmaciejewski.faculty.asu.edu)
Data Visualization Principles
(Some Things I’ve Learned)

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
Three themes
1. Understanding purpose
2. Embracing interaction
3. Identifying value
Talented students
1. Understanding Purpose

Why visualize?
Applications of Visualization

Presentation  (Explanatory)
Analysis      (Exploratory)
1. Presentation
Communicate data and ideas
Explain and inform
Provide evidence and support
Influence and persuade
Infographics
All the rage...
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

Produced by WatchTheAmericans.com and Fan Website

http://visual.ly/nuclear-arms-race
Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

- Hank Aaron: 755 home runs, 23 seasons
- Babe Ruth: 714 home runs, 22 seasons
- Barry Bonds: 708 home runs, 20 seasons

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run rates for each player after age 24.

- Aaron: Actual home runs slightly outpace projected home runs for five seasons.
- Ruth: Averaged 46.4 home runs a season from age 30 to 34. Averaged 46.5 for next four seasons.
- Bonds: From age 35 to 39, he averaged 41 more home runs a season than projected.

Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (257th):

- Hank Aaron: 755
- Babe Ruth: 714
- Barry Bonds: 708
- Willie Mays: 660
- Sandy Sosa: 588
- Frank Robinson: 582
- Mark McGwire: 583
- Ken Griffey Jr.: 580
- Alex Rodriguez: 429
- Albert Pujols: 301

- 10 HR: 14 (2016 SEASONS), 76
- Hit only 20 over first five seasons.
- No one hit more from 1950-69.
- Three 68-home run seasons in record.
- Triple Crown in 1966 (49, 122, 315).
- First to hit 70 in a season.
- Only McGwire hit more in the 90's.
- Youngest to reach 400 home runs.
Hans Rosling
Gapminder

An Inconvenient Truth

Gore made extensive use of data graphics
Frequent presentation goals

Clarify
Focus
Highlight
Simplify
Persuade

May just show a few variables and/or a subset of the data cases
2. Analysis
Explore the data
Assess a situation
Determine how to proceed
Decide what to do
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
Thought
If you can articulate very precisely what you’re seeking, visualization likely isn’t your best approach

OK, so what is visualization good for?
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Frequent analysis goals
Show many/all variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
At Its Heart: Uncertainty

Nothing is perfect
Humans trade off choices, compromise

Cost

Bedroom

Journey time

Attribute Explorer

Spence & Tweedie
Interacting with Computers ‘98
Takeaway 1

Analysis $\neq$ Presentation
Views of the World

Vis researchers

Everyone else
Narrative Visualization: Telling Stories with Data

Edward Segel and Jeffrey Heer

Abstract—Data visualization is regularly promoted for its ability to reveal stories within data, yet these “data stories” differ in important ways from traditional forms of storytelling. Storytellers, especially online journalists, have increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function as a primary story. In this paper, we systematically review the design space of this emerging class of visualization. Drawing on case studies from news media to visualization research, we identify distinct genres of narrative visualization. We characterize these design differences, together with interactivity and storytelling, in terms of the balance between the narrative flow intended by the author (imposed by graphical elements and the narrative) and story discovery on the part of the reader (often through interactive exploration). Our framework suggests design strategies for narrative visualization, including promising under-explored approaches to journalistic storytelling and educational media.

Index Terms—Narrative visualization, storytelling, design methods, case study, journalism, social data analysis.

1 INTRODUCTION

In recent years, many have commented on the storytelling potential of data visualization. News organizations including the New York Times, Washington Post, and the Guardian regularly incorporate dynamic graphics into their journalism. Politicians, activists, and television reportage use interactive visualizations as a backdrop for stories about global health and economics [10] and election results [9]. A recent feature in The Economist [6] explores the proliferation of digital data and notes that visualization designers are “melding the skills of computer science, science, artistic design and storytelling.”

Static visualizations have long been used to support storytelling, usually in the form of diagrams and charts embedded in a larger body of text. In this format, the text conveys the story, and the image typically provides supporting evidence or related details. An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives.

Crafting successful “data stories” requires a diverse set of skills. Giorgini and Page [12] note that effective storytelling “requires skills that those familiar with cinema directors, beyond a technical expert’s knowledge of computer engineering and science.” While techniques from animation, prose, comic books, video games, and film production are applicable to narrative visualization, we should also expect this emerging medium to possess unique attributes. Data stories differ in important ways from traditional storytelling. Stories in text and film typically present a set of events in a tightly controlled progression. While tours through visualized data can be organized in a linear sequence, they can also be interactive, allowing viewers to ask new questions and explore new alternatives.

Currently, most sophisticated visualizations focus on data exploration and analysis. Applications such as spreadsheets and visualization tools support an array of analysis routines and visual renderings, but beyond exporting images for presentation typically provide scant support for crafting stories with analysis results. As such, they provide powerful vehicles for discovering “stories,” but do little to aid in narrative communication of those findings to others. As tools mature and more richly support the web—e.g., Many Eyes [25], Tableau Public [22], GeoTime Stories [8], they are enabling the publication of dynamic graphics with variable constrained levels of interactivity. It remains an open question how the design of such tools might be evolved to support richer and more diverse forms of storytelling.

In this paper, we investigate the design of narrative visualizations and identify techniques for telling stories with data graphs. We take an empirical approach, analyzing visualizations from online journalism, blogs, instructional videos, and visualization research. After reviewing related work, we share five selected case studies which highlight varied design strategies and illustrate our analytic approach. We then formulate a design space constructed from an analysis of 58 examples. Our analysis identifies salient dimensions of visual storytelling, including how graphical techniques and interactivity can enforce various levels of structure and narrative flow. We describe seven genres of narrative visualization: magazine-style, annotated chart, pointed poster, flow chart, comic strip, slide show, and video. These genres can be combined with interactivity and encouraging the production of varying balances of author-driven and reader-driven experiences. Finally, we discuss the implications of our framework, noting recurring design strategies, promoting yet under-explored approaches to integrating visualization with other media, and the potential for improved user interfaces for crafting data stories. By focusing on the graphical and intertextual elements of narrative visualization, our approach gives less attention to the cognitive and emotional experience of the reader. We recognize the importance of these elements, however, and describe directions for future reader-centric research in our conclusion.

2 RELATED WORK

Storytelling and visual expression are integral parts of human culture; storytelling has even been referred to as “the world's second oldest profession” [12]. Without assuming milestones of achievement, we describe a few of the key concepts informing narrative visualization.

2.1 Narrative Structure

The Oxford English Dictionary defines narrative as “an account of a series of events, facts, etc., given in order and with the establishing of connections between them.” Central to this definition is the notion of a chain of causally related events. Stories of this form often have a beginning, middle, and end [3, 241]; an introduction to the situation, a series of events often involving tension or conflict, and a resolution.

Since ancient times, people have used and understood the elements of storytelling. For example, writers (e.g., [5, 15, 218]) have developed protocols of dramatic situations and identified plot lines common to many narratives, such as the “hero’s journey” [5]. This research typically distinguishes between the content of the story and the form in which it is told. While stories of common concerns are often interesting, they may also present a sequence of facts and observations linked together by supplying themes or arguments.

Storytelling strategies vary among media and genres. For instance, stories told through writing have access to a different set of formal mechanisms and narrative structures (e.g., stream of consciousness) than stories told through film (e.g., split-screen sequences [5]). Rumelhart [22] describes narrative devices for journalism such as the anecdote.

Jessica Hullman
Nick Diakopoulos
2. Embrace Interaction

Static ok, Interactive better
Data Visualization 101

Problem:
You have a lot of data (& attributes) to understand

Do you?

Pack all the data into one complex representation

Spread the data into multiple coordinated views

Use interaction to reveal different subsets of the data
Constituents

Two key aspects of data visualization
  Representation
  Interaction

“The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means.”

S. Few, *Now you see it*
Interaction is Vital

Engage in a dialog with your data
Interaction

Why interact?

Intent

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

Yi, Kang, Stasko & Jacko
TVCG (InfoVis) ‘07
Interaction
How manifested today?

Tooltips & selection
Get details

Navigation
Brushing & linking
Interaction
Can we do more?

Employ interaction in a more fundamental manner to strengthen the power of visualization

Some examples...
Node-link graphs

http://www.visualcomplexity.com/vc/
CiteVis

http://www.cc.gatech.edu/gvu/ii/citevis

Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster ‘13
CiteVis2

Visualization design: Panpan Xu, Chad Stolper, Anand Sainath
OnSet

http://www.cc.gatech.edu/gvu/ii/setvis/

Demo

Sadana, Major, Dove, & Stasko
TVCG (InfoVis) ‘14
Interaction
What are the tools of interaction?

Traditional – Desktop: keyboard, mouse

New – Tablet: fingers and multi-touch
Moving to Tablets

Scatterplot

Video

Sadana & Stasko
AVI ‘14

http://www.cc.gatech.edu/gvu/ii/touch/
Scatterplot touch video
Dust and Magnet

Yi, Melton, Stasko & Jacko

*Information Visualization* ‘05
Pushing it Further

Perceptive Pixel D n M
Takeaway 2

Take advantage of interaction
3. Identifying Value

Visualization’s Value
The Value of Information Visualization

Jean-Daniel Fekete¹, Jarke J. van Wijk², John T. Stasko³, and Chris North⁴

¹ Université Paris-Sud, INRIA, Bât 490, F-91405 Orsay Cedex, France, Jean-Daniel.Fekete@inria.fr, http://www.aviz.fr/~fekete/
² Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB EINDHOVEN, The Netherlands, vanwijk@win.tue.nl, http://www.win.tue.nl/~vanwijk/
³ School of Interactive Computing, College of Computing & GVU Center, Georgia Institute of Technology, 85 5th St., NW, Atlanta, GA 30332-0760, USA, stasko@cc.gatech.edu, http://www.cc.gatech.edu/~john.stasko
⁴ Dept of Computer Science, 2202 Kraft Drive, Virginia Tech, Blacksburg, VA 24061-0106, USA, north@vt.edu, http://people.cs.vt.edu/~north/

Abstract. Researchers and users of Information Visualization are convinced that it has value. This value can easily be communicated to others in a face-to-face setting, such that this value is experienced in practice. To convince broader audiences, and also, to understand the intrinsic value of visualization is more difficult. Learning in this process is more

Information Visualization: Human-Centered Issues and Perspectives
Springer, 2008
Simply presenting data *visually* can have a profound impact.
My Class

http://www.cc.gatech.edu/~stasko/7450

CS 7450 - Information Visualization

Instructor: John Stasko
Fall 2013
Mon, Wed 3:00 - 4:30 pm
Whitaker Bldg, room 1103

Information visualization is a research area that focuses on the use of visualization techniques to help people understand and analyze data. While fields such as scientific visualization involve the presentation of data that has some physical or geometric correspondence, information visualization focuses on abstract data without such correspondences such as symbolic, tabular, networked, hierarchical, or textual information sources.

The objectives of the course are

- Learn the principles involved in information visualization
- Learn about the variety of existing techniques and systems in information visualization
- Develop skills in critiquing different visualization techniques as applied to particular tasks
- Learn how to evaluate visualization systems
- Gain a background that will aid the design of new, innovative visualizations

The course will follow a lecture/seminar style with much discussion of assigned readings, as well as viewing of videos and hands-on experience with research and commercial information visualization tools.

We will be reading recent research papers about the different course topics. In addition, we will be using one book for the course: Now You See It by Stephen Few, Analytics Press 2009. Also highly recommended is Envisioning Information by Edward Tufte, Graphics Press 1990.
SalaRead 2008 2009 2010 2011
Top Salaries
Name Salary
1 $225,359.05
2 $222,102.12
3 $216,703.30
4 $200,000.00
5 $191,046.00
6 $185,201.88
7 $180,547.00
8 $179,982.02
9 $178,580.17
10 $178,207.15

Mathematics
Total Expenditure: $7,072,162.95

Average Salary by Department
Salary Range By Title

Average Salary by Title

Average Salary by Gender

SalaRead 2008 2009 2010 2011 Switch to Dashboard
Enter Name: 
Treemap by Title
All Titles
Professor
Associate Professor
Assistant Professor
Dean
Director, College/Div...
Post Doc.

Treemap by Department
All Departments

Electrical & Computer Engr...
Mechanical Engr...
Civil & Engr...
Aerospace
School of Engineering

Biomedical
Psych.
Earth & At.
Public Policy
Provo...

Scheier College of Bus...
Industria...
Mathematics
Materials Sci...
Physics

Inter. Ec.

Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Visualization’s Value?
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]
Visualization’s Value?

\[ \text{Value} = T + I + E + C \]

Ability to minimize the total **time** needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)
What kinds of questions?

“Low-level” tasks
Retrieval value
Filter
Compute derived value
Find extremum
Sort
Determine range
Characterize distribution
Find anomalies
Cluster
Correlate
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)
What is Insight?

An individual observation about the data by the participant, a unit of discovery

Complex
Deep
Qualitative
Relevant
**Unexpected**

Saraiya, North, & Duca
TVCG ‘05

North
IEEE CG&A ‘06
What is Insight?

Sudden grasp of new relationships that are necessary to solve a problem and that were not learned in the past

Bernstein, Penner, Clarke-Stewart & Roy
*Psychology, 6th edition*
What is Insight?

Is not spontaneous “aha!” moments (eg, in cognitive science)
Is knowledge-building and model-confirmation
   Like a substance that people acquire with the aid of systems

Chang, Ziemkiewicz, Green, & Ribarsky

*IEEE CG&A* ‘09
Visualization’s Value?

\( V_{\text{alue}} = T + I + E + C \)

Ability to convey an overall \textit{essence} 
or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)
## Overview and detail

### Focus + context

**Table Lens: Baseball Player Statistics**

<table>
<thead>
<tr>
<th>Player</th>
<th>Avg</th>
<th>Career Avg</th>
<th>Team</th>
<th>Salary 97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larry Herndon</td>
<td>0.247</td>
<td>0.272</td>
<td>Det.</td>
<td>225</td>
</tr>
<tr>
<td>Jose Barfield</td>
<td>0.286</td>
<td>0.276</td>
<td>Tor.</td>
<td>1237.5</td>
</tr>
<tr>
<td>Jeffrey Leonard</td>
<td>0.278</td>
<td>0.272</td>
<td>S. F.</td>
<td>900</td>
</tr>
<tr>
<td>Donnie Hill</td>
<td>0.283</td>
<td>0.272</td>
<td>Oak.</td>
<td>275</td>
</tr>
<tr>
<td>Billy Sample</td>
<td>0.285</td>
<td>0.271</td>
<td>Atl.</td>
<td>Na</td>
</tr>
<tr>
<td>Howard Johnson</td>
<td>0.245</td>
<td>0.252</td>
<td>N. Y.</td>
<td>297.5</td>
</tr>
<tr>
<td>Andru Thomas</td>
<td>0.250</td>
<td>0.256</td>
<td>Atl.</td>
<td>75</td>
</tr>
<tr>
<td>Billy Hatcher</td>
<td>0.257</td>
<td>0.252</td>
<td>Hou.</td>
<td>110</td>
</tr>
<tr>
<td>Omar Moreno</td>
<td>0.233</td>
<td>0.251</td>
<td>Atl.</td>
<td>Na</td>
</tr>
<tr>
<td>Earnell Coles</td>
<td>0.272</td>
<td>0.251</td>
<td>Det.</td>
<td>105</td>
</tr>
</tbody>
</table>

Rao & Card

CHI ’94
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to generate **confidence** and trust about the data, its domain and context

(Beneficial data analysis process side effects)
Some examples
New York City’s Weather for 1980

1.

E. Tufte

The Visual Display of Quantitative Information

1983

Data Values

365  High temp for each day
365  Low temp for each day
365  Avg high temp for each day
365  Avg low temp for each day
365  Precipitation for each day
365  Humidity for each day
12   Precipitation for each month
12   Avg precipitation for each month
1    Precipitation for the year
1    Avg precipitation per year
1    Highest temp (& day) for the year
1    Lowest temp (&day) for the year
1    Avg daily temp for the year
1    Avg daily temp per year
Visualization’s Value

\[ \text{Value} = \text{T} + \text{I} + \text{E} + \text{C} \]
Visualization’s Value

\[ V_{\text{alue}} = T + I + E + C \]

Visualization's Value

\[ \text{Value} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{value}} = T + I + E + C \]
A month-by-month look at Atlanta’s weather in 2013

Atlanta’s 2013 weather can be summed up in one word – soggy. The city ended the year 16.31 inches above normal in rainfall, with much of that rain falling during the waterlogged summer months, which had about twice as many rainy days as dry ones. Rain fell into the city’s official gauge at Hartsfield-Jackson International Airport on 17 days in June, 19 days in July and 24 days in August. The rainfall total for those three months was 23.29 inches, making the period the city’s fourth wettest summer on record. Daily rainfall records were set on April 26 (1.73 inches), May 4 (3.40 inches), June 5 (4.14 inches), Aug. 8 (2.60 inches) and Oct. 6 (2.12 inches). Summer 2013 will also be remembered for a lack of extreme heat, particularly compared to the sweltering summer of 2012. Temperatures climbed into the 90s only 20 times, topping out at 92 on June 28 and Aug. 29. That’s 32 less days of 90+ heat than in 2012, when the all-time record high of 106 was reached.

**Daily temperatures**

Atlanta saw record highs twice in 2013 – on Jan. 12 and Dec. 22, when temperatures reached 76 and 71, respectively.

**Monthly rainfall**

Sources: National Weather Service, National Drought Mitigation Center

**Georgia drought**

LINDA SCOTT, MICHAEL MORRIS / STAFF

Atlanta Journal Constitution
Jan. 3, 2014
CiteVis

Demo

Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster ‘13
Visualization’s Value

\[ V_{\text{alue}} = T + I + E + C \]
Visualization’s Value

\[ V_{alue} = T + I + E + C \]
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
Value
Not evaluation in traditional sense

User study with 12 grad students not so helpful
Takeaway 3

Understand & communicate the value
Key Open Problems in Visualization
Key Open Problems in Visualization

1. Assess and communicate value

\[ \text{Value} = T + I + E + C \]
Key Open Problems in Visualization

1. Assess and communicate value

2. Make the construction of visualizations easier

DataWrapper
https://datawrapper.de/

iVisDesigner
Ren, Hollerer, & Yuan
TVCG (InfoVis) ‘14

SketchStory
Lee, Kazi & Smith
TVCG (InfoVis) ‘13

Lyra
Satyanarayan & Heer
EuroVis ‘14
Key Open Problems in Visualization

1. Assess and communicate value
2. Make the construction of visualizations easier
3. Address real world “big” problems
**Take Aways**

Understand the purpose

Presentation & analysis, related but different

Interaction provides power, use it

Explain the value  \((T + I + E + C)\)
Thanks for your attention
Acknowledgments

• Supported by DARPA’s XDATA program

• Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537
The Value of Visualization for Exploring and Understanding Data

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

Boeing Data Analytics Community of Excellence
Data
Data Visualization
Making pretty pictures?
Data Visualization

Making pretty pictures
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process
  Gain an **understanding**
Challenge
Transform the *data* into *information* (understanding, insight) thus making it useful to people
Visualization
“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visuals help us think
Provides a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Let Me Illustrate
An example...
### Questions:

- Which cereal has the most/least potassium?
- Is there a relationship between potassium and fiber?
- If so, are there any outliers?
- Which manufacturer makes the healthiest cereals?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cereal</td>
<td>Manufacturer</td>
<td>Fiber</td>
<td>Potassium</td>
</tr>
<tr>
<td>2 100% Bran</td>
<td>N</td>
<td>10</td>
<td>280</td>
</tr>
<tr>
<td>3 100% Natural Bran</td>
<td>Q</td>
<td>2</td>
<td>135</td>
</tr>
<tr>
<td>4 All-Bran</td>
<td>K</td>
<td>9</td>
<td>320</td>
</tr>
<tr>
<td>5 All-Bran with Extra Fiber</td>
<td>K</td>
<td>14</td>
<td>330</td>
</tr>
<tr>
<td>6 Almond Delight</td>
<td>R</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7 Apple Cinnamon Cheerios</td>
<td>G</td>
<td>1.5</td>
<td>70</td>
</tr>
<tr>
<td>8 Bran Chex</td>
<td>R</td>
<td>4</td>
<td>125</td>
</tr>
<tr>
<td>9 Bran Flakes</td>
<td>P</td>
<td>5</td>
<td>190</td>
</tr>
<tr>
<td>10 Cap’n’Crunch</td>
<td>Q</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>11 Cheerios</td>
<td>G</td>
<td>2</td>
<td>105</td>
</tr>
<tr>
<td>12 Cocoa Puffs</td>
<td>G</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>13 Corn Chex</td>
<td>R</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>14 Corn Flakes</td>
<td>K</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>15 Count Chocula</td>
<td>G</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>16 Cracklin’ Oat Bran</td>
<td>K</td>
<td>4</td>
<td>160</td>
</tr>
<tr>
<td>17 Cream of Wheat (Quick)</td>
<td>N</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18 Crispy Wheat &amp; Raisins</td>
<td>G</td>
<td>2</td>
<td>120</td>
</tr>
<tr>
<td>19 Double Chex</td>
<td>R</td>
<td>1</td>
<td>80</td>
</tr>
<tr>
<td>20 Froot Loops</td>
<td>K</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>21 Frosted Flakes</td>
<td>K</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>22 Fruit &amp; Fibre Dates, Wal</td>
<td>P</td>
<td>5</td>
<td>200</td>
</tr>
<tr>
<td>23 Fruitful Bran</td>
<td>K</td>
<td>5</td>
<td>190</td>
</tr>
<tr>
<td>24 Fruity Pebbles</td>
<td>P</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>25 Golden Grahams</td>
<td>G</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>26 Grape Nuts Flakes</td>
<td>P</td>
<td>3</td>
<td>85</td>
</tr>
<tr>
<td>27 Honey Nut Cheerios</td>
<td>G</td>
<td>1.5</td>
<td>90</td>
</tr>
<tr>
<td>28 Honey-comb</td>
<td>P</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>29 Just Right Fruit &amp; Nut</td>
<td>K</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>30 Life</td>
<td>Q</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>31 Lucky Charms</td>
<td>G</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>32 Maypo</td>
<td>A</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>33 Muesli Raisins, Dates, &amp; Walnuts</td>
<td>R</td>
<td>3</td>
<td>170</td>
</tr>
<tr>
<td>34 Multi-Grain Cheerios</td>
<td>G</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>35 Nutri-Grain Almond-Raisins</td>
<td>K</td>
<td>3</td>
<td>130</td>
</tr>
<tr>
<td>36 Nutri-grain Wheat</td>
<td>K</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>37 Oatmeal Raisin Crisp</td>
<td>G</td>
<td>1.5</td>
<td>120</td>
</tr>
<tr>
<td>38 Post Nat. Raisin Bran</td>
<td>P</td>
<td>6</td>
<td>260</td>
</tr>
<tr>
<td>39 Product 19</td>
<td>K</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>40 Quaker Oatmeal</td>
<td>Q</td>
<td>2.7</td>
<td>110</td>
</tr>
<tr>
<td>41 Raisin Bran</td>
<td>K</td>
<td>5</td>
<td>240</td>
</tr>
<tr>
<td>42 Raisin Nut Bran</td>
<td>G</td>
<td>2.5</td>
<td>140</td>
</tr>
<tr>
<td>43 Rice Krispies</td>
<td>K</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>44 Shredded Wheat</td>
<td>N</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>45 Shredded Wheat ‘n’Bran</td>
<td>N</td>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>46 Shredded Wheat spoon</td>
<td>N</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>47 Smacks</td>
<td>K</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>48 Special K</td>
<td>K</td>
<td>1</td>
<td>55</td>
</tr>
<tr>
<td>49 Strawberry Fruit Wheats</td>
<td>N</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>50 Total Corn Flakes</td>
<td>G</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>51 Total Raisin Bran</td>
<td>G</td>
<td>4</td>
<td>230</td>
</tr>
<tr>
<td>52 Total Whole Grain</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>53 Trix</td>
<td>G</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>54 Wheaties</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>55 Wheaties Honey Gold</td>
<td>G</td>
<td>1</td>
<td>60</td>
</tr>
</tbody>
</table>
Even Tougher?

What if you could only see one cereal’s data at a time? (e.g. some websites)
What if I read the data to you?
Part of our Culture

“I see what you’re saying”
“Seeing is believing”
“A picture is worth a thousand words”
Academic Communities

Visualization
“Data visualization”

Scientific visualization

Information visualization
Scientific Visualization (SciVis)

Primarily relates to and represents something spatial, physical or geometric

Often 3-D

Examples

- Air flow over a wing
- Stresses on a girder
- Torrents inside a tornado
- Organs in the human body
- Molecular bonding

Not my research focus
Information Visualization (InfoVis)

Visualizing non-spatial data: Items, entities, things which do not have a direct physical correspondence

- Notion of abstractness of the entities is important too
- Examples: baseball statistics, stock trends, connections between criminals, car attributes...

“A key challenge in information visualization is designing a cognitively useful spatial mapping of a dataset that is not inherently spatial and accompanying the mapping by interaction techniques that allow people to intuitively explore the dataset. Information visualization draws on the intellectual history of several traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, statistical graphics, cartography, and art.”

http://conferences.computer.org/infovis/
Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces

Available at http://nvac.pnl.gov/ in PDF form
Applications of Visualization

Presentation
Analysis

Explanatory
Exploratory
1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Infographics

All the rage...
THE NUCLEAR ARMS RACE
It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE
First Atomic Bomb on Hiroshima
First American H Bomb
First U.S. missile submarine launched
SALT I talks on reducing nuclear weapons
SALT II talks on reducing nuclear weapons
Reykjavik summit on reducing nuclear weapons
First Soviet A Bomb
First Soviet H Bomb
USA and USSR InterContinental Ballistic Missiles program
USA withdraw from SALT II
USSR collapses

NUCLEAR STOCKPILE

ROCKET MODELS
SS-9 (USSR)
Titan II (USA)

1966
Year
Warhead
9 Mt
10 Mt
Operational range
16,000 km
15,000 km

A WORLD BREAK IN TWO
NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.

Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva
Steroids or Not, the Pursuit Is On

Barry Bonds is aiming at the career home run record. He needs only six more to be Babe Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons

Babe Ruth
714 homers
22 seasons

Barry Bonds
700 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons
Ruth 567
Aaron 554

Ruth
Actual home runs vs.
projected home runs for
five seasons

Aaron
Actual home runs
slightly outweigh
projected home runs for
five seasons

Bonds
Home run pace at age 35
is near 53 home runs
per season

According to allegations
in a book about Bonds,
began taking steroids
before the 1990 season;
hit 465 in 1996.

Homer Pace After Age 34
If the allegations are correct, Bonds was 34 in his first season on steroids.
Here are projected home run paces for each player after age 34.

29 seasons
Bonds was injured last season. He played 14 games and hit 6 home runs.

Differing Paths to the Top of the Charts
The top seven players on the career home run list, along with a look at Griffey (11th), Rodriguez (37th) and Pujols (9th).

Harry Aaron
23

Babe Ruth
22

Barry Bonds
20

Willie Mays
660

Sandy Sosa
687

Frank Robinson
566

Mark McGwire
583

Ken Griffey Jr.
516

Alex Rodriguez
425

Albert Pujols
361

Note: Ages as of July 1 of each season.
Where We Live...

Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie.

Our families are getting smaller—with one vital exception. Compared with those of Europe and Japan, the U.S. population is younger and more colorful because of the continued arrival of immigrants and their higher-than-average birthrates. Of the 109 million Americans who will join us in the next 37 years, half will be immigrants or their children. In the next few decades, 97% of the world’s population growth will occur in the developing world; the U.S. is the largest developed country in the world that is still growing at a healthy clip. That matters, strategically, economically.

THE ALMIGHTY DOLLAR
MAPPING DISTRIBUTION OF INCOME BY RELIGIOUS BELIEF

It's no secret that the distribution of wealth is inequitable in the United States across racial, regional, and socio-economic groups. But there is a distinct variance among and within America’s faiths as well. Displayed below are the income levels of America’s major religious groups, as compared to the average U.S. income distribution.

INCOME BRACKETS:

- $100,000 +
- $75,000 to $99,999
- $50,000 to $74,999
- $30,000 to $49,999
- LESS THAN $30,000

http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg
Frequent presentation goals

- Clarify
- Focus
- Highlight
- Simplify
- Persuade

May just show a few variables and/or a subset of the data cases
Hans Rosling
Gapminder

Gore made extensive use of data graphics
2. Analysis
Explore the data
Assess a situation
Determine how to proceed
Decide what to do
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data…To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland

*The Elements of Graphing Data*
Frequent analysis goals
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility
Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
So what is visualization most useful for?
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context
Investigative analysis

Intelligence analysts must explore and evaluate volumes of data, from narrative recordings of field agents to open-source news articles. Insights from visual analytics projects and a hyperscalar context show the potential of visual analytics to aid these investigations.

The analysis of overwhelming masses of disparate, conflicting, and incomplete information is central to identifying, organizing, and using data for intelligence analysis. The human process engages human capabilities to make the best possible sense of ambiguous, incomplete, and potentially erroneous information in the face of rapidly changing events. In order to inform the narrative and discover the context.

The investigative process involves the combination of intelligence sources and the use of analytic tools. The investigational process involves the combination of intelligence sources and the use of analytic tools. The investigational process involves the combination of intelligence sources and the use of analytic tools. The investigational process involves the combination of intelligence sources and the use of analytic tools.
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction
Interaction is Vital

Engage in a dialog with your data

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

Yi, Kang, Stasko, Jacko
TVCG (InfoVis) ‘07
3. **Visualization** rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Some examples from my group’s research
Web-based airline booking requests

Help recognize irregularities and analyze what the problem is

http://www.cc.gatech.edu/gvu/ii/selltrend/

Liu, Sullivan, Stasko
TVCG (InfoVis) ‘09
OnSet

Demo

Sadana, Major, Dove, & Stasko

http://www.cc.gatech.edu/gvu/ii/setvis/
NBA Player and Team Data

http://www.cc.gatech.edu/gvu/ii/sportvis/nbaVis/

Project by Fengbo Li

Demo
# PGA Tour Season Recap

Using ShotLink data

<table>
<thead>
<tr>
<th>Event</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Round 4</th>
<th>Position</th>
<th>Total</th>
<th>Strokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humana Challenge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T37</td>
<td>-17</td>
<td>271</td>
</tr>
<tr>
<td>Farmers Insurance Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T51</td>
<td>E</td>
<td>288</td>
</tr>
<tr>
<td>Phoenix Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-28</td>
<td>256</td>
</tr>
<tr>
<td>Pebble Beach Pro-Am</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T60</td>
<td>-1</td>
<td>285</td>
</tr>
<tr>
<td>Northern Trust Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T21</td>
<td>-3</td>
<td>281</td>
</tr>
<tr>
<td>WGC-Cadillac Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T3</td>
<td>-14</td>
<td>274</td>
</tr>
<tr>
<td>Arnold Palmer Invitational</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CUT</td>
<td>+8</td>
<td>152</td>
</tr>
<tr>
<td>Houston Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T16</td>
<td>-10</td>
<td>278</td>
</tr>
<tr>
<td>The Masters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T54</td>
<td>+9</td>
<td>297</td>
</tr>
<tr>
<td>Wells Fargo Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>-7</td>
<td>281</td>
</tr>
<tr>
<td>PLAYERS Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CUT</td>
<td>+1</td>
<td>145</td>
</tr>
<tr>
<td>St. Jude Classic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T2</td>
<td>-10</td>
<td>270</td>
</tr>
<tr>
<td>U.S. Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T2</td>
<td>+3</td>
<td>283</td>
</tr>
<tr>
<td>The Greenbrier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CUT</td>
<td>+2</td>
<td>142</td>
</tr>
<tr>
<td>Open Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-3</td>
<td>281</td>
</tr>
<tr>
<td>Bridgestone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T21</td>
<td>+1</td>
<td>281</td>
</tr>
<tr>
<td>PGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T72</td>
<td>+12</td>
<td>292</td>
</tr>
<tr>
<td>Barclays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T6</td>
<td>-9</td>
<td>275</td>
</tr>
<tr>
<td>Deutsche Bank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T41</td>
<td>-9</td>
<td>276</td>
</tr>
<tr>
<td>BMW Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T33</td>
<td>-1</td>
<td>283</td>
</tr>
</tbody>
</table>

with Anand Sainath, Rahul Basole
Jigsaw
Computational analysis of documents’ text
  Entity identification, document similarity, clustering, summarization, sentiment
Multiple visualizations of documents, analysis results, entities, and their connections
Views are highly coordinated

“Putting the pieces together”
Help “investigators” explore, analyze and understand large document collections

Academic Papers

Product Reviews

Health Forums

Police Reports

2010 Hyundai Genesis Sedan - Consumer Reviews

Great value and excellent performance

Vehicle: 2010 Hyundai Genesis 3.8-L4 Dr. Sedan (3.8L 6cyl 6A) / 4.6-L8cyl 6A

Review:

After driving a few cars in the US, I decided to go for a change due to price. So I have compared the prices of the recalls at the time I bought the car. I bought a Genesis for a test drive. I went to the dealership to get the price. I was able to get the price I wanted. I am very satisfied with the car. I love it. I have been driving it for a few months and I love it. I would highly recommend this car to anyone. It is very quiet on the road. It is a very comfortable car.

Favorite Features

Solid driving experience. Nicely designed.

Suggested Improvements

I would like to see more comfort in the seat. It should be fully powered so it can be raised or lowered.

Recommend

5 stars
Jigsaw

Visual analytics of large text document collections

Görg, Liu, Kihm, Choo, Park, & Stasko

TVCG ‘13
Example

Reviews of wines from Washington state from ‘07-’10

Text: review narrative

Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, ...

Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured

1323 reviews

from database of 150,000 reviews
Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw
Applications

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* ‘08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES ‘10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR ‘10
- Business Intelligence
  - Patents, press releases, corporate agreements, ...
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis ‘10
Moving to Tablets

Scatterplot

Video

Sadana & Stasko
AVI ‘14

http://www.cc.gatech.edu/gvu/ii/touch/
Different benefits of visualization
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
CS 7450 - Information Visualization

Instructor: John Stasko
Fall 2013
Mon, Wed 3:00 - 4:30 pm
Whitaker Bldg, room 1103

Information visualization is a research area that focuses on the use of visualization techniques to help people understand and analyze data. While fields such as scientific visualization involve the presentation of data that has some physical or geometric correspondence, information visualization focuses on abstract data without such correspondences such as symbolic, tabular, networked, hierarchical, or textual information sources.

The objectives of the course are

- Learn the principles involved in information visualization
- Learn about the variety of existing techniques and systems in information visualization
- Develop skills in critiquing different visualization techniques as applied to particular tasks
- Learn how to evaluate visualization systems
- Gain a background that will aid the design of new, innovative visualizations

The course will follow a lecture/seminar style with much discussion of assigned readings, as well as viewing of videos and hands-on experience with research and commercial information visualization tools.

We will be reading recent research papers about the different course topics. In addition, we will be using one book for the course: Now You See It by Stephen Few, Analytics Press 2009. Also highly recommended is Envisioning Information by Edward Tufte, Graphics Press 1990.
# Syllabus

Link to last year's 2012 syllabus with lecture slides

## Overview

<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Topic</th>
<th>Topic</th>
<th>HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aug 19, 21</td>
<td>Introduction</td>
<td>InfoVis overview</td>
<td>HW 1</td>
</tr>
<tr>
<td>2</td>
<td>Aug 26, 28</td>
<td>Visual perception</td>
<td>Value/Benefits of visualization</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Sep 2, 4</td>
<td>No Class -- Labor Day</td>
<td>Multivariate data &amp; table/graph design</td>
<td>HW 2</td>
</tr>
<tr>
<td>4</td>
<td>Sep 9, 11</td>
<td>Few’s design guidance</td>
<td>Multivariate visual representations 1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Sep 16, 18</td>
<td>Multivariate visual representations 2</td>
<td>Tasks and analysis</td>
<td>HW 3a HW 3b</td>
</tr>
<tr>
<td>6</td>
<td>Sep 23, 25</td>
<td>InfoVis systems &amp; toolkits</td>
<td>Commercial systems demos</td>
<td>HW 4</td>
</tr>
<tr>
<td>7</td>
<td>Sep 30, Oct 2</td>
<td>Storytelling</td>
<td>Tufte’s design principles</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Oct 7, 9</td>
<td>Poster session</td>
<td>Casual InfoVis</td>
<td>HW 5</td>
</tr>
<tr>
<td>9</td>
<td>Oct 14, 16</td>
<td>No Class - Fall break</td>
<td>No Class - VIS Conference</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Oct 21, 23</td>
<td>Graphs and networks 1</td>
<td>Graphs and networks 2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Oct 28, 30</td>
<td>Hierarchies &amp; trees 1</td>
<td>Hierarchies &amp; trees 2</td>
<td>HW 6</td>
</tr>
<tr>
<td>12</td>
<td>Nov 4, 6</td>
<td>Interaction</td>
<td>Overview &amp; detail</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Nov 11, 13</td>
<td>Text &amp; documents 1</td>
<td>Text &amp; documents 2</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Nov 18, 20</td>
<td>Visual analytics 1</td>
<td>Visual analytics 2</td>
<td>HW 7</td>
</tr>
<tr>
<td>15</td>
<td>Nov 25, 27</td>
<td>Time series data</td>
<td>Project work day</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Dec 2, 4</td>
<td>Evaluation</td>
<td>Review</td>
<td></td>
</tr>
</tbody>
</table>

## Detail

Below are summaries of all the materials for each class. Lecture slides, references for readings and articles, videos shown, and software demonstrated are listed for each topic. Also available is a more general, alphabetical bibliography.
For more, my EuroVis ‘14 Capstone Talk

http://vimeo.com/98986594
Take Aways
Presentation & analysis

Exploratory data analysis, power from interaction

Developing new questions and insights
Acknowledgments

• Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA program
My Research

http://www.cc.gatech.edu/gvu/ii

The Information Interfaces Group, an HCI research group in the School of Interactive Computing at Georgia Tech, develops computing technologies that help people take advantage of information to enrich their lives.

More about the lab approach

Projects

- **Jigsaw**
  Using visualization and visual analytics to help analysis and sensemaking on text document collections.

- **CiteVis**
  Exploring conference paper citation on data visually.

- **Touch Interaction**
  Developing interaction techniques for visualization on touch-based tablet and mobile devices.

- **Social Media Visual Analytics**
  Helping people analyze and understand social media through interactive visualization and other techniques.

- **Behavis**
  Using visual analytics to help psychologists explore social and communicative behaviors.

- **Rosemary**
  Building novel interactive visualizations without programming.

- **SetVis**
  Exploring boolean set data through visualization and direct manipulation.

- **Sports Data Visualization**
  Visualizing sports data to help people understand, analyze, and predict sporting events.

Hot News

John was recently named to be an IEEE Fellow, effective Jan. 2014.

In November, John was appointed as an Honorary Professor in the School of Computer Science at the Univ of St. Andrews in Scotland.

We presented a paper about visualizing social and communicative behaviors at VHC ’13 in November.

Lab alum Zhicheng “Leo” Liu’s PhD thesis was selected as runner up in the VGTSC Visualization Pioneer Group’s 2013 dissertation awards. See photo.

A journal paper about how Jigsaw combines computational text analysis with interactive visualization appeared in the Oct. ’13 IEEE TVCG issue, and was presented at one of the VIS ’13 TVCG sessions.

We hosted IEEE VIS ’13 here in Atlanta in October. (See photos) Thanks to all the attendees for the positive feedback!

We had a paper about dotlink360 appear at InfoVis ’13 this fall, and we presented 4 other posters at VIS as well.

We co-organized the Public Health Data Visualization and Sports Data Visualization workshops at VIS this fall.

Lab - Visit our lab

CS 7450 - InfoVis Class
VACCINE Center
FOODava Center
The Value of Visualization for Understanding Data and Making Decisions

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

JISIC 2014
Data
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
Data Visualization

Making pretty pictures?
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process

Gain an **understanding**
Visualization

“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visuals help us think
Provide a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Applications of Visualization
Presentation
Analysis
1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Infographics
THE NUCLEAR ARMS RACE
It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE
- First Atomic Bomb on Hiroshima: 1945
- First American H Bomb: 1949
- First U.S. missile submarine launched: 1952
- SALT I talks on reducing nuclear weapons: 1960
- SALT II talks on reducing nuclear weapons: 1979
- Reykjavik summit on reducing nuclear weapons: 1986
- USA and USSR InterContinental Ballistic Missiles program: 1991
- USA withdraw from SALT II: 1987
- USSR collapse: 1991

NUCLEAR STOCKPILE

ROCKET MODELS
- SS-9 (USSR): 1966
- Titan II (USA): 1962
- Warhead: 10 Mt (SS-9), 9 Mt (Titan II)
- Operational range: 16,000 km (SS-9), 15,000 km (Titan II)

AIR CONTROL
- Intercontinental Ballistic Missiles
- USA: 8,000
- USSR: 7,000
- Planes
- USA: 4,000
- USSR: 5,000

A WORLD BREAK IN TWO
NATO and Warsaw Pact were both mutual defense treaties between states. The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.
Gay rights in the US, state by state

Gay rights laws in America have evolved to allow — but in some cases ban — rights for gay, lesbian and transgender people on a range of issues, including marriage, hospital visitation, adoption, housing, employment and school bullying. The handling of gay rights issues vary by state and follow trends by region.

- Obama supports same-sex marriage: share with us what it means to you
- Gay rights: five activists reflect on the history of the movement in the US
- President Obama endorses gay marriage
Simply presenting data *visually* can have a profound impact
CS 7450
Fall ‘12
Design project

Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri
Gun ownership in New York counties

The gun owner next door: What you don't know about the weapons in your neighborhood

http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1

http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood?-gcheck=1&nclick_check=1

17
Frequent presentation goals

Clarify
Focus
Highlight
Simplify

May just show a few variables and/or a subset of the data cases
2. Analysis
Explore the data
Assess a situation
Determine how to proceed
Decide what to do
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland

*The Elements of Graphing Data*
Frequent analysis goals
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
So what is visualization most useful for?
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.
Investigative analysis

Intelligence analysts must explore and evaluate volumes of data, from narrative recordings of field agents to open source news articles. Insights from visual analytics projects and a hypothetical scenario show the potential of visual analytics to aid three investigations.

Visual Analytics Support for Intelligence Analysis

Carsten Görg, University of Colorado
Yenmash Kang, Google
Zhi Cheng Liao, Stanford University
John Stasko, Georgia Tech

Much of this challenge remains, although visual analytics technologies continue to yield great promise. A few commercial visual analytics tools for intelligence analysts are emerging, such as analyst workbenches from Koreograf or open-source solutions like NextMap, which are used for seeing and visualizing data. However, most commercial technologies fall short when it comes to addressing the kinds of visualization challenges that analysts face when working with large, complex datasets.

The investigation here demonstrates an evaluation of a visualization system designed to support intelligence analysis. The system's primary feature is a graph of key network relationships and a visualization of the data that the analyst can examine.

The system is designed to support analysts working with large, complex datasets. It provides visualizations of network relationships and data that the analyst can examine, allowing them to identify patterns and make more informed decisions.

Görg, Kang, Kiu, Stasko

IEEE Computer '13
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

**Q & A dialog through interaction**
Engage in a dialog with your data

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 13, NO. 6, NOVEMBER/DECEMBER 2007

Toward a Deeper Understanding of the Role of Interaction in Information Visualization
Ji Seo Yi, Yuan-kang Kang, John T. Stasko, Member, IEEE, and Jiale A. Jacko

Abstract—Even though interaction is an important part of information visualization (InfoVis), thus far a proposal for a highly level of attention from the InfoVis community. A few frameworks and taxonomies of InfoVis interaction techniques exist, but they typically focus on formal operations and do not address the utility of benefits interaction provides. After conducting an extensive review of InfoVis systems and their interactive capabilities, we propose some general categories of interaction techniques in InfoVis: 1) Select, 2) Explore, 3) Reconfigure, 4) Encode, 5) Abstract/Elaborate, 6) Filter, and 7) Connect. These categories are organized around a user’s intent while interacting with a system rather than the pattern of interaction technique provided to a user. The categories can act as a framework to help discuss and evaluate interaction techniques, as well as an interaction technique toward a deeper understanding of the role of interaction.

Index Terms—Information visualization, interaction, interaction technique, taxonomy, visualization approach

1 INTRODUCTION

Interaction visualization (InfoVis) systems, at their core, appear to have two main components: representation and interaction. The representation component, whose work lies in the field of computer graphics, is responsible for how the data is modeled, visualized, and presented on the display. The interaction component involves the dialog between the user and the system in the user explores the data set to uncover insights. The interaction component’s work is in the area of human-computer interaction (HCI). Although discussed as two separate components, representation and interaction clearly are not mutually exclusive. For instance, interaction with a system may activate a change in representation. Alternatively, the representation serves to communicate information to the user in a way that is meaningful and user-friendly. We want to consider each such contribution to an end-user’s experience.

We use the term representation component to have received the vast majority of attention in InfoVis research. A variety of recent conferences, proceedings, or in situ issues in the area will unaware of many articles about new representation of data sets, but interaction often is the main focus of research efforts in the field, especially in the "InfoVis" track of InfoVis. In other words, it is overshadowed by far more recently representation aspects. A few papers have already focused on the interactive aspects of InfoVis (e.g., [34, 55, 47]), but these are relatively recent. As in previous papers, we make no attempt at a comprehensive survey of the more recent literature. However, one paper on the importance of exploration and interactive exploration in the design of interactive visualization systems ([34, 55, 46]) indicates that interactive exploration is of comparable importance to visualization in general (e.g., [47]). We used these data mainly for analysis and supportive purposes (e.g., [8, 11, 9]).

The main purpose of this paper is to discuss the interactive aspects of InfoVis, focusing on the issue of how the user interacts with the data, and the capabilities of the system to support this interaction. We propose a new approach to organizing the different types of interaction techniques, which we believe is needed to provide a deeper understanding of the role and utility of interaction in InfoVis. We will discuss the different types of interaction techniques and how they can be applied in InfoVis systems.

For this reason, we might object to the interaction of information itself. Is the concept of "interaction" a misunderstanding of the interaction relationship? Or is it a misunderstanding of the interaction behavior? Or is it a misunderstanding of the interaction technique? Instead, think of interaction as a dynamic query to the user of an interactive system. A user might ask the system to perform a specific task, such as filtering a data set, and the system would respond accordingly. This approach to understanding interaction is more realistic and useful for understanding the role of interaction in InfoVis.
3. **Visualization** rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Visualization of different data types

• Text & documents
• Statistics
• Financial/business data
• Internet information
• Software
• ...

Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

32
Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?
Example Tasks and Goals

- Which documents contain text on topic XYZ?
- Which documents are of interest to me?
- Are there other documents that are similar to this one (so they are worthwhile)?
- How are different words used in a document or a document collection?
- What are the main themes and ideas in a document or a collection?
- Which documents have an angry tone?
- How are certain words or themes distributed through a document?
- Identify “hidden” messages or stories in this document collection.
- Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
- Find connections between documents.
The Challenge

Text is nominal data
Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data
- Bar charts, line charts, scatterplots, etc.
“Putting the pieces together”

Jigsaw
Computational analysis of documents’ text
  Entity identification, document similarity, clustering, summarization, sentiment
Multiple visualizations of documents, analysis results, entities, and their connections
  Views are highly coordinated
Help “investigators” explore, analyze and understand large document collections

Academic Papers

Product Reviews

Health Forums

Police Reports

40

2010 Hyundai Genesis Sedan - Consumer Reviews

38
Jigsaw

Visual analytics of large text document collections

Görg, Liu, Kihm, Choo, Park, & Stasko
TVCG ‘13
Example

Made each page into a separate “document”
585 in total
Entities: Person, Location, Organization, Date, Money
Graph View
## Calendar View

<table>
<thead>
<tr>
<th>Year</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
</tr>
<tr>
<td>2005</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
</tr>
<tr>
<td>2006</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
</tr>
<tr>
<td>2007</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
</tr>
<tr>
<td>2008</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
</tr>
<tr>
<td>2009</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
</tr>
<tr>
<td>2010</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
<td>![Calendar Entry]</td>
</tr>
</tbody>
</table>

*Note: Calendar entries represent specific events or tasks.*
Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw
Applications

• Intelligence & law enforcement
  – Police cases
  – Won 2007 VAST Contest
  – Stasko et al, Information Visualization ‘08
• Academic papers, PubMed
  – All InfoVis & VAST papers
  – CHI papers
  – Görg et al, KES ‘10
• Investigative reporting
• Fraud
  – Finance, accounting, banking
• Grants
  – NSF CISE awards from 2000
• Topics on the web (medical condition)
  – Autism
• Consumer reviews
  – Amazon product reviews, edmunds.com, wine reviews
  – Görg et al, HCIR ’10
• Business Intelligence
  – Patents, press releases, corporate agreements, ...
• Emails
  – White House logs
• Software
  – Source code repositories
  – Ruan et al, SoftVis ‘10
Further benefits of visualization
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
EuroVis ‘14 Capstone Talk

http://vimeo.com/98986594
Take Aways
Presentation & analysis

Interaction provides the power

Exploring & developing questions
Acknowledgments

• Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA Program
Value-Driven Evaluation of Visualizations

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
“No user studies are reported in this paper, thus it is unclear if the visualization will be useful.”
Really?
· Which cars have the best miles-per-gallon?
· How much does a Ford Taurus cost?
· Is there a correlation between car weight and torque?
· How many different countries manufacture cars?
· What is the range of car’s horsepower?
Evaluation

Explanation(s) and Illustration(s) of value
Visualization’s Value?
Visualization’s Value?

\[ V_{\text{value}} = T + I + E + C \]
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to minimize the total **time** needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)
Visualization’s Value?

\[ V_{\text{value}} = T + I + E + C \]

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to convey an overall *essence* or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)
Visualization’s Value?

\[ \text{Value} = T + I + E + C \]

Ability to generate confidence and trust about the data, its domain and context

(Beneficial data analysis process side effects)
Map of the Market

http://www.cc.gatech.edu/gvu/ii/citevis
Data Visualization 101

Problem:
You have a lot of data (& attributes) to understand

Do you?

Pack all the data into one complex representation

Spread the data into multiple coordinated views

Use interaction to reveal different subsets of the data
Interaction is Vital

Engage in a dialog with your data

“The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means.”

S. Few, *Now you see it*
Interaction

Understand the benefits (science of interaction)

Explore new interactive operations

Visualization on other display devices (touch)
For more, see my EuroVis ‘14 Capstone Talk

http://vimeo.com/98986594
Acknowledgments

• Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA program
Visual Analytics for Investigative Analysis and Exploration of Documents and Data

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
Data Visualization
Making pretty pictures?
Data Visualization

Making pretty pictures
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process

Gain an **understanding**
Visualization
“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visuals help us think
Provide a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Applications of Visualization
Presentation
Analysis
1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Infographics
THE NUCLEAR ARMS RACE
It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE

<table>
<thead>
<tr>
<th>Event</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Atomic Bombs on Hiroshima</td>
<td>45</td>
</tr>
<tr>
<td>First American H Bomb</td>
<td>49</td>
</tr>
<tr>
<td>First Soviet A Bomb</td>
<td>52</td>
</tr>
<tr>
<td>First Soviet H Bomb</td>
<td>53</td>
</tr>
<tr>
<td>First U.S. missile submarine launched</td>
<td>57</td>
</tr>
<tr>
<td>SALT I talks on reducing nuclear weapons</td>
<td>60</td>
</tr>
<tr>
<td>USSR launched the first satellite</td>
<td>63</td>
</tr>
<tr>
<td>USA and USSR InterContinental Ballistic Missiles program</td>
<td>69</td>
</tr>
<tr>
<td>Reykjavik summit on reducing nuclear weapons</td>
<td>79</td>
</tr>
<tr>
<td>USA withdraws from SALT II</td>
<td>86</td>
</tr>
<tr>
<td>USSR collapses</td>
<td>91</td>
</tr>
</tbody>
</table>

NUCLEAR STOCKPILE

ROCKET MODELS

SS-9 (USSR)
- Year: 1966
- Warhead: 10 Mt
- Operational range: 16,000 km

Titan II (USA)
- Year: 1962
- Warhead: 9 Mt
- Operational range: 15,000 km

AIR CONTROL

Intercontinental Ballistic Missiles
- USA: 8,000
- USSR: 7,000

Planes
- USA: 4,000
- USSR: 5,000

A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states. The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.

Troops
- NATO: 2.6 million
- Warsaw Pact: 4 million

Tanks
- NATO: 13,000
- Warsaw Pact: 42,500

Artillery
- NATO: 10,750
- Warsaw Pact: 31,500

Produced by WatchTheAmericans.com a Fan Website
The Beernomic Table

http://thebeermongers.com/beers/
Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

755

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paced for each player after age 34.

- Aaron
- Ruth
- Bonds

Bonds was injured last season. He played 14 games and hit 5 homers.

Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (3rd).

NY Times 2006
Where We Live...

Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie.

80% of the U.S. population lives in a metropolitan area.

The entire state of Wyoming (pop. 505,300) has fewer people than the Manhattan, Pa., metro area.

Lyon County, Texas, is the least populated county in the lower 48 states, with 62 residents.

Alaska is the most sparsely populated state, with 1 person per square mile.

Sources: U.S. Census Bureau, LexisNexis.2009.03.
THE ALMIGHTY DOLLAR
MAPPING DISTRIBUTION OF INCOME BY RELIGIOUS BELIEF

It's no secret that the distribution of wealth is inequitable in the United States across racial, regional, and socio-economic groups. But there is a distinct variance among and within America’s faiths as well. Displayed below are the income levels of America’s major religious groups, as compared to the average U.S. income distribution.

INCOME BRACKETS:
- **$100,000 +**
- **$75,000 to $99,999**
- **$50,000 to $74,999**
- **$30,000 to $49,999**
- **LESS THAN $30,000**

A COLLABORATION BETWEEN GOOD AND COLUMN FIVE.

SOURCE:
THE PEN FORUM

http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg
Simply presenting data *visually* can have a profound impact
Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri
Gun ownership in New York counties

http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1
http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood--?gcheck=1&nclick_check=1
Frequent presentation goals
Clarify
Focus
Highlight
Simplify

May just show a few variables and/or a subset of the data cases
2. Analysis
understand, compare, decide, judge, evaluate, assess, determine, ...
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland
The Elements of Graphing Data
Frequent analysis goals
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
So what is visualization most useful for?
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

**Learning, awareness, trust, context**
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning

Use them when they get the job done!
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction
Visualization

“The use of computer-supported, interactive visual representations of data to amplify cognition”
3. **Visualization** rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

*Broad and deep understanding quickly*
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...


Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

...
Text is Everywhere

We use documents as primary information artifacts in our lives

Our access to documents has grown tremendously in recent years due to the networking infrastructure
  - WWW
  - Digital libraries
  - ...

Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?
Example Tasks and Goals

• Which documents contain text on topic XYZ?
• Which documents are of interest to me?
• Are there other documents that are similar to this one (so they are worthwhile)?
• How are different words used in a document or a document collection?
• What are the main themes and ideas in a document or a collection?
• Which documents have an angry tone?
• How are certain words or themes distributed through a document?
• Identify “hidden” messages or stories in this document collection.
• Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
• Find connections between documents.
The Challenge

Text is nominal data

Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data
- Bar charts, line charts, scatterplots, etc.
“Putting the pieces together”

Jigsaw
Computational analysis of documents’ text
   Entity identification, document similarity, clustering, summarization, sentiment
Multiple visualizations of documents, analysis results, entities, and their connections
   Views are highly coordinated
Example

Reviews of wines from Washington state from ‘07–’10

Text: review narrative

Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, ...

Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured

1323 reviews

from database of 150,000 reviews

Demo
Graph View
Calendar View

A calendar view with years 2004 to 2010 is displayed. Each month is represented with days marked by colors or symbols, indicating various events or data points. The left sidebar shows filters and keywords, with options for multi-agent systems, intelligent agent, agent-based, computer graphics, graphics, amount, cop, directorate, division, TS, field of application, organization, and Michigan Technological University.
Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw
Applications

• Intelligence & law enforcement
  – Police cases
  – Won 2007 VAST Contest
  – Stasko et al, *Information Visualization* ‘08

• Academic papers, PubMed
  – All InfoVis & VAST papers
  – CHI papers
  – Görg et al, KES ‘10

• Investigative reporting

• Fraud
  – Finance, accounting, banking

• Grants
  – NSF CISE awards from 2000

• Topics on the web (medical condition)
  – Autism

• Consumer reviews
  – Amazon product reviews, edmunds.com, wine reviews
  – Görg et al, HCIR ‘10

• Business Intelligence
  – Patents, press releases, corporate agreements, …

• Emails
  – White House logs

• Software
  – Source code repositories
  – Ruan et al, SoftVis ‘10
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

**Learning, awareness, trust, context**

Data cleaning example
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

**Q & A dialog through interaction**

Repeated interactive queries
3. **Visualization** rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Varieties, colors, reviewers, producers, concepts, ...
Insights
No great aha! – More learning & knowledge building
Domain knowledge (wines) matters
Further benefits of visualization
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
EuroVis ‘14 Capstone Talk

http://vimeo.com/98986594
Take Aways
Presentation & analysis

Interaction provides the power
Exploring & developing questions
Acknowledgments

• Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA Program
My Research

http://www.cc.gatech.edu/gvu/ii
Data Visualization Principles
(Some Things I’ve Learned)

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
Three themes
1. Understanding purpose
2. Embracing interaction
3. Identifying value
Talented students
1. Understanding Purpose

Why visualize?
Applications of Visualization

Presentation (Explanatory)
Analysis (Exploratory)
1. Presentation
Communicate data and ideas
Explain and inform
Provide evidence and support
Influence and persuade
Infographics
All the rage...
THE NUCLEAR ARMS RACE
It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE

<table>
<thead>
<tr>
<th>Event</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Atomic Bomb on Hiroshima</td>
<td>45</td>
</tr>
<tr>
<td>First American H Bomb</td>
<td>69</td>
</tr>
<tr>
<td>First U.S. missile submarine launched</td>
<td>86</td>
</tr>
<tr>
<td>USA and USSR InterContinental Ballistic Missiles program</td>
<td>91</td>
</tr>
</tbody>
</table>

NUCLEAR STOCKPILE

<table>
<thead>
<tr>
<th>Year</th>
<th>USA</th>
<th>USSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>10,000</td>
<td>5,000</td>
</tr>
<tr>
<td>1970</td>
<td>30,000</td>
<td>25,000</td>
</tr>
<tr>
<td>1980</td>
<td>50,000</td>
<td>40,000</td>
</tr>
</tbody>
</table>

ROCKET MODELS

<table>
<thead>
<tr>
<th>Model</th>
<th>Year</th>
<th>Warhead</th>
<th>Operational range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS-9</td>
<td>1966</td>
<td>10 Mt</td>
<td>16,000 km</td>
</tr>
<tr>
<td>Titan II</td>
<td>1962</td>
<td>9 Mt</td>
<td>15,000 km</td>
</tr>
</tbody>
</table>

AIR CONTROL

<table>
<thead>
<tr>
<th>Type</th>
<th>USA</th>
<th>USSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercontinental Ballistic Missiles</td>
<td>8,000</td>
<td>7,000</td>
</tr>
<tr>
<td>Planes</td>
<td>4,000</td>
<td>5,000</td>
</tr>
</tbody>
</table>

A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states. The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.

 Troops: NATO = 2.6 million, Warsaw Pact = 4 million
 Tanks: NATO = 13,000, Warsaw Pact = 42,500
 Artillery: NATO = 10,750, Warsaw Pact = 31,500

Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva
Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs

- Hank Aaron
  755 home runs
  23 seasons

- Babe Ruth
  714 home runs
  22 seasons

- Barry Bonds
  708 home runs
  20 seasons

Bonds took lead
Home runs after 16 seasons
Bonds: 567
Aaron: 554
Ruth: 519

2004

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run pages for each player after age 24.

**Aaron**
Actual home runs slightly outpace projected home runs for five seasons.

**Ruth**
Averaged 46.4 home runs a season from age 30 to 34. Averaged 46.5 for next four seasons.

**Bonds**
From age 35 to 39, he averaged 4.1 more home runs a season than projected.

According to allegations in a book about Bonds, he began taking steroids before the 1999 season. He was 14th in the league two seasons later. He hit 73 home runs, surpassing Aaron’s career pace.

Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th)

- Hank Aaron
  755
  23 seasons
  15 times hit 30 or more (M.L. most)

- Babe Ruth
  714
  22 seasons
  Hit only 20 over first five seasons

- Barry Bonds
  708
  22 seasons
  Averaged 52 from 2000 to 2004

- Willie Mays
  660
  22 seasons
  No one hit more from 1950-59

- Sammy Sosa
  689
  18 seasons
  Three 60-home run seasons is record

- Frank Robinson
  586
  16 seasons
  Triple Crown in ’66 (49, 122, 311)

- Mark McGwire
  583
  16 seasons
  First to hit 70 in a season

- Ken Griffey Jr.
  536
  17 seasons
  Only McGwire hit more in the 90’s

- Alex Rodriguez
  429
  15 seasons
  Youngest to reach 400 home runs

- Albert Pujols
  261
  14 seasons
  Second most ever in first five seasons
Hans Rosling
Gapminder

An Inconvenient Truth

Gore made extensive use of data graphics
Frequent presentation goals

Clarify
Focus
Highlight
Simplify
Persuade

May just show a few variables and/or a subset of the data cases
2. Analysis
Explore the data
Assess a situation
Determine how to proceed
Decide what to do
Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning
Thought

If you can articulate very precisely what you’re seeking, visualization likely isn’t your best approach.

OK, so what is visualization good for?
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Frequent analysis goals
Show many/all variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
At Its Heart: Uncertainty

Nothing is perfect
Humans trade off choices, compromise

![Graphs showing cost, journey time, and bedroom attributes](image)

Attribute Explorer

Spence & Tweedie
Interacting with Computers ‘98
Takeaway 1

Analysis $\neq$ Presentation
Views of the World

Vis researchers

- Analysis
- Presentation

Everyone else

- Analysis
- Presentation
Narrative Visualization: Telling Stories with Data

Edward Segel and Jeffrey Heer

Abstract—Data visualization is rapidly growing in importance due to its ability to reveal stories within data, yet these "data stories" differ from traditional forms of storytelling. Storytellers, especially online journalists, have increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function as a piece of the story itself. In this paper, we systematically review the design space of this emerging class of visualizations. Drawing on case studies from news media to journalism research, we identify distinct genres of narrative visualization. We characterize these design differences, together with interactivity and aesthetics, in terms of the balance between the narrative flow intended by the author (imposed by graph structure and the narrative structure) and story discovery on the part of the reader (driven through navigational exploration. Our framework suggests design strategies for narrative visualization, including presenting user-defined visualizations to journalists, storytelling, and educational modules.

Index Terms—Narrative visualization, storytelling, design methods, case study, journalism, social data analysis.

1 INTRODUCTION

In recent years, many have commented on the storytelling potential of data visualization. News organizations including the New York Times, Washington Post, and the Guardian regularly incorporate dynamic graphics into their journalism. Politicians, activists, and television reporters use interactive visualizations as a backdrop for stories about global health and economics [10] and election results [9]. A recent frame in The Economist [6] explores the proliferation of digital data and notes that visualization designers are "mastering the skills of a computer scientist, statistician, artist, designer and writer."

Static visualizations have long been used to support storytelling, usually in the form of diagrams and charts embedded in a larger body of text. In this format, the text conveys the story, and the image typically provides supporting evidence or related details. An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives.

Crafting successful "data stories" requires a diverse set of skills. Goodwin and Page [12] note that effective story-telling "requires" skills that "are familiar to some directors, beyond a technical expert's knowledge of computer engineering and science." While techniques from motion, prose, comic books, video games, and film production are applicable to narrative visualization, we should also expect this emerging medium to possess unique attributes. Data stories differ in important ways from traditional storytelling. Stories in text and film typically present a set of events in a tightly controlled progression. While tours through visualized data can be organized in a linear sequence, they can also be interactive, allowing verification, new questions, and alternative explanations.

Currently, most sophisticated visualization tools focus on data exploration and analysis. Applications such as spreadsheets and visualization tools support an array of analysis routines and visualization tools, but beyond exporting images for presentation typically provide some support for crafting stories with analysis results. As such, they provide powerful vehicles for discovering "stories" but do little to aid the communication of those findings to others. As tools mature and more richly integrate with the websites, with many authors (e.g., Tableau Public [22], GeoTime Stories [8]), they are enabling the production of narrative graphics with variable constrained levels of interactivity. It remains an open question how the design of such tools might be evolved to support richer and more diverse forms of storytelling.

2 RELATED WORK

Storytelling and visual expression are integral parts of human culture; storytelling has even been referred to as "the world's second-oldest profession" [12]. Without summarizing milestones of achievement, we describe a few of the key concepts informing narrative visualization.

2.1 Narrative Structure

The Oxford English Dictionary defines narrative as an "account of a series of events, facts, etc., given in order and with the establishing of connections between them." Central to this definition is the notion of a chain of causally related events. Stories of this form often have a beginning, middle, and end [2, 24]; an introduction to the situation, a series of events often involving tension, conflict, and a resolution. Since ancient times, people have tried to understand and formalize the elements of storytelling. For example, writers (e.g., [5, 15, 21]) have developed tropes of dramatic situations and identified plot lines common to many narratives, such as the "hero's journey" [5].

This research typically distinguishes between the content of the story and the form in which it is told. While stories often contain interacting characters, they may also present a sequence of facts and observations linked together by a satisfying theme or narrative.

Storytelling strategies vary among media and genres. For instance, stories told through writing have access to a different set of formal mechanisms and narrative structures (e.g., streams of consciousness) than stories told through film (e.g., screen sequences) [5]. Bramhall [2] describes narrative devices for journalism such as the accumu-
2. Embrace Interaction

Static ok, Interactive better
Data Visualization 101

Problem:
You have a lot of data (& attributes) to understand

Do you?

Pack all the data into one complex representation

Spread the data into multiple coordinated views

Use interaction to reveal different subsets of the data
Constituents

Two key aspects of data visualization

- Representation
- Interaction

“The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means.”

S. Few, *Now you see it*
Interaction is Vital

Engage in a dialog with your data
Interaction
Why interact?

Intent

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

Yi, Kang, Stasko & Jacko
TVCG (InfoVis) ‘07
Interaction
How manifested today?

Tooltips & selection
Get details
Navigation
Brushing & linking
Interaction
Can we do more?

Employ interaction in a more fundamental manner to strengthen the power of visualization

Some examples...
Node-link graphs

http://www.visualcomplexity.com/vc/
CiteVis

Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster ‘13
Visualization design: Panpan Xu, Chad Stolper, Anand Sainath
http://vislists.anandsainath.com/list/vis25

<table>
<thead>
<tr>
<th>Author</th>
<th>Conference</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roth, S.F. 6</td>
<td>InfoVis</td>
<td>1990</td>
</tr>
<tr>
<td>Chuah, M.C. 4</td>
<td>SciVis</td>
<td>1991</td>
</tr>
<tr>
<td>Thomas, J. 3</td>
<td>VAST</td>
<td>1992</td>
</tr>
<tr>
<td>Munzner, T. 3</td>
<td></td>
<td>1993</td>
</tr>
<tr>
<td>Keahey, T.A. 3</td>
<td></td>
<td>1994</td>
</tr>
<tr>
<td>Sprenger, T.C. 2</td>
<td></td>
<td>1995</td>
</tr>
<tr>
<td>Carriere, J. 2</td>
<td></td>
<td>1996</td>
</tr>
<tr>
<td>Gross, M.H. 2</td>
<td></td>
<td>1997</td>
</tr>
<tr>
<td>Riedl, J. 2</td>
<td></td>
<td>1998</td>
</tr>
<tr>
<td>Ed Huai-Hsin Chi 2</td>
<td></td>
<td>1999</td>
</tr>
<tr>
<td>Wills, G.J. 2</td>
<td></td>
<td>2000</td>
</tr>
<tr>
<td>Derthick, M. 2</td>
<td></td>
<td>2001</td>
</tr>
<tr>
<td>Becker, B. 2</td>
<td></td>
<td>2002</td>
</tr>
<tr>
<td>Whitney, P. 2</td>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>Pang, A. 2</td>
<td></td>
<td>2004</td>
</tr>
<tr>
<td>Fracchia, F.D. 2</td>
<td></td>
<td>2005</td>
</tr>
<tr>
<td>Eick, S.G. 2</td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Inselberg, A. 2</td>
<td></td>
<td>2007</td>
</tr>
<tr>
<td>Kolojejchick, J. 2</td>
<td></td>
<td>2008</td>
</tr>
</tbody>
</table>
OnSet

http://www.cc.gatech.edu/gvu/ii/setvis/

Demo

Sadana, Major, Dove, & Stasko

TVCG (InfoVis) ‘14
Interaction
What are the tools of interaction?

Traditional – Desktop: keyboard, mouse
New – Tablet: fingers and multi-touch
Moving to Tablets

Scatterplot

Video

Sadana & Stasko
AVI ‘14

http://www.cc.gatech.edu/gvu/ii/touch/
Scatterplot touch video
Dust and Magnet

Yi, Melton, Stasko & Jacko

Information Visualization ‘05
Pushing it Further

Perceptive Pixel D n M
Takeaway 2

Take advantage of interaction
3. Identifying Value

Visualization’s Value
The Value of Information Visualization

Jean-Daniel Fekete¹, Jarke J. van Wijk², John T. Stasko³, and Chris North⁴

¹ Université Paris-Sud, INRIA, Bât 490, F-91405 Orsay Cedex, France, Jean-Daniel.Fekete@inria.fr, http://www.aviz.fr/~fekete/

² Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB EINDHOVEN, The Netherlands, vanwijk@win.tue.nl, http://www.win.tue.nl/~vanwijk/

³ School of Interactive Computing, College of Computing & GVU Center, Georgia Institute of Technology, 85 5th St., NW, Atlanta, GA 30332-0760, USA, stasko@cc.gatech.edu, http://www.cc.gatech.edu/~john.stasko

⁴ Dept of Computer Science, 2202 Kraft Drive, Virginia Tech, Blacksburg, VA 24061-0106, USA, north@vt.edu, http://people.cs.vt.edu/~north/

Abstract. Researchers and users of Information Visualization are convinced that it has value. This value can easily be communicated to others in a face-to-face setting, such that this value is experienced in practice. To convince broader audiences, and also, to understand the intrinsic qualities of visualization is an important endeavor. In this paper we review

Information Visualization: Human-Centered Issues and Perspectives
Springer, 2008
Simply presenting data *visually* can have a profound impact
My Class

http://www.cc.gatech.edu/~stasko/7450

CS 7450 - Information Visualization

Instructor: John Stasko
Fall 2013
Mon,Wed 3:00 - 4:30 pm
Whitaker Bldg. room 1103

Information visualization is a research area that focuses on the use of visualization techniques to help people understand and analyze data. While fields such as scientific visualization involve the presentation of data that has some physical or geometric correspondence, information visualization focuses on abstract data without such correspondences such as symbolic, tabular, networked, hierarchical, or textual information sources.

The objectives of the course are

- Learn the principles involved in information visualization
- Learn about the variety of existing techniques and systems in information visualization
- Develop skills in critiquing different visualization techniques as applied to particular tasks
- Learn how to evaluate visualization systems
- Gain a background that will aid the design of new, innovative visualizations

The course will follow a lecture/seminar style with much discussion of assigned readings, as well as viewing of videos and hands-on experience with research and commercial information visualization tools.

We will be reading recent research papers about the different course topics. In addition, we will be using one book for the course: *Now You See It* by Stephen Few, Analytics Press 2009. Also highly recommended is *Envisioning Information* by Edward Tufte, Graphics Press 1990.
Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Visualization’s Value?
Visualization’s Value?

\[ V_{\text{value}} = T + I + E + C \]
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to minimize the total **time** needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)
What kinds of questions?

“Low-level” tasks
Retrieval value
Filter
Compute derived value
Find extremum
Sort
Determine range
Characterize distribution
Find anomalies
Cluster
Correlate
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)
What is Insight?

An individual observation about the data by the participant, a unit of discovery

Complex
Deep
Qualitative
Relevant
**Unexpected**

Saraiya, North, & Duca
TVCG ‘05

North
IEEE CG&A ‘06
What is Insight?

Sudden grasp of new relationships that are necessary to solve a problem and that were not learned in the past

Bernstein, Penner, Clarke-Stewart & Roy
*Psychology, 6th edition*
What is Insight?

Is not spontaneous “aha!” moments (eg, in cognitive science)
Is knowledge-building and model-confirmation
  Like a substance that people acquire with the aid of systems

Chang, Ziemkiewicz, Green, & Ribarsky

*IEEE CG&A ‘09*
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to convey an overall essence or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)
# Table Lens: Baseball Player Statistics

Calculate: \(\text{Hits} / \text{At Bats} = \text{Avg}\)

<table>
<thead>
<tr>
<th>Player</th>
<th>Avg</th>
<th>Career Avg</th>
<th>Team</th>
<th>Salary 97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larry Herndon</td>
<td>0.24734903</td>
<td>0.27282076</td>
<td>Det.</td>
<td>225</td>
</tr>
<tr>
<td>Jesse Barfield</td>
<td>0.2806248</td>
<td>0.27630189</td>
<td>Tor.</td>
<td>1237.5</td>
</tr>
<tr>
<td>Jeffrey Leonar</td>
<td>0.27859238</td>
<td>0.27260458</td>
<td>S. P.</td>
<td>900</td>
</tr>
<tr>
<td>Donnie Hill</td>
<td>0.28318584</td>
<td>0.2725564</td>
<td>Oak.</td>
<td>275</td>
</tr>
<tr>
<td>Billy Sample</td>
<td>0.285</td>
<td>0.2718061</td>
<td>Atl.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

| Howard Johnson | 0.24545455 | 0.2523068 | N. Y. | 297.5     |
| Andre Thomas   | 0.250774  | 0.2561994 | Atl.  | 75        |
| Billy Hatcher  | 0.2575656 | 0.25211507 | Hou. | 110       |
| Omar Moreno    | 0.2334833 | 0.2518029 | Atl.  | N/A       |
| Earnell Coles  | 0.2725523 | 0.25153375 | Det. | 105       |

Rao & Card
CHI ‘94
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to generate confidence and trust about the data, its domain and context

(Beneficial data analysis process side effects)
Some examples
1.

E. Tufte  
*The Visual Display of Quantitative Information*  
1983
Data Values

365  High temp for each day
365  Low temp for each day
365  Avg high temp for each day
365  Avg low temp for each day
365  Precipitation for each day
365  Humidity for each day
12   Precipitation for each month
12   Avg precipitation for each month
   1  Precipitation for the year
   1  Avg precipitation per year
   1  Highest temp (& day) for the year
   1  Lowest temp (&day) for the year
   1  Avg daily temp for the year
   1  Avg daily temp per year
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{alue}} = T + I + E + C \]
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
A month-by-month look at Atlanta's weather in 2013

Atlanta’s 2013 weather can be summed up in one word – soggy. The city ended the year 16.31 inches above normal in rainfall, with much of that rain falling during the waterlogged summer months, which had about twice as many rainy days as dry ones. Rain fell into the city’s official gauge at Hartsfield-Jackson International Airport on 17 days in June, 19 days in July and 24 days in August. The rainfall total for those three months was 23.29 inches, making the period the city’s fourth wettest summer on record. Daily rainfall records were set on April 28 (1.73 inches), May 4 (3.40 inches), June 5 (4.14 inches), Aug. 8 (2.60 inches) and Oct. 8 (2.12 inches). Summer 2013 will also be remembered for a lack of extreme heat, particularly compared to the sweltering summer of 2012. Temperatures climbed into the 90s only 20 times, topping out at 92 on June 28 and Aug. 29. That’s 32 less days of 90+ heat than in 2012, when the all-time record high of 106 was reached.

**Daily temperatures**

*Atlanta saw record highs twice in 2013 – on Jan. 12 and Dec. 22, when temperatures reached 76 and 71, respectively.*

**Monthly rainfall**

Sources: National Weather Service, National Drought Mitigation Center

---

Atlanta Journal Constitution
Jan. 3, 2014
CiteVis

Demo

Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster ‘13
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
Visualization’s Value

\[ \text{Value} = T + I \hspace{1cm} + E \hspace{1cm} + C \]
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
Visualization’s Value

\[ V_{value} = T + I + E + C \]
Value
Not evaluation in traditional sense

User study with 12 grad students not so helpful
Takeaway 3

Understand & communicate the value
Key Open Problems in Visualization
Key Open Problems in Visualization

1. Assess and communicate value

\[ V_{\text{value}} = T + I + E + C \]
Key Open Problems in Visualization

1. Assess and communicate value
2. Make the construction of visualizations easier

DataWrapper
https://datawrapper.de/

iVisDesigner
Ren, Hollerer, & Yuan
TVCG (InfoVis) ‘14

SketchStory
Lee, Kazi & Smith
TVCG (InfoVis) ‘13

Lyra
Satyanarayan & Heer
EuroVis ‘14
Key Open Problems in Visualization

1. Assess and communicate value
2. Make the construction of visualizations easier
3. Address real world “big” problems
Take Aways

Understand the purpose

Presentation & analysis, related but different

Interaction provides power, use it

Explain the value \((T + I + E + C)\)
Thanks for your attention
Acknowledgments

- Supported by DARPA’s XDATA program

- Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537
The Value of Visualization for Exploring and Understanding Data

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

Boeing Data Analytics Community of Excellence
Data
Data Visualization

Making pretty pictures?
Data Visualization

Making pretty pictures
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process
Gain an understanding
Challenge
Transform the data into information (understanding, insight) thus making it useful to people.
Visualization

“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visuals help us think
Provides a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Let Me Illustrate

An example...
Questions:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cereal</td>
<td>Manufacturer</td>
<td>Fiber</td>
<td>Potassium</td>
</tr>
<tr>
<td>2</td>
<td>100% Bran</td>
<td>N</td>
<td>10</td>
<td>280</td>
</tr>
<tr>
<td>3</td>
<td>100% Natural Bran</td>
<td>Q</td>
<td>2</td>
<td>135</td>
</tr>
<tr>
<td>4</td>
<td>All-Bran</td>
<td>K</td>
<td>9</td>
<td>320</td>
</tr>
<tr>
<td>5</td>
<td>All-Bran with Extra Fiber</td>
<td>K</td>
<td>14</td>
<td>330</td>
</tr>
<tr>
<td>6</td>
<td>Almond Delight</td>
<td>R</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Apple Cinnamon Cheerios</td>
<td>G</td>
<td>1.5</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>Bran Chex</td>
<td>R</td>
<td>4</td>
<td>125</td>
</tr>
<tr>
<td>9</td>
<td>Bran Flakes</td>
<td>P</td>
<td>5</td>
<td>190</td>
</tr>
<tr>
<td>10</td>
<td>Cap'n'Crumble</td>
<td>Q</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>Cheerios</td>
<td>G</td>
<td>2</td>
<td>105</td>
</tr>
<tr>
<td>12</td>
<td>Cocoa Puffs</td>
<td>G</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>13</td>
<td>Corn Chex</td>
<td>R</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>14</td>
<td>Corn Flakes</td>
<td>K</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>15</td>
<td>Count Chocula</td>
<td>G</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>16</td>
<td>Cracklin' Oat Bran</td>
<td>K</td>
<td>4</td>
<td>160</td>
</tr>
<tr>
<td>17</td>
<td>Cream of Wheat (Quick)</td>
<td>N</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>Crispy Wheat &amp; Raisins</td>
<td>G</td>
<td>2</td>
<td>120</td>
</tr>
<tr>
<td>19</td>
<td>Double Chex</td>
<td>R</td>
<td>1</td>
<td>80</td>
</tr>
<tr>
<td>20</td>
<td>Froot Loops</td>
<td>K</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>21</td>
<td>Frosted Flakes</td>
<td>K</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>22</td>
<td>Fruit &amp; Fibre Dates, Walnuts</td>
<td>P</td>
<td>5</td>
<td>200</td>
</tr>
<tr>
<td>23</td>
<td>Fruitful Bran</td>
<td>K</td>
<td>5</td>
<td>190</td>
</tr>
<tr>
<td>24</td>
<td>Fruity Pebbles</td>
<td>P</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>Golden Grahams</td>
<td>G</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>26</td>
<td>Grape Nuts Flakes</td>
<td>P</td>
<td>3</td>
<td>85</td>
</tr>
<tr>
<td>27</td>
<td>Honey Nut Cheerios</td>
<td>G</td>
<td>1.5</td>
<td>90</td>
</tr>
<tr>
<td>28</td>
<td>Honey-comb</td>
<td>P</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>29</td>
<td>Just Right Fruit &amp; Nut</td>
<td>K</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>30</td>
<td>Life</td>
<td>Q</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>31</td>
<td>Lucky Charms</td>
<td>G</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>32</td>
<td>Maypo</td>
<td>A</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>33</td>
<td>Muesli Raisins, Dates, &amp; Walnuts</td>
<td>R</td>
<td>3</td>
<td>170</td>
</tr>
<tr>
<td>34</td>
<td>Multi-Grain Cheerios</td>
<td>G</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>35</td>
<td>Nutri-Grain Almond-Raisins</td>
<td>K</td>
<td>3</td>
<td>130</td>
</tr>
<tr>
<td>36</td>
<td>Nutri-grain Wheat</td>
<td>K</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>37</td>
<td>Oatmeal Raisin Crisp</td>
<td>G</td>
<td>1.5</td>
<td>120</td>
</tr>
<tr>
<td>38</td>
<td>Post Nat. Raisin Bran</td>
<td>P</td>
<td>6</td>
<td>260</td>
</tr>
<tr>
<td>39</td>
<td>Product 19</td>
<td>K</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>40</td>
<td>Quaker Oatmeal</td>
<td>Q</td>
<td>2.7</td>
<td>110</td>
</tr>
<tr>
<td>41</td>
<td>Raisin Bran</td>
<td>K</td>
<td>5</td>
<td>240</td>
</tr>
<tr>
<td>42</td>
<td>Raisin Nut Bran</td>
<td>G</td>
<td>2.5</td>
<td>140</td>
</tr>
<tr>
<td>43</td>
<td>Rice Krispies</td>
<td>K</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>44</td>
<td>Shredded Wheat</td>
<td>N</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>45</td>
<td>Shredded Wheat 'n'Bran</td>
<td>N</td>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>46</td>
<td>Shredded Wheat spoon</td>
<td>N</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>47</td>
<td>Smacks</td>
<td>K</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>48</td>
<td>Special K</td>
<td>K</td>
<td>1</td>
<td>55</td>
</tr>
<tr>
<td>49</td>
<td>Strawberry Fruit Wheats</td>
<td>N</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>50</td>
<td>Total Corn Flakes</td>
<td>G</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>51</td>
<td>Total Raisin Bran</td>
<td>G</td>
<td>4</td>
<td>230</td>
</tr>
<tr>
<td>52</td>
<td>Total Whole Grain</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>53</td>
<td>Trix</td>
<td>G</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>54</td>
<td>Wheaties</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>55</td>
<td>Wheaties Honey Gold</td>
<td>G</td>
<td>1</td>
<td>60</td>
</tr>
</tbody>
</table>
Even Tougher?

What if you could only see one cereal’s data at a time? (e.g. some websites)
What if I read the data to you?
Part of our Culture

“I see what you’re saying”
“Seeing is believing”
“A picture is worth a thousand words”
Visualization
“Data visualization”

Scientific visualization

Information visualization
Scientific Visualization (SciVis)

Primarily relates to and represents something spatial, physical or geometric

- Often 3-D
- Examples
  - Air flow over a wing
  - Stresses on a girder
  - Torrents inside a tornado
  - Organs in the human body
  - Molecular bonding

Not my research focus
Information Visualization (InfoVis)

Visualizing *non-spatial* data: Items, entities, things which do not have a direct physical correspondence

- Notion of *abstractness* of the entities is important too
- Examples: baseball statistics, stock trends, connections between criminals, car attributes...

“A key challenge in information visualization is designing a cognitively useful spatial mapping of a dataset that is not inherently spatial and accompanying the mapping by interaction techniques that allow people to intuitively explore the dataset. Information visualization draws on the intellectual history of several traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, statistical graphics, cartography, and art.”

http://conferences.computer.org/infovis/
Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces

Applications of Visualization

Presentation  Exploratory
Analysis      Exploratory
1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Infographics

All the rage...
THE NUCLEAR ARMS RACE
It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE
- First Atomic Bombs on Hiroshima 1945
- First American H Bomb 1952
- First Soviet A Bomb 1953
- First Soviet H Bomb 1957
- US and USSR InterContinental Ballistic Missiles program 1960
- USA and USSR InterContinental Ballistic Missiles program 1960
- USA withdraw from Soviet 1979
- USA withdraw from Soviet 1986
- USSR collapse 1991

NUCLEAR STOCKPILE
- USA
- USSR

ROCKET MODELS
- SS-9 (USSR)
- Titan II (USA)

A WORLD BREAK IN TWO
NATO and Warsaw Pact were both mutual defense treaties between states. The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.

Produced by WatchTheAmericans.com a Fan Website
http://visual.ly/nuclear-arms-race
Steroids or Not, the Pursuit Is On

Barry Bonds is aiming at the career home run record. He needs only six more to be Babe Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs.

Hank Aaron
755 home runs
23 seasons

Babe Ruth
714 home runs
22 seasons

Barry Bonds
708 home runs
20 seasons

According to allegations in a book about Bonds, he began taking steroids before the 1998 season, hit 42 home runs in 1998. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the allegations are correct, Bonds was 34 in his first season on steroids. Here are projected home run rates for each player after age 34.

Aaron
Actual home run slightly understated. Projected home run rate for five seasons.

Ruth
Averaged 46.4 home runs a season from age 30 to 34. Averaged 42.9 for next four seasons.

Bonds
Hit age 35 to 37, he averaged 14 more home runs in a season than projected.

Note: Ages as of July 1 of each season.

Differing Paths to the Top of the Charts

The seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (33rd).
The Beeriodic Table

Click on any image below to see full description

http://thebeermongers.com/beers/
Where We Live...

Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie.

80% of the U.S. population lives in a metropolitan area.

The entire state of Wyoming (pop. 505,300) has fewer people than the Manhattan, Pa., metro area.

3. Chicago metro area (pop. 9,443,400)

4. Philadelphia metro area (pop. 6,293,208)

1. New York City metro area (pop. 18,747,360)

2. Los Angeles metro area (pop. 12,923,500)

5. Dallas-Fort Worth metro area (pop. 5,810,500)

Loving County, Texas, is the least populated county in the lower 48 states, with 62 residents.

Alaska is the most sparsely populated state, with 1 person per square mile.

THE ALMIGHTY DOLLAR
MAPPING DISTRIBUTION OF INCOME BY RELIGIOUS BELIEF

It's no secret that the distribution of wealth is inequitable in the United States across racial, regional, and socio-economic groups. But there is a distinct variance among and within America's faiths as well. Displayed below are the income levels of America's major religious groups, as compared to the average U.S. income distribution.

INCOME BRACKETS:

- $100,000 +
- $75,000 to $99,999
- $50,000 to $74,999
- $30,000 to $49,999
- LESS THAN $30,000

A COLLABORATION BETWEEN GOOD AND COLUMN FIVE.

SOURCE: THE PEN FORUM

http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg
Frequent presentation goals
Clarify
Focus
Highlight
Simplify
Persuade

May just show a few variables and/or a subset of the data cases
Gore made extensive use of data graphics
2. Analysis
Explore the data
Assess a situation
Determine how to proceed
Decide what to do
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland

The Elements of Graphing Data
Frequent analysis goals
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility
Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
So what is visualization most useful for?
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context
Investigative analysis

Intelligence analysts must explore and evaluate voluminous data, from narrative recordings of fluid agents to open-source news articles. Insights from visual analytics projects and the graphical formats show the potential of visual analytics to aid these investigations.

Investigative analysis has been a key application domain for visual analytics due to the exponential growth of intelligence-related data. Görg, Kang, Kiu, and Stasko's work describes the challenges and gaps of the new visual analytics tools, as they enable analysts to focus on actionable intelligence and support the analysts' need to navigate and understand complex data. Their work identifies opportunities for visual analytics support in the intelligence process, the risk of overlooking important data, and the ability of visual analytics to assist in hypothesis-driven analysis.

Görg, Kang, Kiu, Stasko

*IEEE Computer* ‘13
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction
Interaction is Vital

Engage in a dialog with your data

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

Yi, Kang, Stasko, Jacko
TVCG (InfoVis) ‘07
3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.
Visualization most useful in **exploratory data analysis**

Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Some examples from my group’s research
SellTrend

Web-based airline booking requests

Help recognize irregularities and analyze what the problem is

http://www.cc.gatech.edu/gvu/ii/selltrend/

Liu, Sullivan, Stasko
TVCG (InfoVis) ‘09
OnSet

http://www.cc.gatech.edu/gvu/ii/setvis/

Demo

Sadana, Major, Dove, & Stasko
NBA Player and Team Data

http://www.cc.gatech.edu/gvu/ii/sportvis/nbaVis/

Project by Fengbo Li

Demo
PGA Tour Season Recap

<table>
<thead>
<tr>
<th>Event</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Round 4</th>
<th>Position</th>
<th>Total</th>
<th>Strokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humana Challenge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T37</td>
<td>-17</td>
<td>271</td>
</tr>
<tr>
<td>Farmers Insurance Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T51</td>
<td>E</td>
<td>288</td>
</tr>
<tr>
<td>Phoenix Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-28</td>
<td>256</td>
</tr>
<tr>
<td>Pebble Beach Pro-Am</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T60</td>
<td>-1</td>
<td>285</td>
</tr>
<tr>
<td>Northern Trust Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T21</td>
<td>-3</td>
<td>281</td>
</tr>
<tr>
<td>WGC-Cadillac Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T3</td>
<td>-14</td>
<td>274</td>
</tr>
<tr>
<td>Arnold Palmer Invitational</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CUT</td>
<td>+8</td>
<td>152</td>
</tr>
<tr>
<td>Houston Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T16</td>
<td>-10</td>
<td>278</td>
</tr>
<tr>
<td>The Masters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T54</td>
<td>+9</td>
<td>257</td>
</tr>
<tr>
<td>Wells Fargo Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>-7</td>
<td>281</td>
</tr>
<tr>
<td>PLAYERS Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CUT</td>
<td>+1</td>
<td>145</td>
</tr>
<tr>
<td>St. Jude Classic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T2</td>
<td>-10</td>
<td>270</td>
</tr>
<tr>
<td>U.S. Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T2</td>
<td>+3</td>
<td>283</td>
</tr>
<tr>
<td>The Greenbrier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CUT</td>
<td>+2</td>
<td>142</td>
</tr>
<tr>
<td>Open Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-3</td>
<td>281</td>
</tr>
<tr>
<td>Bridgestone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T21</td>
<td>+1</td>
<td>281</td>
</tr>
<tr>
<td>PGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T72</td>
<td>+12</td>
<td>292</td>
</tr>
<tr>
<td>Barclays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T8</td>
<td>-9</td>
<td>275</td>
</tr>
<tr>
<td>Deutsche Bank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T41</td>
<td>-8</td>
<td>276</td>
</tr>
<tr>
<td>BMW Champ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T33</td>
<td>-1</td>
<td>283</td>
</tr>
</tbody>
</table>

with Anand Sainath, Rahul Basole
"Putting the pieces together"

**Jigsaw**

Computational analysis of documents’ text
- Entity identification, document similarity, clustering, summarization, sentiment
- Multiple visualizations of documents, analysis results, entities, and their connections
- Views are highly coordinated
Help “investigators” explore, analyze and understand large document collections

Academic Papers

Product Reviews

Health Forums

Police Reports

2010 Hyundai Genesis Sedan - Consumer Reviews

Great value and excellent performance

By Javen on 12/9/10 10:55 AM (EST)

Vehicle

2010 Hyundai Genesis 3.8 L V6 Sedan (3.8L V6 6A) 4.6 L V8 6A

Review

After driving a few cars I decided on a change due to price. Toyota/Lexus were having all the recall problems at the time, so I got a Genesis for a test drive. What I saw is that the car handles beautifully in all driving conditions, the sound proofing is good. I would highly recommend this car to any sedan like ride at a very affordable price.

Favorite Features

Solid driving experience, nicely designed.

Suggested Improvements

Drivens seat could be a bit more comfortable, similar to the seat should be fully powered so its can be raised or lowered.

Recommend: 4/5
Jigsaw

Visual analytics of large text document collections

Görg, Liu, Kihm, Choo, Park, & Stasko
TVCG ‘13
Example

Reviews of wines from Washington state from ‘07-’10

Text: review narrative

Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, ...

Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured

1323 reviews

from database of 150,000 reviews
Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw
Applications

• Intelligence & law enforcement
  – Police cases
  – Won 2007 VAST Contest
  – Stasko et al, Information Visualization ‘08

• Academic papers, PubMed
  – All InfoVis & VAST papers
  – CHI papers
  – Görg et al, KES ‘10

• Investigative reporting

• Fraud
  – Finance, accounting, banking

• Grants
  – NSF CISE awards from 2000

• Topics on the web (medical condition)
  – Autism

• Consumer reviews
  – Amazon product reviews, edmunds.com, wine reviews
  – Görg et al, HCIR ‘10

• Business Intelligence
  – Patents, press releases, corporate agreements, ...

• Emails
  – White House logs

• Software
  – Source code repositories
  – Ruan et al, SoftVis ‘10
Moving to Tablets

Scatterplot

Video

Sadana & Stasko
AVI ‘14

http://www.cc.gatech.edu/gvu/ii/touch/
Different benefits of visualization
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
CS 7450 - Information Visualization

Instructor: John Stasko
Fall 2013
Mon, Wed 3:00 - 4:30 pm
Whitaker Bldg, room 1103

Information visualization is a research area that focuses on the use of visualization techniques to help people understand and analyze data. While fields such as scientific visualization involve the presentation of data that has some physical or geometric correspondence, information visualization focuses on abstract data without such correspondences such as symbolic, tabular, networked, hierarchical, or textual information sources.

The objectives of the course are

- Learn the principles involved in information visualization
- Learn about the variety of existing techniques and systems in information visualization
- Develop skills in critiquing different visualization techniques as applied to particular tasks
- Learn how to evaluate visualization systems
- Gain a background that will aid the design of new, innovative visualizations

The course will follow a lecture/seminar style with much discussion of assigned readings, as well as viewing of videos and hands-on experience with research and commercial information visualization tools.

We will be reading recent research papers about the different course topics. In addition, we will be using one book for the course: *Now You See It* by Stephen Few, Analytics Press 2009. Also highly recommended is *Envisioning Information* by Edward Tufte, Graphics Press 1990.
Syllabus

Link to last year's 2012 syllabus with lecture slides

Overview

<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Topic</th>
<th>Topic</th>
<th>HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aug 19, 21</td>
<td>Introduction</td>
<td>InfoVis overview</td>
<td>HW 1</td>
</tr>
<tr>
<td>2</td>
<td>Aug 26, 28</td>
<td>Visual perception</td>
<td>Value/Benefits of visualization</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Sep 2, 4</td>
<td>No Class -- Labor Day</td>
<td>Multivariate data &amp; table/graph design</td>
<td>HW 2</td>
</tr>
<tr>
<td>4</td>
<td>Sep 9, 11</td>
<td>Few's design guidance</td>
<td>Multivariate visual representations 1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Sep 16, 18</td>
<td>Multivariate visual representations 2</td>
<td>Tasks and analysis</td>
<td>HW 3a HW 3b</td>
</tr>
<tr>
<td>6</td>
<td>Sep 23, 25</td>
<td>InfoVis systems &amp; toolkits</td>
<td>Commercial systems demos</td>
<td>HW 4</td>
</tr>
<tr>
<td>7</td>
<td>Sep 30, Oct 2</td>
<td>Storytelling</td>
<td>Tufte’s design principles</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Oct 7, 9</td>
<td>Poster session</td>
<td>Casual InfoVis</td>
<td>HW 5</td>
</tr>
<tr>
<td>9</td>
<td>Oct 14, 16</td>
<td>No Class - Fall break</td>
<td>No Class - VIS Conference</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Oct 21, 23</td>
<td>Graphs and networks 1</td>
<td>Graphs and networks 2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Oct 28, 30</td>
<td>Hierarchies &amp; trees 1</td>
<td>Hierarchies &amp; trees 2</td>
<td>HW 6</td>
</tr>
<tr>
<td>12</td>
<td>Nov 4, 6</td>
<td>Interaction</td>
<td>Overview &amp; detail</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Nov 11, 13</td>
<td>Text &amp; documents 1</td>
<td>Text &amp; documents 2</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Nov 18, 20</td>
<td>Visual analytics 1</td>
<td>Visual analytics 2</td>
<td>HW 7</td>
</tr>
<tr>
<td>15</td>
<td>Nov 25, 27</td>
<td>Time series data</td>
<td>Project work day</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Dec 2, 4</td>
<td>Evaluation</td>
<td>Review</td>
<td></td>
</tr>
</tbody>
</table>

Detail

Below are summaries of all the materials for each class. Lecture slides, references for readings and articles, videos shown, and software demonstrated are listed for each topic. Also available is a more general, alphabetical bibliography.
For more, my EuroVis ‘14 Capstone Talk

http://vimeo.com/98986594

The Value of Visualization... and Why Interaction Matters

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

EuroVis 2014: Capstone: The Value of Visualization... and Why Interaction Matters

Capstone Speaker: John Stasko, Georgia Institute of Technology

Abstract: Visualization researchers need to do a better job communicating the value of our field externally. Visualization, by its very nature, provides inherent challenges to doing this. In this talk I will explain these challenges and articulate my views on the value of visualization, including its unique capabilities for data presentation and analysis. I will describe the advantages of interaction, and discuss in detail why interaction is so important to our field and how it has been under-utilized to date. Finally, I will present a number of new interaction ideas and techniques that can be integrated into our future systems.

Read More...
**Take Aways**
Presentation & analysis

Exploratory data analysis, power from interaction

Developing new questions and insights
Acknowledgments

• Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA program
My Research

http://www.cc.gatech.edu/gvu/ii
The Value of Visualization for Understanding Data and Making Decisions

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

JISIC 2014
Data
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
Data Visualization

Making pretty pictures?
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process
Gain an understanding
Visualization
“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visually help us think
Provide a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Applications of Visualization
Presentation
Analysis
1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Infographics
THE NUCLEAR ARMS RACE
It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE
- First Atomic Bomb on Hiroshima: 1945
- First American H Bomb: 1949
- First U.S. missile submarine launched: 1952
- SALT I talks on reducing nuclear weapons: 1960
- SALT II talks on reducing nuclear weapons: 1979
- Reykjavik summit on reducing nuclear weapons: 1986

USA and USSR intercontinental ballistic missiles program: 91
USA withdraws from SALT II: 86
USSR collapses: 91

NUCLEAR STOCKPILE

ROCKET MODELS
SS-9 (USSR) vs. Titan II (USA)
- Year: 1966 vs. 1962
- Warhead: 10 Mt. vs. 9 Mt.
- Operational range: 16,000 km vs. 15,000 km

AIR CONTROL
- Intercontinental Ballistic Missiles
  - USA: 8,000 vs. USSR: 7,000
- Planes
  - USA: 4,000 vs. USSR: 5,000

A WORLD BREAK IN TWO
NATO and Warsaw Pact were both mutual defense treaties between states. The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.

Troops
- NATO: 2.6 million vs. Warsaw Pact: 4 million

Tanks
- NATO: 13,000 vs. Warsaw Pact: 42,500

Artillery
- NATO: 10,750 vs. Warsaw Pact: 31,500

http://visual.ly/nuclear-arms-race
Gay rights in the US, state by state

Gay rights laws in America have evolved to allow — but in some cases ban — rights for gay, lesbian and transgender people on a range of issues, including marriage, hospital visitation, adoption, housing, employment and school bullying. The handling of gay rights issues vary by state and follow trends by region.

- Obama supports same-sex marriage: share with us what it means to you
- Gay rights: five activists reflect on the history of the movement in the US
- President Obama endorses gay marriage
Simply presenting data *visually* can have a profound impact
Gun ownership in New York counties

The gun owner next door: What you don't know about the weapons in your neighborhood

http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1

http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/Map-Where-gun-permits-your-neighborhood-gcheck=1&nclick_check=1
Frequent presentation goals

Clarify
Focus
Highlight
Simplify

May just show a few variables and/or a subset of the data cases
2. Analysis
Explore the data
Assess a situation
Determine how to proceed
Decide what to do
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland  
*The Elements of Graphing Data*
Frequent analysis goals
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility
Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
So what is visualization most useful for?
1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.
Investigative analysis has been a key application domain for visualization since the establishment of the National Visualization and Analytics Information Technology (NVAIT) program in 2005. The program's goal is to develop visual analytics tools and techniques that can support the cognitive processes of analysts who are working to understand complex, multidimensional data sets. This research is important because it can help analysts identify patterns and trends that are not immediately apparent in the raw data. By providing analysts with a more complete understanding of the data, these tools can help them make better decisions and improve the overall effectiveness of their work.
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction
Engage in a dialog with your data

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect
3. **Visualization** rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly
Visualization most useful in exploratory data analysis
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...
Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?
Example Tasks and Goals

• Which documents contain text on topic XYZ?
• Which documents are of interest to me?
• Are there other documents that are similar to this one (so they are worthwhile)?
• How are different words used in a document or a document collection?
• What are the main themes and ideas in a document or a collection?
• Which documents have an angry tone?
• How are certain words or themes distributed through a document?
• Identify “hidden” messages or stories in this document collection.
• Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
• Find connections between documents.
The Challenge

Text is nominal data

Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data

- Bar charts, line charts, scatterplots, etc.
Jigsaw
Computational analysis of documents’ text
   Entity identification, document similarity, clustering, summarization, sentiment
Multiple visualizations of documents, analysis results, entities, and their connections
   Views are highly coordinated

“Putting the pieces together”
Help “investigators” explore, analyze and understand large document collections

**Academic Papers**  
**Product Reviews**  
**Health Forums**  
**Police Reports**

---

2010 Hyundai Genesis Sedan - Consumer Reviews

**Vehicle**
2010 Hyundai Genesis 3.8 4dr Sedan (3.8L 6cyl 6A) 4.6 kW (4.6L 8cyl 6A)

**Review**
After owning a 4 LS series Levu's I decided on a change due to price. Toyota/Lexus mix having all the recall problems at the automotive reviews. I took a Genesis for a test drive. What I say that the car handles beautifully in all driving conditions, it sounds proof cable. I would highly recommend this car to any sedan like ride at a very affordable price.

**Favorite Features**
Solid driving experience, nicely designed.

**Suggested Improvements**
Driving seat could be a bit more comfortable, similar to the seat should be fully powered so it can be raised or lowered.

**Recommend**

---

Fort Worth Police Department
WANTED PERSON REPORT

---

Jewelry worth over $25,000 stolen

---

Additional Information
Large diamond ring on left hand
Money carried-Woman usually carries large amounts

---

Subject, Karen Lynn Bennett, came to her grandmother's house on the evening of 1-10-64, and left all of her clothes and property. The subject stayed with her grandmother for a few nights and then left with a white male, James Berlin grandmother for a few nights and then left with James Berlin grandson, who she would be born the destination unknown. Subject, Kari Lynn Bennett, is 5'2", weight 115 lbs, blond, blue eyes.

---

38
Jigsaw

Visual analytics of large text document collections

Görg, Liu, Kihm, Choo, Park, & Stasko
TVCG ‘13
Example

Made each page into a separate “document”
585 in total
Entities: Person, Location, Organization, Date, Money
Graph View
## Calendar View

<table>
<thead>
<tr>
<th></th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw
Applications

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* ‘08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES ‘10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR ‘10
- Business Intelligence
  - Patents, press releases, corporate agreements, ...
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis ‘10
Further benefits of visualization
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
EuroVis ‘14 Capstone Talk

http://vimeo.com/98986594
Take Aways
Presentation & analysis

Interaction provides the power

Exploring & developing questions
Acknowledgments

• Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA Program
My Research

http://www.cc.gatech.edu/gvu/ii
Value-Driven Evaluation of Visualizations

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
“No user studies are reported in this paper, thus it is unclear if the visualization will be useful.”
Really?
· Which cars have the best miles-per-gallon?
· How much does a Ford Taurus cost?
· Is there a correlation between car weight and torque?
· How many different countries manufacture cars?
· What is the range of car’s horsepower?
Explanation(s) and Illustration(s) of value
Visualization’s Value?
Visualization’s Value?

$V_{\text{alue}} = T + I + E + C$
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to minimize the total time needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)
Visualization’s Value?

\[ V_{\text{value}} = T + I + E + C \]

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to convey an overall \textit{essence} or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)
Visualization’s Value?

\[ \text{Value} = \text{T} + \text{I} + \text{E} + \text{C} \]

Ability to generate confidence and trust about the data, its domain and context

(Beneficial data analysis process side effects)
Map of the Market

CiteVis

http://www.marketwatch.com/tools/stockresearch/marketmap

http://www.cc.gatech.edu/gvu/ii/citevis
Data Visualization 101

Problem:
You have a lot of data (& attributes) to understand

Do you?

Pack all the data into one complex representation

Spread the data into multiple coordinated views

Use interaction to reveal different subsets of the data
Interaction is Vital

Engage in a dialog with your data

“The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means.”

S. Few, Now you see it
Interaction

Understand the benefits (science of interaction)

Explore new interactive operations

Visualization on other display devices (touch)
For more, see my EuroVis ‘14 Capstone Talk

http://vimeo.com/98986594
Acknowledgments

• Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA program
Visual Analytics for Investigative Analysis and Exploration of Documents and Data

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
Data Visualization

Making pretty pictures?
Data Visualization

Making pretty pictures
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process

Gain an **understanding**
Visualization
“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visuals help us think
Provide a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Applications of Visualization
Presentation
Analysis
1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Infographics
THE NUCLEAR ARMS RACE
It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE
- First Atomic Bomb on Hiroshima
- First Soviet A Bomb
- First U.S. missile submarine launched
- SALT I talks on reducing nuclear weapons
- SALT II talks on reducing nuclear weapons
- Reykjavik summit on reducing nuclear weapons
- USA and USSR InterContinental Ballistic Missiles program
- USA withdraws from SALT II
- USSR collapse

NUCLEAR STOCKPILE

ROCKET MODELS
- SS-9 (USSR)
  - Year: 1966
  - Warhead: 10 Mt
  - Operational range: 16,000 km
- Titan II (USA)
  - Year: 1962
  - Warhead: 10 Mt
  - Operational range: 15,000 km

AIR CONTROL
- Intercontinental Ballistic Missiles
  - USA: 8,000
  - USSR: 7,000
- Planes
  - USA: 4,000
  - USSR: 5,000

A WORLD BREAK IN TWO
NATO and Warsaw Pact were both mutual defense treaties between states. The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.

Troops
- NATO: 2.6 million
- Warsaw Pact: 4 million

 Tanks
- NATO: 13,000
- Warsaw Pact: 42,500

Artillery
- NATO: 10,750
- Warsaw Pact: 31,500

Produced by WatchTheAmericans.com a Fan Website
755 Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs.

Hank Aaron
755 home runs
23 seasons

Babe Ruth
714 home runs
23 seasons

Barry Bonds
708 home runs
20 seasons

Bonds took lead
Home runs after 16 seasons
Bonds: 567
Aaron: 544
Ruth: 516

According to allegations in a book about Bonds, he began taking steroids before the 1990 season. He hit 44 in 1990, the year before the book was published. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the allegations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

Aaron
Actual home runs: 544
Projected home runs: 40

Ruth
Averaged 40.4 home runs a season from age 30 to 34.
Averaged 42.5 for next four seasons.

Bonds
From age 35 to 39, he averaged 44 home runs a season. He was projected to average 41.

Note: Ages as of July 1 of each season.

Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (128), Rodriguez (97th) and Pujols (267th).

NY Times 2006
Where We Live...

Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie.

80% of the U.S. population lives in a metropolitan area.

The entire state of Wyoming (pop. 505,300) has fewer people than the Harrisburg, Pa., metro area.

Our families are getting smaller—with one vital exception. Compared with those of Europe and Japan, the U.S. population is younger and more colorful because of the continued arrival of immigrants and their higher-than-average birthrates. Of the 100 million Americans who will join us in the next 37 years, half will be immigrants or their children. In the next few decades, 97% of the world's population growth will occur in the developing world; the U.S. is the largest developed country in the world that is still growing at a healthy clip. That matters, strategically, economically.

1. New York City metro area (pop. 18,741,300)
2. Los Angeles metro area (pop. 12,923,500)
3. Chicago metro area (pop. 9,443,400)
4. Philadelphia metro area (pop. 6,829,300)
5. Dallas-Fort Worth metro area (pop. 5,810,600)

New Jersey is the most densely populated state, with 1,134 people per square mile.

Sources: U.S. Census Bureau; LexisNexis 2009/11 Information LLC.

THE ALMIGHTY DOLLAR
MAPPING DISTRIBUTION OF INCOME BY RELIGIOUS BELIEF

It's no secret that the distribution of wealth is inequitable in the United States across racial, regional, and socio-economic groups. But there is a distinct variance among and within America's faiths as well. Displayed below are the income levels of America's major religious groups, as compared to the average U.S. income distribution.

INCOME BRACKETS:

- **$100,000 +**
- **$75,000 to $99,999**
- **$50,000 to $74,999**
- **$30,000 to $49,999**
- **LESS THAN $30,000**

A COLLABORATION BETWEEN GOOD AND COLUMN FIVE.

http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg
Simply presenting data *visually* can have a profound impact
CS 7450
Fall ‘12
Design project

Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri
Gun ownership in New York counties

http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1
http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/Map-Where-gun-permits-your-neighborhood?-gcheck=1&nclick_check=1
Frequent presentation goals
Clarify
Focus
Highlight
Simplify

May just show a few variables and/or a subset of the data cases
2. Analysis
understand, compare, decide, judge, evaluate, assess, determine, ...
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland
The Elements of Graphing Data
Frequent analysis goals
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility
Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
So what is visualization most useful for?
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

**Learning, awareness, trust, context**
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning

Use them when they get the job done!
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

**Q & A dialog through interaction**
Visualization

“The use of computer-supported, interactive visual representations of data to amplify cognition”
3. **Visualization** rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...


Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

35
Text is Everywhere

We use documents as primary information artifacts in our lives.

Our access to documents has grown tremendously in recent years due to the networking infrastructure:  
- WWW  
- Digital libraries  
- ...
Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?
Example Tasks and Goals

• Which documents contain text on topic XYZ?
• Which documents are of interest to me?
• Are there other documents that are similar to this one (so they are worthwhile)?
• How are different words used in a document or a document collection?
• What are the main themes and ideas in a document or a collection?
• Which documents have an angry tone?
• How are certain words or themes distributed through a document?
• Identify “hidden” messages or stories in this document collection.
• Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
• Find connections between documents.
The Challenge

Text is nominal data

Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data

- Bar charts, line charts, scatterplots, etc.
“Putting the pieces together”

**Jigsaw**

Computational analysis of documents’ text
- Entity identification, document similarity, clustering, summarization, sentiment
- Multiple visualizations of documents, analysis results, entities, and their connections
- Views are highly coordinated
Example
Reviews of wines from Washington state from ‘07–’10
Text: review narrative
Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, ...
Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured
1323 reviews
from database of 150,000 reviews
Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw
Applications

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* ‘08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES ‘10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR ’10
- Business Intelligence
  - Patents, press releases, corporate agreements, ...
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis ‘10
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context

Data cleaning example
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

**Q & A dialog through interaction**

Repeated interactive queries
3. **Visualization** rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Varieties, colors, reviewers, producers, concepts, ...
Insights
No great aha! – More learning & knowledge building
Domain knowledge (wines) matters
Further benefits of visualization
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
Take Aways
Presentation & analysis

Interaction provides the power

Exploring & developing questions
Acknowledgments

• Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA Program
My Research

http://www.cc.gatech.edu/gvu/ii

The Information Interfaces Group, an HCI research group in the School of Interactive Computing at Georgia Tech, develops computing technologies that help people take advantage of information to enrich their lives.

More about the lab approach

Projects
- Jigsaw
  - Using visualization and visual analytics to help analysis and sensemaking on text document collections.
- CiteVis
  - Exploring conference paper citation data visually.

Hot News
In June, John gave the Capstone lecture (replay, slides) at the EuroVis 2014 Conference.
We presented a paper about our work on touch interfaces for Infovis on tablet computers at AVI 2014 in Lago Como, Italy in May.
This summer, lab students are interning at Microsoft Research, Google, and Twitter, among other places.
We had papers about Ploucues and our study of visual analytics for intelligence analysis appear in the first two 2014 issues of the Information Visualization Journal.
John was named an IEEE Fellow in Jan. 2014.
In November, John was appointed as an Honorary Professor in the School of Computer Science at the Univ. of St. Andrews in Scotland.
We presented a paper about visualizing social and communicative behaviors at VAHC '13 in November.
A journal paper about how jigsaw combines computational text analysis with interactive visualization appeared in the Oct. '13 IEEE TVCG issue, and was presented at one of the VIS '13 TVCG sessions.

Lab - Visit our lab
- GCS 7450 - Infovis Class
- VACCINE Center
- FODAVA Center

Research supported in part by grants from the National Science Foundation (NSF-1320037), DARPA's XDATA program, from the Dept. of Homeland Security's Center of Excellence in Command, Control, and Interoperability (VACCINE), an IBM PhD Fellowship, and donations from the Georgia Automotive and Shapes Research.
iLeaps

Android Native App Development for iLaw Enforcement Application Program for Students

Kofi Nyarko | Benjamin Hall | Alfred Shoetan | Leonardo Ferreira | Roberta Virgil

June 30, 2015
Incidents on college campuses have made safety a central issue.

Need a platform that can improve response time for first-responders while simultaneously providing situational awareness.

Platform enables individuals to report incidents that can be geo-referenced and cataloged for forensic analysis as well as predictive analytics.
• Decrease the response time of emergency responders to an incident
## Dispatch

### iLEAPS (Emergency Response System)

#### Recent Events
<table>
<thead>
<tr>
<th>Emergency Type</th>
<th>Incident Type</th>
<th>Date and Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency</td>
<td>Emergency</td>
<td>2014-12-04 17:06:27</td>
</tr>
<tr>
<td>Emergency</td>
<td>Emergency</td>
<td>2014-12-04 16:35:36</td>
</tr>
</tbody>
</table>

#### All Events
<table>
<thead>
<tr>
<th>Emergency Type</th>
<th>Incident Type</th>
<th>Date and Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency</td>
<td>Emergency</td>
<td>2014-12-04 17:06:27</td>
</tr>
<tr>
<td>Emergency</td>
<td>Emergency</td>
<td>2014-12-04 16:35:36</td>
</tr>
</tbody>
</table>
• Officer and User apps
  • Android push notifications with Google Cloud Messaging
  • Geo-tagging
  • Taking pictures of incidents
• **Android push notifications with Google Cloud Messaging**
  • Send data from your server to your users' devices, and receive messages from devices on the same connection
  • The GCM service handles all aspects of queueing of messages and delivery to client applications running on target devices
  • Distribute messages to single devices, to groups of devices, or to devices subscribed to topics.
  • Battery-efficient connection and free
• Android push notifications with Google Cloud Messaging. Example:

```java
@Override
public void onMessageReceived(String from, Bundle data) {
    String message = data.getString("message");
    Log.d(TAG, "From: " + from);
    Log.d(TAG, "Message: " + message);
    // Handle received message here.
}
```

https://gcm-http.googleapis.com/gcm/send
Content-Type: application/json
Authorization: key=AIzaSyZ-1u...0GBYzPu7Udno5aA
{
    "to": "/topics/foo-bar",
    "data": {
        "message": "This is a GCM Topic Message!",
    }
}
• **Geo-tagging**
  • Google Location Services API, part of Google Play Services
  • Provides a more powerful, high-level framework that automatically handles location providers, user movement, and location accuracy.
  • It also handles location update scheduling based on power consumption parameters you provide.
  • In most cases, you'll get better battery performance, as well as more appropriate accuracy, by using the Location Services API.
• Taking pictures of incidents

- Handles the event
- Calls a Android Method
- Invokes the Activity for taking photos
- Store photo on server
- Upload photo
- Send a GCM message
- GCM
- PHP
- Send a GCM message
- Upload photo
- Store photo on server
- Invokes the Activity for taking photos
- Calls a Android Method
- Handles the event

Overview
• Core functionality on all three interfaces almost complete

• The team is also testing and fixing issues

• Future work:
  • to perform predictive analytics on forensic data (discovery of trends and patterns in collected data)
  • expand to other organizations

Questions ?
Construction and First Analysis of a Corpus for the Evaluation and Training of Microblog/Twitter Geoparsers

JAN OLIVER WALLGRÜN, FRANK HARDISTY, ALAN M. MACEACHREN, MORTEZA KARIMZADEH (PRESENTER), YITING JU, SCOTT PEZANOWSKI
Manually annotated Corpus

Goal: Building a gold standard corpus of microblog posts with manually annotated place references.
◦ Evaluation, comparison, training of geoparsers.
◦ Cognitive studies. What is a place name? Analysis of disagreement between lay persons’ annotations.

Microblog posts (Tweets)
◦ Suitable for monitoring social activities, elections and society’s sentiment, situational awareness in crisis times, etc.
◦ ~1-2 geotagged.
◦ ~10 contain references to places.

Corpus Building Framework
◦ Crowdsourcing
◦ GeoVisual Analytics

Corpus Status:
◦ 6000 tweets
◦ %30 with at least one place name, annotated with GeoNames ID and Geographic Coordinates.
General Corpus Building Framework

“Where are the best #steakhouses in New York City? Find out: http://t.co/KJKLJL #nyc”

Each occurrence would be annotated with the information:

(1) Toponym: New York City;
(2) Geonames ID: 5128581;
(3) Latitude: 40.71427;
(4) Longitude: -74.00597.
Tweet Collection and Sampling

Twitter Streaming API
- ~1 billion Tweets collected, from Jan 2014 to Sept 2014.
- Keywords based on current (crisis related) events.
- Random Sample of 6000 tweets (500 for each keyword).
  - dengue, malaria, earthquake, measles, fire, protest, flood, rebels, flu, riot, gun, tornado.

More place-focused tweets that talk about
- Natural Disasters.
- Infections disease.
- Human threats/violence.

- Filter for “the”.
  - The most common word in the English language.
Crowdsourcing Place Identification

- Amazon Mechanical Turk (AMT)
  - Distributing small tasks: Human Intelligence Tasks (HITs).
  - Recruitment and payment handled by AMT.

- Web Interface for AMT workers to identify place names
  - Detailed instructions for workers.
  - Only allowing for workers with a HIT acceptance rate of 97% or higher from previous HITs.
  - 2 tweets per HIT
  - Each Tweet tagged by at least 5 different AMT workers.
  - 4 cents per assignment.
  - Overall costs of $600 to collect 5 AMT results for each of the 6000 tweets.
Place Identification Instructions

- Annotate as a place reference:
  1. Any named town, city, county, state, or country (e.g., Los Angeles, Jefferson County, NY, Italy)
  2. Named buildings (e.g., Eiffel Tower, Dodgers Stadium, Alcatraz, James J. Ferris High School);
  3. Named areas (e.g., Grand Canyon, Pacific Ocean, Washington Mall, Hyde Park);
  4. Street and highway names (Atherton Street, 1st Ave, Highway 1, I-70);
  5. Place names inside hashtags (#newyork).

- Do NOT annotate as a place reference:
  1. Businesses (Starbucks, Microsoft, Texas Steak House, Baltimore Ravens)
  2. Organizations (Lutheran Church, Red Cross, United Nations, Grand Canyon Historical Society);
  3. Place names used as descriptors (U.S. dollar, Philadelphia cheesesteak).
Agreement Criteria

1. ≥ 5 results from different workers exist for this tweet.

2. For each word in the tweet, there is at least a 70% majority agreement on whether that word is part of a place name and if so together with which other words.

Alternate Generation

Generate the alternates using relevance ranking and geocoding component from our own geoparser, GeoTxt*.

- Name similarity
- Population
- Geographic prominence

Manual Geocoding

- “Expert” Geocoding: Graduate Students and faculty members in Geography.
  1. Only tweets with place names (%30).
  2. Either having local knowledge or general knowledge and willingness to research.

![Image of geocoding interface with maps and input fields]
Corpus Analysis

- Of the event-based sample of tweets making up the corpus
  - 70.0% contain no place names;
  - 22.3% contain one place name;
  - 5.7% contain two place names;
  - 1.3% more than two place names.

- Reached agreement for 91.6% of the 6000 tweets in the corpus.

- Disagreement analysis useful for
  - Place conceptualization.
  - Improving corpus building and NER.
Disagreement Analysis

Resolution of disagreements:
1. Manual resolution by experts that decide what are correct submissions.
2. Automatic resolution via algorithms that can correct or discard incorrect submissions.
3. Introducing new rules to instruct workers more clearly and collect more results via AMT.

Disagreement Categories:
1. 1% - Tweeter errors (TWE) → Manual resolution, new instructions, auto-correction.
   - “It’s been 20 years since the north ridge earthquake? Wow time flies”

2. Worker errors
   - workers make mistakes that clearly were against the instructions given to them.

3. Place Definition
   - What constitutes a place name? Reasonable people may disagree.
   - The most common of the sources of disagreement.
Worker Errors

- 32% - Familiarity (FAM) → Manual resolution.
  - “... help us find missing white large dog Leechan http://t.co/pxZYwmpov Sendai #Japan. one of ...”

- 5% - Misidentified Tokens (MTN) → Manual resolution.
  - “@lakeline @HockeenightsCT great. Now you’ve done it CT. Youve ...”

- 20% - Non-merging Errors (NME) → Automatic resolution.
  - “… Earthquake Shakes Mexicos Pacific Coast: A strong earthquake has shaken the southern Pacific...”

- 6% - Skipped repeats (REP) → Automatic resolution.
  - “… Text: No measles outbreak in Cavite TRECE MARTIRES CITY, Cavite There is still no ...”
Place Definition

- 23% - Adjectival Usage (ADJ) → Improve instructions, manual resolution.
  - “S’ Sudan rebels reinforce captured city defences: South Sudan’s rebels are strengthening ...”

- 16% - Organization with a place in name (ORG) → Improve instructions, manual resolution.
  - “#Patriots How the Indianapolis Colts Defense Can Keep Tom Brady in Check [link]

- 11% - Place Hierarchy Related Issue (PHR) → Improve Instructions, automatic resolution.
  - RT @NewEarthquake: 5.6 earthquake, 297km ENE of Grytviken, South Georgia and the ...

- 5% - Non-proper name (NPN) → Improve instructions.
  - “... Flash flood watch has now been trimmed back to just include the south shore through 9am.”

- 10% - Vaguely qualified place name (VQP) → Improve instructions, potentially special annotation.
  - “... wildfires in south #California: At least two structures burned to the ground an...”
  - “... Theres an outbreak of measles in the Bronx and upper Manhattan, officials warn: ...”

- 6% - Inclusion or exclusion of kind of place (KND) → Improve instructions, mandatory for physical features, flexibility in comparison in the benchmarking framework, automatic resolution.
  - “... How Adelia, 9, and the island of Sabang, #Indonesia, triumphed over #malaria ...”

- 16% - Hashtag Related Issues (HAS) → Improve the annotation platform, improve instructions, automatic resolution.
  - “... everyone please pray for the people there. #prayforindonesia”
Outlook

- Find appropriate licensing/terms for corpus release.
- Implementing our three strategies to deal with disagreements in annotations.
- Use the same framework for other kinds of text.
- Evaluate and improve our own Geoparsing service.

- Summary of disagreement:

<table>
<thead>
<tr>
<th></th>
<th>TWE</th>
<th>FAM</th>
<th>MTN</th>
<th>NME</th>
<th>REP</th>
<th>ADJ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>32%</td>
<td>5%</td>
<td>20%</td>
<td>6%</td>
<td>23%</td>
</tr>
<tr>
<td>ORG</td>
<td>16%</td>
<td>11%</td>
<td>5%</td>
<td>16%</td>
<td>6%</td>
<td>16%</td>
</tr>
</tbody>
</table>
Suggestions? 😊

Karimzadeh@psu.edu
An Intelligent Crowdsourcing System for Forensic Analysis of Surveillance Video

Khalid Tahboub, Neeraj J. Gadgil, Javier Ribera, Blanca Delgado and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana USA
Outline

• Motivation

• System Requirements

• Proposed Approach

• Experimental Results

• Conclusion and Future Work
Motivation

• Hundreds of search and rescue missions are carried out in North America and many more worldwide

• Search and rescue missions and “after-the-event” investigations require careful analysis of videos of tens or hundreds of hours of duration

• Law enforcement authorities have used some forms of crowdsourcing to reach out to volunteers
Motivation

• Crowdsourcing is often considered as an effective solution to problems that involve cognitive tasks

• Use the collective intelligence of crowds to do tasks that machines find very difficult

• Goal: Build a system to enhance crowdsourcing by incorporating machine learning methods and distinguish the ability of crowd members
System Requirements

- Web-based system
- Enable training and teaching
- Upload a set of videos as one investigation task
- Accelerate the process using machine learning methods
- No intervention required – autonomous from the perspective of law enforcement
- Produce the final result in the form of a set of video segments specifying the events of interest as one storyline
Proposed Approach

- Use a hierarchal pyramid model
- Incorporate human detection and re-identification
- Enhance the annotation platform using object tracking
Proposed Approach

Surveillance Video

Automatic Human Detection

Offense Detection (members of the crowd)

Verification of Positives

Confirmed Positives

Positives (Candidates)

Segments with humans detected

Automatic & Manual Person Re-identification
Proposed Approach

http://lorenz.ecn.purdue.edu/~vat/
Proposed Approach: Enhanced Annotation Platform

- Standard video player functionalities
- Implemented using JavaScript and HTML5
- User-friendly interface
- Temporal and spatial annotations
Proposed Approach: Enhanced Annotation Platform

• Corners are detected according to the features from accelerated segment test (FAST)

• Corner point description and matching are based on binary robust independent elementary features (BRIEF)
Proposed Approach: Training Management

- Add a label with training modules
- Assign training to workers
- Assess training performance (Pass/Fail)
Proposed Approach: Human Detection

- An OpenCV implementation of the histograms of oriented gradients (HOG)
- Linear support vector machine (SVM) classifier
- The classifier is trained with the INRIA person dataset
Proposed Approach: Person Re-identification

- Aims to find all the reappearances of a confirmed suspect in the surveillance video
- Based on color and texture features
Proposed Approach: Person Re-identification

1. Target to be re-identified selected
   - (resize, extract horizontal strips (7))
   - Extract 17 features / strip (9 HSV & 8 texture) 119-D vector
   - Similarity metric & ranking

2. Human Detection (HoG)
   - Candidate (resize, extract horizontal strips (7))
   - Extract 17 features / strip (9 HSV & 8 texture) 119-D vector

Similarity metric & ranking
Proposed Approach: Person Re-identification

Target to be re-identified

Similarity metric & Ranking

Candidates
Experimental Results

• Used public datasets: BEHAVE, Lunds University traffic, PETS 2006, PETS 2007 series, i-Lids-AVSS 2007 and videos from public safety communication research (PSCR) laboratory

• Performance of trained versus non-trained members of the crowd

• Proposed approach performance under a staged investigation
Experimental Results: Training

- Two labels for training (Suspicious bag activity, Traffic violation)
- For trained crowd, tasks were only assigned after successful completion of training modules
- Total 8 workers participated

<table>
<thead>
<tr>
<th>Event type</th>
<th>Trained group</th>
<th></th>
<th></th>
<th>Untrained group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Annotators</td>
<td>Precision</td>
<td>Recall</td>
<td>No. of Annotators</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>Suspicious bag activity</td>
<td>4</td>
<td>0.89</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Traffic violation</td>
<td>3</td>
<td>0.5</td>
<td>0.67</td>
<td>4</td>
<td>0.2</td>
<td>0.25</td>
</tr>
</tbody>
</table>

- Increase in the number of true event detection
Experimental Results: Staged Investigation

- Video content includes crowded scenes in stadiums, hallways, parking lots, stores and building entrances

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of surveillance videos</td>
<td>113</td>
</tr>
<tr>
<td>Total duration</td>
<td>115 minutes</td>
</tr>
<tr>
<td>Offense type</td>
<td>Abandoned baggage</td>
</tr>
<tr>
<td>Number of videos showing offense</td>
<td>1</td>
</tr>
<tr>
<td>Number of suspect reappearances</td>
<td>3</td>
</tr>
</tbody>
</table>
Experimental Results: Staged Investigation

- Human detection:
  - Number of true positives: 83
  - Number of false positives: 13
  - Number of true negatives: 15
  - Number of false negatives: 2

- Person Re-id:
  - Number of true positives: 1
  - Number of false positives: 5
  - Number of true negatives: 89
  - Number of false negatives: 1
Experimental Results: Staged Investigation

- Manual Person Re-id:
  - All false negatives were rejected by experienced crowd members
  - Final output is the correct identification of the offense and the three reappearances
  - Total execution time is 17 minutes

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of true positives</td>
<td>3</td>
</tr>
<tr>
<td>Number of false positives</td>
<td>2</td>
</tr>
<tr>
<td>Number of true negatives</td>
<td>91</td>
</tr>
<tr>
<td>Number of false negatives</td>
<td>0</td>
</tr>
</tbody>
</table>
Conclusion and Future work

• Proposed a crowdsourcing system for forensic analysis of surveillance video

• Used a hierarchal pyramid model to distinguish the crowd members based on their ability, experience and performance record

• Enhanced crowdsourcing by incorporating human detection, re-identification and tracking methods

• In future, we plan to incorporate more machine learning methods with our model
An Intelligent Crowdsourcing System for Forensic Analysis of Surveillance Video

Khalid Tahboub, Neeraj J. Gadgil, Javier Ribera, Blanca Delgado and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana USA
Automatic Detection of Abnormal Human Events on Train Platforms
Blanca Delgado¹, Khalid Tahboub² and Edward J. Delp²

¹Universitat Politècnica de Catalunya, Barcelona, Spain
²Purdue University, West Lafayette, IN, U.S.A.

Motivation
- Video surveillance systems that contain a large number of cameras makes the continuous monitoring of the video feeds nearly impossible task.
- Many video analytic techniques have been developed to provide automatic analysis of the video data.
- Many cities witnessed fatal incidents where people were hit and killed by trains.
- Automatic detection of abnormal human events on train platforms can save lives.

Proposed Method
- We detect moving objects and analyze their location in relative to the edge of the track bed. Train presence is investigated by analyzing motion vectors in the track bed region.

  - Background Subtraction (whole frame)
  - Final Contours of Foreground Bases
  - Approximate contours by rectangles
  - If out of the rectangular area of the track bed
    - Yes
    - No
  - Yes
    - No
  - Positive: moving object into track bed area
    - Yes
    - Waving proximity alarm
  - In motion parallel to the edge?
  - Estimate motion vectors (track bed area)
  - Is in motion parallel to the edge?
    - Yes
    - No
  - Analyze motion vectors to activate speed
  - If speed > threshold?
    - Yes
    - No
  - Train Presence
    - To detect the presence of the train, its motion information is used, taking into account that the motion of the train is parallel to the edge of the platform.
    - To estimate motion vectors, we find the most prominent corners of each frame constrained to the track bed region.
    - We use a point-based tracker to track the prominent corners which is based on Lucas Kanade Feature Tracker.
    - Displacement vectors serve as an indicator of the train speed. When the train is moving at a high speed, the displacement distance is large.

Example of prominent corners:

Results
- The experimental evaluation was conducted using our new dataset: videos have been recorded from three different views and cameras were placed at a height similar to typical surveillance cameras.

  - The dataset has been divided in short clips containing the following events:
    - No events or trains
    - Person crossing the track bed
    - Train arriving
    - Train waiting or discharging passengers
    - Train departing
  - Precision was found to be 90%, while recall was 100%.
  - Train presence was correctly identified at all times.

Conclusion and Future Work
- An experimental evaluation was conducted using a video dataset that was recorded at a local train station, precision and recall were found to be 90% and 100% respectively.
- Train presence was correctly identified at all times.
- For future work, we want to investigate:
  - Multi-camera scenarios.
  - Study the impact of objects moving in the immediate vicinity to the camera.
  - Estimate the depth of moving objects using motion information.

Publications

Acknowledgements
This work is partially funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0001.
Automatic Recognition and Interpretation of Gang Graffiti (GARI)

Edward J. Delp

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu
Acknowledgements

• This work is funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0001

• Indiana Gang Intelligence Network (INGangNetwork)

• Cook County Sheriff Department
GARI System Goals

- System that allows first responders to:
  - Collect gang graffiti and gang tattoos images
  - Analyze and interpret the images
  - Image retrieval
  - Interact with a database of images
  - Browse database using an interactive map
  - SQL database

- Platforms
  - Android and iPhone application
  - Web-based interface
  - CGAP (Civilian Gang Application Program)
GARI System Overview

1. Original image
2. Offline automatic analysis and labeling
   - Geoposition
   - Date and time
   - Extracted Features
3. Original Database
4. Filtered results
   - Info + thumbnails
5. Offline manual filtering
6. Manual labeling
   - Additional Features
7. Labeled image
   - Addition to Database

Filtered Database

Server

Purdue University
Gang Graffiti Interpretation

**SHAPE**
Simple, Straightforward

**NUMBERS**
42nd street gang

**SYMBOLS**
6-point star, pitchforks

**COLOR**
Goon Squad: Red/Black
Gang Graffiti Interpretation

**LETTERS**
East Side Gang

**TIME**
Black: 18 ST (18th Street Gang)
Red: 13 SUR (Sureños 13)

**POSITION/ALIGNMENT**
Letters at star points
Numbers in the middle
Letters at the bottom
Pitchforks upright
Image Analysis – Color Recognition

MEXICANOS MALDITOS SUREÑOS 13

18 STREET GANG

SUREÑOS 13
Image Analysis – Scene Analysis

GANG RIVALRY TRACKING

• Date: 08/19/2010
• Time: 3.25 PM
• Geo: 41.387917, 2.169919

• Date: 01/03/2011
• Time: 5.11 PM
• Geo: 41.387917, 2.169919
Gang Tattoos
User Interface

Main menu

User options

Analysis results

Image matching
Browse Database

- Browse general results
- Show graffiti on map
- Inspect specific graffiti
Augmented Reality Map
## Multiple Image Upload

### Desktop Version

### Upload Images

<table>
<thead>
<tr>
<th>Image</th>
<th>First Responder ID: 000101</th>
<th>Date and Time: 2011-06-07 09:50:06</th>
<th>Location embedded on image</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comments:**

- Unknown
- Unknown
- Unknown
- Unknown

**Gang:** Unknown

**Gang member:** No prison

<table>
<thead>
<tr>
<th>Image</th>
<th>First Responder ID: 000101</th>
<th>Date and Time: 2011-06-07 10:08:01</th>
<th>Location embedded on image</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image6.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image8.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comments:**

- Unknown
- Unknown
- Unknown
- Unknown

**Gang:** Unknown

**Gang member:** No prison

<table>
<thead>
<tr>
<th>Image</th>
<th>First Responder ID: 000101</th>
<th>Date and Time: 2011-06-07 10:16:13</th>
<th>Location embedded on image</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image9.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image10.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image11.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image12.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comments:**

- Unknown
- Unknown
- Unknown
- Unknown

**Gang:** Unknown

**Gang member:** No prison

<table>
<thead>
<tr>
<th>Image</th>
<th>First Responder ID: 000101</th>
<th>Date and Time: 2011-06-07 10:18:33</th>
<th>Location embedded on image</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image13.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image14.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image15.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image16.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comments:**

- Unknown
- Unknown
- Unknown
- Unknown

**Gang:** Unknown

**Gang member:** No prison

<table>
<thead>
<tr>
<th>Image</th>
<th>First Responder ID: 000101</th>
<th>Date and Time: 2011-06-07 10:20:12</th>
<th>Location embedded on image</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image17.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image18.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image19.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image20.png" alt="Image" /></td>
<td>Graffiti, Tattoo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comments:**

- Unknown
- Unknown
- Unknown
- Unknown

**Gang:** Unknown

**Gang member:** No prison
Browse Database
Desktop Version

- **Browse by**
  - Radius
  - Date
  - Address
  - Graffiti/Tattoo

**Browse Database**
Total: 305 graffiti
Show all images in map
Show images with real GPS in map (208)

**Database browsing screenshots**
Browse Database

Desktop Version

Date and Time: 2011-01-28 09:17:21
GPS latitude: 40.429325
GPS longitude: -86.9126441667
File Size: 664501 bytes
Height: 1552 px
Width: 2592 px
Focal length: 4.31 mm
Camera make: HTC
Camera model: HTC Desire
Image id: 1111
Show in map
Browse Database
Desktop Version

Specific image
All images in radius
Browse Database
Desktop Version
Gang Graffiti – Image Analysis

- Blur Prevention and Detection
- Light Sensor Value
- Input Image
  - Color Correction
  - Touchscreen Tracing
  - Gaussian Color Segmentation
  - Segmentation Enhancement
  - Background Stripe Removal
  - Graffiti Component Extraction
  - Graffiti Component Reconnection
  - Content-Based Image Retrieval
  - Gang Graffiti – Image Analysis

Input Image:
- Color Correction
- Touchscreen Tracing
- Gaussian Color Segmentation
- Segmentation Enhancement
- Background Stripe Removal
- Graffiti Component Extraction
- Graffiti Component Reconnection
- Content-Based Image Retrieval
- Gang Graffiti

Output Image:
- Input Image
- Color Correction
- Touchscreen Tracing
- Gaussian Color Segmentation
- Segmentation Enhancement
- Background Stripe Removal
- Graffiti Component Extraction
- Graffiti Component Reconnection
- Content-Based Image Retrieval
- Gang Graffiti
Gang Graffiti – Blur Prevention

• Camera automatically takes image after 4 second countdown
• Using smartphone’s accelerometer to detect shaking
  – If shaking, counter restarts
• Forcing camera to autofocus before taking image
Gang Graffiti – Color Correction

- Allows color correction without fiducial marker
- Non-intrusive

Light Sensor

- Automatically assigns color correction matrix to light sensor value

Before

After
Gang Graffiti – Color Segmentation
Gang Graffiti – Segmentation Enhancement

Original Image

Initial Segmentation
Block-Wise Luma
Enhanced
Gang Graffiti – Component Extraction
Gang Graffiti – Component Classification

- Symbol: 6-Point Star
  Confidence: 84%

- Symbol: Crown
  Confidence: 80%

- Symbol: Pitchfork
  Confidence: 75%

- Symbol: Unknown

- Symbol: Unknown
Gang Graffiti – Find Similar Images

Retrieval: Accuracy: 99.10%, Time: 70 ms
Gang Tattoo – Image Analysis

ARCHIVING

Database image → SIFT feature extraction → Set of SIFT Descriptors → Database (Images, Descriptors) → Images, Gang symbol meaning

RETRIEVAL

Cropped image → SIFT feature extraction → Set of SIFT descriptors → SIFT descriptor matching → Scores → Top rank-20 images
Gang Tattoo – Image Analysis

Original Image

Cropped Image

Image Analysis

Image Matching

Analysis Results

GARI/VIPER

August 2014
Gang Tattoo – Find Similar Images

Retrieval Accuracy: 81.37% Top-10 (57.84% Top-1)
GARI Availability

• The GARI system is available free to any law enforcement agency

• INGang Members can receive GARI now (all images are shared)

• Not in Indiana
  – MOU with Purdue University
  – Agency must agree to host a GARI server or cooperate with an agency currently using GARI
  – Software and server setup (and support) provided by VACCINE/Purdue University

For more information: gari@ecn.purdue.edu
Characterizing the Uncertainty of Classification Methods and Its Impact on the Performance of Crowdsourcing

Javier Ribera, Khalid Tahboub, Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana
Outline

1. Motivation
2. Uncertainty characterizations
3. Crowdsourcing
4. Crowd flow estimation
5. Results
6. Conclusions
Motivation

• Video surveillance applications demand robust video analytics

• Changing environments, occlusions, shadows, video distortions

• Uncertainty in classification methods is quantified
Motivation

- Incorporate crowdsourcing
- Online active learning
- How to aggregate the contributions of the crowd?
Uncertainty

- Classification is crucial to many methods

- New data points need to be labeled using the given model
Uncertainty Characterizations

• When is the unlabeled data point “close enough” to the nearest labeled data point?

• Uncertainty quantification of the classification: $\mu$

• Quantify and threshold (maximum uncertainty allowed)

$$\mu < \alpha$$
Uncertainty Characterizations

– 1st uncertainty characterization

\[ \mu_1 = \frac{d_1}{d_2} < \alpha \]
Uncertainty Characterizations

- 1st uncertainty characterization

\[ \mu_1 = \frac{d_1}{d_2} < \alpha \]

\[ \alpha = 1 \]

\[ \alpha = 0.6 \]

\[ \alpha = 0.4 \]

\[ \alpha = 0.8 \]
Uncertainty Characterizations

– 2nd uncertainty characterization

$$\mu_2 = \sum_{i=2}^{L} \frac{d_1}{d_i} < \alpha$$
Uncertainty Characterizations

– 2nd uncertainty characterization

\[ \mu_2 = \sum_{i=2}^L \frac{d_1}{d_i} < \alpha \]
Uncertainty Characterizations

- 3rd uncertainty characterization

\[ \mu_3' = \frac{d_{12}}{d_b} \]

\[ \mu_3 = \frac{1}{1 + \frac{1}{\mu_3'}} \]
Crowdsourcing

- Crowdsourcing is incorporated when the classification uncertainty is estimated to be high

$$\mu > \alpha \Rightarrow \text{crowdsourcing}$$
Crowdsourcing

• How to aggregate the labels of the crowd?

labels

$\tau_3$

$\tau_1$

$\tau_2$

Unique label for $t$

$l_1$

$l_2$

$l_3$

$l_4$

$l_5$
Crowdsourcing

1. Average crowd labels for the same data point
2. Use the subset with lowest variance (discard the outliers)
Active Learning

- New labeled data point is incorporated into the model to enhance future classifications
Crowd Flow Estimation

• Application: estimate the number of people that have crossed a specific region

Tripwire (red)
Region of Interest (blue)
Crowd Flow Estimation

- Estimation: proportional to the number of foreground pixels in the Tripwire

\[ \nu = \frac{S_N}{C} \]

- \( C \) is the number of pixels seen per person: depends on the level of crowdedness (occlusions)
Crowd Flow Estimation

1. Compute foreground segmentation in the Tripwire and accumulate foreground pixels.
2. Estimate level of crowdedness in the Region of Interest by means of texture features.
   - Feature vector: 16-D texture feature vector built using the Gray Level Co-occurrence Matrix.
Crowd Flow Estimation

- The texture feature vector $t$ is the testing data point for our model in a 16-D texture feature space.
- The “reference” feature vectors $\tau_i$ have been manually labeled in the training stage with their corresponding $C$, based on their level of crowdedness.
Experiments

• UCSD pedestrian dataset

• Experiments run with degraded quality

• Which uncertainty characterization leads to better results?

• How crowd labels aggregation methods impact the performance?
Results

• With no crowdsourcing ($\alpha=1$):

<table>
<thead>
<tr>
<th></th>
<th>Average error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original quality</td>
<td>8.1%</td>
</tr>
<tr>
<td>Degraded quality</td>
<td>31.7%</td>
</tr>
</tbody>
</table>

• Degradation consisted of boosting contrast and compressing the video file to 40%
Results

- Comparison of uncertainty characterizations:

<table>
<thead>
<tr>
<th>Degraded quality</th>
<th>Uncertainty characterization</th>
<th>$\alpha$</th>
<th>Average error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># 1</td>
<td>0.6</td>
<td>15.2 %</td>
</tr>
<tr>
<td></td>
<td># 2</td>
<td>0.5</td>
<td>12.5 %</td>
</tr>
</tbody>
</table>

- Degradation consisted in boosting contrast and compressing the video file to a 40%
Results

• Comparison of crowd aggregations:

<table>
<thead>
<tr>
<th>Aggregation of K of the crowd members</th>
<th>Average error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>K = 5</td>
<td>26 %</td>
</tr>
<tr>
<td>K = 3</td>
<td>19.5 %</td>
</tr>
</tbody>
</table>
Results

- Active learning
- The method learns from crowdsourcing and the model adapts to the new conditions.
Conclusions

• Video analytics can be enhanced by characterizing uncertainties and use crowdsourcing to address it.

• Different characterizations of the uncertainty of a classification are proposed and evaluated.

• Aggregating results from the labelers with smallest variance yields better results.

• Future work includes adding the history of the performance of the crowd members into the model.

• Further performance analysis.
Characterizing the Uncertainty of Classification Methods and Its Impact on the Performance of Crowdsourcing

Javier Ribera, Khalid Tahboub, Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana
Efficient Graph-Cut Tattoo Segmentation

Joonsoo Kim, Albert Parra, He Li, and Edward J. Delp
Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu
Outline

• Motivation

• Overview of Existing Methods

• Proposed Method

• Experimental Results

• Conclusions and Discussion
Motivation

• Law enforcement is interested in exploiting tattoos as an information source to identify, track and prevent gang-related crimes

• The use of tattoos for person identification and tattoo symbol interpretation is an interesting problem for law enforcement
Motivation

• In a retrieval system tattoo segmentation is an important step for retrieval accuracy since segmentation removes background information in a tattoo image.

• Many existing methods for tattoo segmentation have problems:
  – Some existing methods use pre-cropped images whose background is skin only.
  – Existing methods using uncropped image are not robust when background has color similar to skin tones, there is an illumination change, and an image has diverse skin tones.
Overview of Existing Methods


- The HSV color space is used to find skin regions
- Suggested a fixed color range to detect skin pixels - restricted to Asian people
- Can be affected by illumination, background, and camera characteristics
Overview of Existing Methods


- Used histogram on image intensity (gray-scale) and multiple clusters
- Merge most of the clusters
- K-means clustering is used in RGB space to distinguish a tattoo/skin region
Overview of Existing Methods


• Region near the tattoo is detected based on a visual saliency model
• GrabCut segmentation is used in the saliency region
• Edge map with QCC (Quasi Connected Components) is combined with the Grabcut segmentation result
• The accuracy of this method highly depends on the accuracy of a visual saliency map
Proposed Method

• Two important assumptions
  – Tattoos are surrounded by skin (closed loop)
  – Edges between tattoo and skin
• Tattoo detection becomes detecting skin pixels around tattoo
  – Skin model
  – Regions near image edges are only considered for “possible segmentation regions”
  – Segmentation errors outside of the possible segmentation regions can be avoided
  – Computationally efficient
Approach

(a) Original image

(b) Possible Segmentation Regions

(c) Skin Probability map

(d) Visual Saliency map

(e) Graph-Cut Segmentation

(f) Tattoo segmentation
Our Method

- Canny edge detector is used to find edges for each RGB color channel
- The edge regions from each color channel are combined
- Morphological dilation is used to the combined edge regions for finding the possible segmentation regions

Skin Color Model (GMM) → Detection of Possible Segmentation Regions → Graph-Cut Segmentation → Saliency Probability Map (GBVS) → Image → Post-Processing → Segmented Tattoo Regions
Segmentation

- Graph–cut segmentation – minimize the Gibbs energy

\[ E(x) = \sum_{i \in I} D(x_i) + \lambda \sum_{i \in I, j \in N_i} V(x_i, x_j), \]

- \( D(x_i) \): the data term
- \( V(x_i, x_j) \): the smoothness term
- \( I \): image
- \( N_i \): neighborhood pixels of pixel \( I \)
- \( x_i \in \{ 0(\text{background}), 1(\text{foreground}) \} \)
**Segmentation**

- **$D(x_i)$**: the data term
  \[ D(x_i) = w_1 D_1(x_i) + w_2 D_2(x_i), \]

- **$D_1(x_i)$**: energy for skin or non-skin model based on GMM

- **$D_2(x_i)$**: energy for normalized saliency map
  \[ V(x_i, x_j) = |x_i - x_j| \cdot f(C_{ij}) \]

- **$V(x_i, x_j)$**: the smoothness term
  \[ f(\varepsilon) = 1/(1+\varepsilon), \]
  \[ C_{ij} = \|C_i - C_j\|^2 \]
  \[ C_i : \text{YCbCr color of pixel } i \]
Skin Model

- GMM skin model - train on diverse skin tones
  \[ D(x_i) = w_1 D_1(x_i) + w_2 D_2(x_i), \]
  \[ D_1(x_i) : \text{energy for skin/non-skin} \]
  \[ D_1(x_i = 1) = -\log(p_1(x_i = 1)) \]

\[ p(x_i = 1) = \sum_{j=1}^{M} \pi_j g(C_i | u_j, \Sigma_j) \]

\( C_i \): YCbCr color of \( i^{th} \) pixel
\( \pi_j \): \( j^{th} \) mixture weight
\( M \): \# of mixture components
\( g() \): Gaussian density function
\( u_j, \Sigma_j \): Gaussian parameters
- \( \pi_j u_j, \Sigma_j \) are estimated using EM
Skin Model

- Cannot estimate all non-skin colors - use adaptive threshold for non-skin model

\[ D_1(x_i = 0) = -\log(wt_1 \times \frac{1}{n_I} \sum_{i \in I} p_1(x_i = 1)) \]

- Adaptive threshold helps segment skin when the skin colors in an image do not fit the skin model
Saliency

- Case for color similar to skin in background, use visual saliency map to focus on skin near tattoo
  \[
  D(x_i) = w_1D_1(x_i) + w_2D_2(x_i),
  \]

- \( D_2(x_i) \) : energy for normalized saliency map

- Normalized GBVS saliency map is used for \( p_2(x_i=1) \)
  \[
  D_2(x_i = 1) = -\log(p_2(x_i = 1))
  \]

- Adaptive threshold is used for \( p_2(x_i=0) \)
  \[
  D_2(x_i = 0) = -\log(wt_2 * \frac{1}{n_I} \sum_{i \in I} p_2(x_i = 1))
  \]
Post Processing

- Detect both skin pixels near boundaries of the body and skin pixels near tattoos
- For skin pixels near tattoos, examine which set of skin pixels connected with each other forms a closed contour with holes inside
- Graph-cut segmentation errors - small holes (false contours)
- We check

\[
\frac{n_c + n_h}{n_b} \geq t_f
\]

- \( n_c \): the number of pixels - closed contour
- \( n_h \): the number of pixels for holes inside the closed contour
- \( n_b \): the number of pixels in the minimum bounding rectangular box including the closed contour
Post Processing

- Example of Post processing

- Minimum bounding box including false contour
- Minimum bounding box including a tattoo

(a) Original image  (b) Graph-Cut Segmentation  (c) Tattoo segmentation after post processing
Experimental Results

- Tattoo images - Indiana Gang Network (INGangNetwork)

- TDSD dataset used to train skin model - includes diverse skin tones

- GMM - M=5

- Image retrieval system - do not segment the tattoo inside the closed contour found after post processing

- Compared with existing methods: Allen’s method and Heflin’s method
Experimental Results

(a) Original image    (b) Allen’s method    (c) Heflin’s method    (d) Proposed method
Experimental Results

(a) Original image  (b) Allen’s method  (c) Heflin’s method  (d) Proposed method
Experimental Results

(a) Original image  (b) Allen’s method  (c) Heflin’s method  (d) Proposed method
Experimental Results

(a) Original image  (b) Allen’s method  (c) Heflin’s method  (d) Proposed method
Experimental Results

(a) Original image  (b) Allen’s method  (c) Heflin’s method  (d) Proposed method
Experimental Results

- Our method segmented tattoo regions incorrectly when tattoo regions are very close to boundaries of the body and there are hair regions with strong edges inside a body.
Conclusions and Discussion

• Using skin pixels near image edges allows one to segment tattoo regions more efficiently

• Our method sometimes fails to extract tattoo regions correctly when there are tattoo regions close to boundaries of the body and/or near “hair regions”

• In the future we will investigate techniques to detect these regions
Efficient Graph-Cut Tattoo Segmentation

Joonsoo Kim, Albert Parra, He Li, and Edward J. Delp
Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu
GARI
A Gang Tattoo and Graffiti System

He Li, Joonsoo Kim, Jiaju Yue, Edward J. Delp
(ace@ecn.purdue.edu)

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
Acknowledgements

• This work is funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0001

• Cook County Sheriff Department

• Indiana Gang Intelligence Network (INGangNetwork)
GARI System Goals

• System that allows first responders to:
  – Collect gang graffiti and gang tattoos images
  – Analyze and interpret the images
  – Image retrieval
  – Interact with a database of images
  – Browse database using an interactive map
  – SQL database

• Platforms
  – Android and iPhone application
  – Web-based interface (report generation, bulk image upload, same features as mobile app)
  – CGAP (Civilian Gang Application Program)
GARI System Overview

1. Original image
2. Offline automatic analysis and labeling
   - Geoposition
   - Date and time
   - Extracted Features
3. Original Database
4. Filtered results Info + thumbnails
5. Offline manual filtering
6. Manual labeling
   - Additional Features
7. Labeled image
   - Addition to Database
Gang Graffiti Interpretation

SHAPE
Simple, Straightforward

NUMBERS
42nd street gang

SYMBOLS
6-point star, pitchforks

COLOR
Goon Squad: Red/Black
Gang Graffiti Interpretation

**LETTERS**
East Side Gang

**TIME**
Black: 18 ST (18th Street Gang)
Red: 13 SUR (Sureños 13)

**POSITION/ALIGNMENT**
Letters at star points
Numbers in the middle
Letters at the bottom
Pitchforks upright
Image Analysis – Color Recognition

MEXICANOS MALDITOS SUREÑOS 13

18 STREET GANG

SUREÑOS 13
Image Analysis – Scene Analysis

GANG RIVALRY TRACKING

- Date: 08/19/2010
- Time: 3.25 PM
- Geo: 41.387917, 2.169919

- Date: 01/03/2011
- Time: 5.11 PM
- Geo: 41.387917, 2.169919
Gang Tattoos
Mobile App User Interface

Main menu

User options

Analysis results

Image matching
Browse Database

Browse general results

Show graffiti on map

Inspect specific graffiti
Augmented Reality Map
Browse Database

Desktop Version

• Browse by
  – Radius
  – Date
  – Address
  – Zipcode
  – Firstresponder ID

Browse Database
Total: 305 graffiti

Show all images in map
Show images with real GPS in map (208)

Date/Time captured: 2011-01-28 10:47:47
Address: 2413 Yandes St, Indianapolis, IN 46205, USA
More information
Image ID: 1682

Database browsing screenshots
Browse Database

Desktop Version

Date and Time: 2011-01-28 09:17:21
GPS latitude: 40.429325
GPS longitude: -86.9126441667
File Size: 664501 bytes
Height: 1552 px
Width: 2592 px
Focal length: 4.31 mm
Camera make: HTC
Camera model: HTC Desire
Image id: 1111
Show in map
Browse Database

Desktop Version

Specific image

All images in radius
Browse Database

Desktop Version
Create Database Report

Use Shift/Ctrl keys to select multiple fields.

Table Fields (Columns):

- Image ID
- First responder name
- First responder ID
- Upload date and time
- Image size (bytes)
- Image height (pixels)
- Image width (pixels)
- Camera make
- Camera model
- GPS longitude
- GPS latitude
- Address (given GPS coordinates)
- City (given GPS coordinates)
- County (given GPS coordinates)
- State (given GPS coordinates)
- ZIP code (given GPS coordinates)
- Gang Name
- Body Part ID
- Comments

# of Entries (Rows):
Select one of the following sorting options:

- Sort Entries by Gangname:
- Sort Entries by zip code:
- Sort Entries by Date and Time:
- Sort Entries by range of Date:

Select:

- Z-A
- A-Z
- DESC
- ASC

Newest first
Oldest first

All From To:

Submit
Create Database Report

<table>
<thead>
<tr>
<th>zip</th>
<th>firstrespo</th>
<th>filedatetimeuplo</th>
<th>gpslong</th>
<th>gpslatitude</th>
<th>address</th>
<th>zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:09</td>
<td>-77.034</td>
<td>38.904</td>
<td>1424 L St NW</td>
<td>20005</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.885</td>
<td>40.43</td>
<td>1018 Roberts St</td>
<td>47904</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.875</td>
<td>40.416</td>
<td>26 Grant St</td>
<td>47904</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.865</td>
<td>40.41</td>
<td>2810 Kossuth St</td>
<td>47904</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-122.41</td>
<td>37.786</td>
<td>3319 Longlois Dr</td>
<td>47904</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.885</td>
<td>40.407</td>
<td>905 S 10th St</td>
<td>47905</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.872</td>
<td>40.405</td>
<td>1141 S 22nd St</td>
<td>47905</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.885</td>
<td>40.402</td>
<td>1000 Potomac Ave</td>
<td>47905</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:09</td>
<td>-86.846</td>
<td>40.422</td>
<td>3607 Driftwood Dr S</td>
<td>47905</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.921</td>
<td>40.451</td>
<td>2601-2699 Trace 26</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.912</td>
<td>40.428</td>
<td>Memorial Mall</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.92</td>
<td>40.431</td>
<td>1165 W Stadium Ave</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.929</td>
<td>40.423</td>
<td>136-174 Nimitz Dr</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.936</td>
<td>40.418</td>
<td>2-98 Aviation Drive.</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.927</td>
<td>40.417</td>
<td>148 Arnold Dr</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:09</td>
<td>-86.931</td>
<td>40.452</td>
<td>1390-1398 Palmer Dr</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:09</td>
<td>-86.928</td>
<td>40.452</td>
<td>1300-1306 Palmer Dr</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:09</td>
<td>-86.928</td>
<td>40.454</td>
<td>2562-2582 Neil Armstrong</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.908</td>
<td>40.433</td>
<td>812 Vine St</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-77.08</td>
<td>38.939</td>
<td>465 Northwestern Av</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:09</td>
<td>-77.076</td>
<td>38.936</td>
<td>465 Northwestern Av</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-122.405</td>
<td>37.785</td>
<td>510 Hayes St</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-122.399</td>
<td>37.786</td>
<td>610 Purdue Mall</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>0</td>
<td>0</td>
<td>422 Vine St</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>0</td>
<td>0</td>
<td>773-799 Northwestern</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>0</td>
<td>0</td>
<td>1203 W State St</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:09</td>
<td>0</td>
<td>0</td>
<td>465 Northwestern Av</td>
<td>47906</td>
</tr>
<tr>
<td>1101</td>
<td>5/16/2011</td>
<td>22:04</td>
<td>-86.919</td>
<td>40.424</td>
<td>S Russell Dr</td>
<td>47907</td>
</tr>
<tr>
<td>1112</td>
<td>12/12/2014</td>
<td>13:34</td>
<td>-86.9125</td>
<td>40.42934</td>
<td>501 Purdue Mall</td>
<td>47907</td>
</tr>
<tr>
<td>1112</td>
<td>12/12/2014</td>
<td>15:10</td>
<td>-86.9125</td>
<td>40.42933</td>
<td>501 Purdue Mall</td>
<td>47907</td>
</tr>
<tr>
<td>200069</td>
<td>8/7/2014</td>
<td>8:37</td>
<td>-87.9708</td>
<td>42.07957</td>
<td>213 South Dryden Pl</td>
<td>60004</td>
</tr>
<tr>
<td>200069</td>
<td>11/20/2014</td>
<td>9:53</td>
<td>-88.0135</td>
<td>42.15278</td>
<td>2400-2498 Crystal Ln</td>
<td>60004</td>
</tr>
</tbody>
</table>
Gang Graffiti – Image Analysis

- Input Image
- Color Correction
- Touchscreen Tracing
- Gaussian Color Segmentation
- Segmentation Enhancement
- Background Stripe Removal
- Graffiti Component Extraction
- Graffiti Component Reconnection
- Content-Based Image Retrieval
- Gang Graffiti

Light Sensor Value

Blur Prevention and Detection
Gang Graffiti – Color Correction

- Allows color correction without fiducial marker
- Non-intrusive

- Automatically assigns color correction matrix to light sensor value

Before | After
--- | ---
![Before Image] | ![After Image]
Gang Graffiti – Color Segmentation
Gang Graffiti – Segmentation Enhancement

Original Image

Initial Segmentation

Block-Wise Luma

Enhanced
Gang Graffiti – Stripe Removal

Segments to be removed (green) and ignored (blue)
Gang Graffiti – Component Reconnection

Before

After
Gang Graffiti – Component Extraction
Gang Graffiti – Component Classification

Symbol: 6-Point Star
Confidence: 84%

Symbol: Crown
Confidence: 80%

Symbol: Pitchfork
Confidence: 75%

Symbol: Unknown

Symbol: Unknown
Gang Graffiti – Find Similar Images

Retrieval: Accuracy: 99.10%, Time: 70 ms
Gang Tattoo – Image Analysis

Original Image

Cropped Image

Image Analysis

Image Matching

Analysis Results

GARI/VIPER

June 8, 2015
Gang Tattoo – Find Similar Images

[Images of different tattoo designs related to gang symbols, including swastikas and symbolic motifs]
Tattoo Recognition Technology – Challenge (Tatt-C)

Tattoo Identification And Region of Interest
Introduction for Recognition Technology-Challenge

• The Tattoo Recognition Technology-Challenge (Tatt-C) : a competition to test image-based tattoo matching technology. This challenge will assess the capability of tattoo recognition methods to detect and retrieve tattoos.

• We have participated in
  – Tattoo Identification : matching different instances of the same tattoo images from the same subject over time
  – Region of Interest : matching a small region of interest that is contained in a larger image
Tattoo Identification

• Block Diagram

- Database (Gallery) Image
  - SIFT Feature Extraction
  - Set of MHLC Descriptors
  - Visual Saliency Map (GBVS)
  - Database (Images, Descriptors)

- Query (Probe) Image
  - SIFT Feature Extraction
  - Set of MHLC Descriptors
  - Robust Similarity Local Descriptor Matching
  - Visual Saliency Map (GBVS)

- Number of Extracted SIFT Feature Points

- Scores
  - Top Rank-20 Images

GARI/VIPER  June 8, 2015
Region of Interest

- Block Diagram

**ARCHIVING**

**RETRIEVAL**

Database (Gallery) Image → SIFT Feature Extraction → Set of LSSIM+SIFT Descriptors → Database (Images, Descriptors) → Images

Query (Probe) Image → SIFT Feature Extraction → Set of LSSIM+SIFT Descriptors → Robust Weighted Distance Similarity Matching → Scores → Top Rank-20 Images

LSSIM : Local Self-SIMilarities
Experimental Results

- Definition of probe, gallery, and background image

  - Probe image: A query image used to search against the dataset – ideally find the true matches for a probe image

  - Gallery: A set of images that will be searched against to find the matches for the probe image

  - Background image: Irrelevant images that can be added to the gallery to test the robustness of the method – “distractor”
Experimental Results

Dataset Information

<table>
<thead>
<tr>
<th></th>
<th>Tattoo Identification</th>
<th>Region of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of gallery images</td>
<td>215</td>
<td>157</td>
</tr>
<tr>
<td>Number of probe images</td>
<td>157</td>
<td>297</td>
</tr>
<tr>
<td>Number of background images (&quot;Distractor&quot;)</td>
<td>4332</td>
<td>4332</td>
</tr>
</tbody>
</table>

In some experiments the gallery and background are merged
Experimental Results

• Performance Metrics
  – Cumulative Match Characteristic (CMC)
    • CMC = The probability that one or more correctly matched images for a probe is observed within the top K ranks
Experimental Results

- **Cumulative Match Characteristic (CMC):**
  - $CMC(k) = \sum P(r)$ where $r \leq k$

- **Rank Probability Mass Function**
  - $P(r) = \text{at rank } r$, the number of correctly matched probe images in the candidate list divided by the total number of probe images

- Given a database and set of queries (probes), this probability can be estimated

| Query(probe) | Image set | rank1 | rank2 | rank3 | rank4 | ...
|--------------|-----------|-------|-------|-------|-------|-------
| P(r)         | 1/3       | 2/3   | 0     | 0     |       |
| CMC(k)       | 1/3       | $1/3 + 2/3 = 1$ | 1     | 1     |       |

: Correctly retrieved images
Experimental Results

5-fold Cross Validation

• We conducted 5 separate experiment using leave-one-out-cross validation

• In each experiment, 4 of subset were combined to form a training set, with the 5th subset used for testing

• Using this scheme, we determined the weight between SIFT descriptor and MHLC descriptor in tattoo identification and the weight between SIFT descriptor and self similarity descriptor in region of interest
Experimental Results (Tattoo Identification)

• Example of Matched Results
  (Without Background Images)

Top 10 ranks (left to right, top to bottom)

Probe Image probe_004.jpg

Matched
Experimental Results (Tattoo Identification)

• Example of Matched Results (Without Background Images)

Top 13 ranks (left to right, top to bottom)

Probe Image: probe_053.jpg

Matched
Experimental Results (Tattoo Identification)

- Example of Matched Results (Without Background Images)

Probe Image: `probe_056.jpg`

Top 10 ranks (left to right, top to bottom): Matched
Experimental Results
(Tattoo Identification)

- Example of Matched Results
(With Background Images)

Probe Image
probe_004.jp
g

Top 10 ranks (left to right, top to bottom)
Experimental Results (Tattoo Identification)

• Example of Matched Results (With Background Images)

Top 10 ranks (left to right, top to bottom)

No Match

Probe Image
probe_053.jpg
Experimental Results (Tattoo Identification)

- Example of Matched Results (With Background Images)

Top 10 ranks (left to right, top to bottom)

Probe Image probe_056.jpg

Matched
Experimental Results
(Tattoo Identification)

• Comparison Between Different Methods

<table>
<thead>
<tr>
<th>Top rank</th>
<th>Proposed</th>
<th>MSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.99376</td>
<td>0.96814</td>
</tr>
<tr>
<td>3</td>
<td>0.99376</td>
<td>0.99376</td>
</tr>
<tr>
<td>5</td>
<td>0.99376</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0.99376</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

CMC vs Top K rank Table
(K=1,3,5,10,15,20)
(Without background images)

CMC vs Top K rank Plot
(Without background images)
Experimental Results (Tattoo Identification)

• Comparison with Phase 1 results (Without Background Images)

![Graph showing Cumulative Accuracy vs Rank for different algorithms.]

- Purdue: 0.994
Experimental Results (Tattoo Identification)

• Comparison with Phase 1 results (With Background Images)

Purdue: 0.987
Experimental Results (Region of Interest)

- Example of Matched Results (Without Background Images)
  Top 10 ranks (left to right, top to bottom)

Matched!!

Probe Image
probe_001.jpg
Experimental Results (Region of Interest)

- Example of Matched Results (Without Background Images)

Top 10 ranks (left to right, top to bottom)

Probe Image
probe_002.jpg

Matched!!
Experimental Results (Region of Interest)

• Example of Matched Results (Without Background Images)

Top 10 ranks (left to right, top to bottom)

Matched!!

Probe Image
probe_0013.jpg
Experimental Results (Region of Interest)

- Example of Matched Results (With Background Images)

Top 10 ranks (left to right, top to bottom)

Probe Image
probe_001.jpg
Experimental Results (Region of Interest)

• Example of Matched Results (With Background Images)

   Top 10 ranks (left to right, top to bottom)
   No Matching…

Probe Image
probe_002.jpg
Experimental Results (Region of Interest)

- Example of Matched Results (With Background Images)

Top 10 ranks (left to right, top to bottom)

Probe Image probe_013.jpg

Matched!!
Experimental Results  
(Region of Interest)

• Comparison Between Different Methods

<table>
<thead>
<tr>
<th>Top rank</th>
<th>Proposed</th>
<th>MSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.96614</td>
<td>0.94262</td>
</tr>
<tr>
<td>3</td>
<td>0.96964</td>
<td>0.95262</td>
</tr>
<tr>
<td>5</td>
<td>0.97962</td>
<td>0.96262</td>
</tr>
<tr>
<td>10</td>
<td>0.98648</td>
<td>0.98296</td>
</tr>
<tr>
<td>15</td>
<td>0.98982</td>
<td>0.98296</td>
</tr>
<tr>
<td>20</td>
<td>0.99316</td>
<td>0.98648</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Top rank</th>
<th>Proposed</th>
<th>MSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9156</td>
<td>0.88858</td>
</tr>
<tr>
<td>3</td>
<td>0.92912</td>
<td>0.90542</td>
</tr>
<tr>
<td>5</td>
<td>0.93914</td>
<td>0.9156</td>
</tr>
<tr>
<td>10</td>
<td>0.94246</td>
<td>0.9358</td>
</tr>
<tr>
<td>15</td>
<td>0.94596</td>
<td>0.9358</td>
</tr>
<tr>
<td>20</td>
<td>0.94596</td>
<td>0.93912</td>
</tr>
</tbody>
</table>

CMC vs Top K rank Table  
(K=1,3,5,6,10,15,20)  
(Without background images)
Experimental Results (Region of Interest)

• Comparison with Phase 1 results (Without Background Images)

![Graph showing cumulative accuracy vs. rank for different algorithms]

- CEA_1: 0.926
- CEA_2: 0.929
- MorphoTrak: 0.973
- Purdue: 0.983
Experimental Results
(Region of Interest)

- Comparison with Phase 1 results
  (Without Background Images)

![Graph showing cumulative accuracy comparison between algorithms]

- Purdue: 0.942
Issues and Problems

• How do the methods scale?

• How much domain specific/location specific information is needed – it not not just an image analysis problem

• User interface!

• How does an graffiti or tattoo image analysis system fit into existing law enforcement work flow
GARI Availability

• The GARI system is available free to any law enforcement agency
  – MOU required
  – Hosting of the GARI server

For more information: gari@ecn.purdue.edu
Video-Based Re-Identification of Subjects for The Greater Cleveland Regional Transit Authority

eBOLO

These slides supersede the slides dated March 6, 2015
We acknowledge support of eBOLO from the U.S. Department of Homeland Security and the DHS Centers of Excellence
eBOLO Goals

• Demonstrate a proof of concept of a surveillance system that would allow GCTRA to identify persons of interest from a video and determine if and when they re-appear at a later date

• Add a subject to the eBOLO list by manually highlighting them; use computer vision and image processing techniques to generate the “eBOLO fingerprint”
eBOLO Goals

• Use eBOLO fingerprint to continuously monitor the video surveillance system to determine when that subject reappears in the video => flag or alarm GCRTA that the subject has reappeared and provide location details

• Provide recommendations to GCTRA with respect to infrastructure improvements (e.g., camera types, camera placement, background networking and data storage) so that an eBOLO system would be more effective
eBOLO Research Team

- Octavia I. Camps
  - Mengran Gou, Tom Hebble, Sadjad Asgari Esfeden
- Edward J. Delp
  - Khalid Tahboub, Blanca Delgado
- Patrick J. Flynn
  - Joel Brogan, Nikita Amelchenko, Amy Li
- Yun Fu
  - Ming Shao
- Richard J. Radke
  - Srikrishna Karanam, Yang Li, Gyanendra Sharma
Body Matching

Octavia I. Camps
  • Mengran Gou, Tom Hebble, Sadjad Asgari Esfeden

Edward J. Delp
  • Khalid Tahboub, Blanca Delgado

Richard J. Radke
  • Srikrishna Karanam, Yang Li, Gyanendra Sharma
Dataset Collection

• In October 2014, the team conducted experiments and collected extensive data from 5 stations: Brookpark, Puritas, West Park, Triskett, and West 117

• A total of 266 hours of footage from 9 cameras was collected
# Camera Fields of View

<table>
<thead>
<tr>
<th>Location</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>BrookPark Bus Wait</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>BrookPark Tunnel 1</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>BrookPark Tunnel 2</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>Puritas Tunnel</td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>Triskett Bridge</td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>Triskett Main Entrance</td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>West 117th Lobby West</td>
<td><img src="image7.png" alt="Image" /></td>
</tr>
<tr>
<td>West 117th Tunnel</td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td>West Park Tunnel</td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>
# Resolutions and Frame Rate

<table>
<thead>
<tr>
<th>Station</th>
<th>Resolution</th>
<th>Max Resolution</th>
<th>Frame Rate (fps)</th>
<th>Max Frame Rate (fps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brookpark</td>
<td>352 x 240</td>
<td>720 x 480</td>
<td>3.75</td>
<td>30</td>
</tr>
<tr>
<td>Puritas</td>
<td>640 x 480</td>
<td>N/A</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>West Park</td>
<td>352 x 240</td>
<td>N/A</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>Triskett</td>
<td>352 x 240</td>
<td>720 x 480</td>
<td>3.75</td>
<td>30</td>
</tr>
<tr>
<td>W117</td>
<td>352 x 240</td>
<td>720 x 480</td>
<td>3</td>
<td>30</td>
</tr>
</tbody>
</table>

- Due to system constraints, the maximum resolution and frame rate for each camera is rarely achievable
Ground Truth Generation

- Students manually outlined re-appearances of subjects from the data collection - this created what is known as “ground truth” data – this data will be used for training and validation of the analysis methods

- The sizes and appearances of the subjects vary considerably

- Several “distractor” pedestrian bounding boxes were also generated
Camera Locations For Ground Truth Data

• Two views were selected for ground truthing:
  – Triskett Bridge
  – West117 Tunnel

• These are camera locations with significant illumination changes and a clear view of people as they enter/leave the camera field of view

• The selected segments are for different times of the day over two separate days
Camera Locations For Ground Truth Data

• Summary of the videos ground truthed for the two cameras:

  West117 Tunnel

<table>
<thead>
<tr>
<th>Video File Name</th>
<th>Date</th>
<th>Start Time</th>
<th>End time</th>
</tr>
</thead>
<tbody>
<tr>
<td>117_tunnel_10a-11a</td>
<td>16-Oct-14</td>
<td>10:00 AM</td>
<td>11:00 AM</td>
</tr>
<tr>
<td>117_tunnel_5p-6p</td>
<td>16-Oct-14</td>
<td>5:00 PM</td>
<td>6:00 PM</td>
</tr>
<tr>
<td>117_tunnel_oct17_1145a-1245p</td>
<td>17-Oct-14</td>
<td>11:45 AM</td>
<td>12:45 PM</td>
</tr>
</tbody>
</table>

  Triskett Bridge

<table>
<thead>
<tr>
<th>Video File Name</th>
<th>Date</th>
<th>Start Time</th>
<th>End time</th>
</tr>
</thead>
<tbody>
<tr>
<td>tsktt_bridge_oct17_1140a-1240p</td>
<td>17-Oct-14</td>
<td>11:40 AM</td>
<td>12:20 PM</td>
</tr>
<tr>
<td>tskt_bridge_1030a-1130a</td>
<td>16-Oct-14</td>
<td>5:30 PM</td>
<td>6:30 PM</td>
</tr>
<tr>
<td>tskt_bridge_1031a-1131a</td>
<td>16-Oct-14</td>
<td>10:30 AM</td>
<td>11:30 AM</td>
</tr>
</tbody>
</table>
Ground Truth Data

• The ground truth data was organized in terms of appearances - an appearance is a video segment outlining one person as he or she enters the camera field of view and walks/loiters until he/she leaves the camera field of view

• Multi-shot images outlining the person during the entire time he/she is within the camera field of view are generated and stored as part of the ground truth data

• An example of Ground Truth data- single subject appearance at West177-tunnel:
Ground Truth Data Organization

- Ground truth data was divided into training and testing sets
- Testing data is used for method evaluation
- Training data is used to tune method parameters
- Summary of testing data:

<table>
<thead>
<tr>
<th>Triskett-Bridge</th>
<th>West177-Tunnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1 16 appearances</td>
<td>Student 1 13 appearances</td>
</tr>
<tr>
<td>Student 2 28 appearances</td>
<td>Student 2 19 appearances</td>
</tr>
<tr>
<td>Student 3 27 appearances</td>
<td>Student 3 20 appearances</td>
</tr>
<tr>
<td>Student 4 18 appearances</td>
<td>Student 4 9 appearances</td>
</tr>
<tr>
<td>Student 5 16 appearances</td>
<td>Student 5 11 appearances</td>
</tr>
<tr>
<td>Student 6 16 appearances</td>
<td>Student 6 12 appearances</td>
</tr>
<tr>
<td>Non-students 60 appearances</td>
<td>Non-students 36 appearances</td>
</tr>
</tbody>
</table>

- Summary of training data:

<table>
<thead>
<tr>
<th>Triskett-Bridge</th>
<th>West177-Tunnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 7 28 appearances</td>
<td>Student 7 19 appearances</td>
</tr>
<tr>
<td>Non-students 7 appearances</td>
<td>Non-students 11 appearances</td>
</tr>
</tbody>
</table>
Body Matching Re-Identification Pipeline

Target specification:

1. Manually outline target
2. Extract features for eBOLO signature
3. Begin scanning for target

Target re-identification:

1. Human detection and tracking
2. Candidate filtering
3. Feature extraction
4. Feature comparison
5. Threshold / rank order top matches
6. Present results to user

Manually outline target → Extract features for eBOLO signature → Begin scanning for target

Human detection and tracking → Candidate filtering → Feature extraction

Feature comparison → Threshold / rank order top matches → Present results to user
Human Detection and Tracking

- Foreground detection to get regions of interest; mixture of Gaussians
- Person detection based on improvements to aggregated channel features (Dollar et al. 2014) in each region of interest
- Algorithm is based on sliding windows, multi-scale feature pyramids, and boosted decision trees
Human Detection

- MOG Background Subtraction allows for reduced search space
- HOG and HAAR Cascades used to detect bodies within search space

Triskett Bridge busy scene (right) and MOG areas of foreground (left, in white)
Color Features

Each target and candidate is divided into several horizontal strips. Features are extracted from each strip based on the distribution of colors.

RGB

HSV

YCbCr
Texture Features

Texture features based on the patterns of intensity in each strip are also calculated.

We are currently investigating the combinations of color and texture feature that yield the best performance.
SURF Features

- Spectral analysis performed using SURF features on ETHZ dataset and annotated version of Triskett Bridge video (matching descriptor Cartesian distances (“Global Distance”))
  - Intra-Person Matching
  - Inter-Person Matching

Examples of Intra-Person (true) matching, and Inter-Person (false matching) on pedestrians
SURF Feature - Body Analysis

- Histograms of intra-person match distances (green) vs. inter-person match distances (blue).
We are investigating re-id methods including RankSVM (Prosser et al. 2010) and LFDA (Pedagadi 2013).

In each case, offline training is slow, but online comparison is instantaneous.
Metric Learning Re-Identification

$S$: Image pairs from the same person

$D$: Image pairs from different people

Classifier

Pairwise Constraints
Discriminative Content Using Adaptive Boosting

- Find localized patches of pixels that contain discriminative content for a specific target
- Color features are extracted within these confined patches of pixels
- Build the classifier to look for these features

‘Target’

‘Non-Targets’
Recognition Of Person-Based Attributes

• Identify the target based on its attributes - backpacks, hats, eyewear, ...

• The main attributes we are investigating:
  – carried objects such as backpacks and other types of luggage
  – irregularities in the silhouette of a person
Carried Objects

• We use prior information relative to the expected locations of carried objects
• Create templates for various poses

• Candidates are matched against pose templates
• Unmatched regions are expected to correspond to carried objects
Preliminary Results: RPI group

Station 1: Triskett Bridge
Input Video: tskt_bridge_1030a-1130a
Target Video: tskt_bridge_530p-630p

Station 2: West 117
Input Video: 117_tunnel_10a-11a
Target Video: 117_tunnel_5p-6p
Preliminary Results: RPI group

Ranked Gallery

Probe/Tagged Image
Preliminary Results: RPI group

Ranked Gallery

Probe/Tagged Image
## Summary of RPI preliminary results

<table>
<thead>
<tr>
<th>Input Video</th>
<th>Target Video</th>
<th>Person</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>tskt_bridge_1030a-1130a</td>
<td>tskt_bridge_530p-630p</td>
<td>Subject1</td>
<td>&gt;5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject6</td>
<td>3</td>
</tr>
</tbody>
</table>

% in rank 1: 50%
% in rank 5: 83%

<table>
<thead>
<tr>
<th>Input Video</th>
<th>Target Video</th>
<th>Person</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>117_tunnel_10a-11a</td>
<td>117_tunnel_5p-6p</td>
<td>Subject1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subject6</td>
<td>2</td>
</tr>
</tbody>
</table>

% in rank 1: 50%
% in rank 5: 100%
Preliminary results: NEU group

Basic information:
Station: West Park (highest FPS, ~17)
Video: w_pk_11a-12p; w_pk_530p-630p

Data:
Each person is represented by a video
Probe set: basic outfits
Gallery set: complement outfits + 25 randomly picked distractors

Features:
15 frame length dense trajectories + HOG3D; Fisher Vector Encoded

Metric learning:
Kernel Linear Fisher Discriminative Analysis (kLFDA)

Training data:
SAIVT Soft-bio dataset
Re-identification via dynamics

Key idea: use *dynamics* of subject’s motion for re-identification instead of their appearance

(Click images to play videos)
Preliminary results: NEU group

(Click images to play videos)

Robust to appearance change!
Preliminary results: NEU group

(Ranked Gallery)

(Click images to play videos)

Probe/Tagged Image

Robust to appearance change!
Preliminary results: NEU group

![Graph showing matching rate vs. rank with gallery size 38]

<table>
<thead>
<tr>
<th>Rank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28.6</td>
<td>42.9</td>
<td>57.1</td>
<td>71.4</td>
<td>100</td>
</tr>
</tbody>
</table>

Matching Rate (%)
Preliminary Results: Purdue Group

Target (Probe)  Ranked Gallery

Top matches: 1  2  3  4  5
Preliminary Results: Purdue Group

Target (Probe)  
Top matches: 1

Ranked Gallery

2  3  4  5
### Preliminary Results: Purdue Group

#### Triskett Bridge

<table>
<thead>
<tr>
<th>Rank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching Accuracy</td>
<td>53%</td>
<td>63%</td>
<td>77%</td>
<td>87%</td>
<td>93%</td>
<td>100%</td>
</tr>
</tbody>
</table>

#### West 177 - Tunnel

<table>
<thead>
<tr>
<th>Rank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching Accuracy</td>
<td>53%</td>
<td>77%</td>
<td>90%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Prototype Graphical User Interface

This slide is a playable video
Current and Future Work (Body Re-id)

- Hybrid generative/discriminative approach for fast, robust human detection and tracking

- Further investigation of effective features (e.g., visual saliency and subjective attributes)

- Per-camera tuning of feature extraction and re-id metric

- Boosting approach for offline training

- Subject-discriminative features and weights learned on-the-fly

- Intelligent selection of distinctive target and candidate snapshots for multi-shot re-identification
Face Recognition

Patrick J. Flynn, Kevin Bowyer
  • Joel Brogan, Amy Li, Nikita Amelchenko
Yun Fu
  • Ming Shao
Motivation

• Face detection in low-resolution images is challenging
• Detection becomes harder for both machine and human from resolutions of $24 \times 24$ to $6 \times 6$
• Results below are produced by combination of “Viola-Jones detector” + color heuristics

Resolution of faces

$24 \times 24$

$16 \times 16$

$8 \times 8$

$6 \times 6$
Motivation

• Poor lighting and low frame rate will fool detectors relying on skin color and face tracking
• In eBOLO dataset, lighting and frame rate are poor
• Commercial software is likely to fail
• To address these problems, we propose to:
  – Use body bounding box, and detect faces only on upper body
  – Empirically enlarge images (by image interpolation) to achieve better results
Face Detection: Notre Dame

- Custom face detector implemented based on literature
- 3 stage detector, robust to illumination changes
- To prevent false negatives, we allow false positives to contribute to the final face detection
- A density filter of face detection points allows us to robustly filter out the actual facial area from all TP and FP detections from a given pedestrian

Left: 10x10 Face. Right: 3 cascading classifier stages. White areas are pixels used for classification of a frame determined by a modified adaboost method
Preliminary Results: Notre Dame
Preliminary Results: Notre Dame
Preliminary Results: Notre Dame
Preliminary Results: Notre Dame

Video Demo
Face Detection: NEU Group

• Viola-Jones face detector
  – Haar feature + weak classifiers + cascade structure
• Implementation: OpenCV Library
  – Version 2.4
  – Requirement on minimum resolution: $20 \times 20$
• Randomly sample 100 images from
  – West 117 Tunnel 5p-6p
  – Triskett Bridge 1030a-1130a
Preliminary Results: NEU Group

- **Correct** detections in low-resolution images (after image interpolation)
Preliminary Results: NEU Group

- Incorrect detections in low-resolution images, but can be corrected by heuristics (face is on top of upper body)
Preliminary Results: NEU Group

- Detection results of correct/incorrect and missed detections
- Using interpolation to increase the image resolution up to 4 times the original
Face Recognition: Notre Dame

- Transforming a simple face space using Modified Census Transform with local colors allows for significant illumination/color invariance
- Due to Very Low Resolution (VLR) conditions, localized feature extraction becomes less practical
- Using training data, the face-space relationships between cameras can be calculated as homographies
  - Face data still being annotated
Face Recognition

• Face Hallucination
  – Constructing higher-resolution faces for better feature extraction and matching
  – Currently determining viability for the Very Low Resolution case
Video Summarization: Notre Dame

- Personnel: Nikita Amelchenko (CS undergraduate)
- Concept: remove video where nothing is happening
- Challenge: something is always happening
  - Noise
  - Timestamp/annotation
  - Other nonstationarities (sunlight)
- Approach
  - Baseline noise distribution
  - Masking
  - Thresholding difference
Baseline

Aggregate frame difference vs. Frame number
Motion
Strong motion
Annotation Tools: NEU Group

FDViewer: A Face Detector Interface

• **Single program** to detect and analyze faces.

• **Visualize** images with labels.

• **Well organized** using Corpus structures.
  – Handle image collections in a fluent fashion.

• **Quickly** browse results and record metrics.
  – Count FP, FN, Missed Detections w/out ground truth
  – Note that with ground truth manual counting is minimal
  – Annotations are stored as CSV files:
    
    \[
    <\text{image}_\text{name}, \text{numFN}, \text{numFN}, \text{numMissed}>\]

• **Easy** to create, save and load Corpus.
Annotation Tools

FDViewer: Video Demo
Annotation Tools

FDViewer: An Extendable (Modular) Framework

- **Easily attach** any object detector of choice.
  - **Change** detectors via dropdown menu.
  - **Analyze** and **Compare** different detectors.
Annotation Tools

Future Work

1. Enable Corpus merging.
   - Add user bookmarks.
2. Provide views (exemplar) filter.
   - View type specific results.
3. Allow images to be rescaled.
   - Reprocess scaled images.
4. Add ‘Delete’ to drop unwanted samples.
Future Work (face re-id)

- Fine tune Viola-Jones detector with appropriate image pre-processing
- Try different open source/commercial detectors (e.g., face plus plus)
- Compute quantitative results for detection, e.g., ROC, based on our face annotation tools (FDViewer)
- Investigate applicability of face recognition and video summarization to eBOLO application
- Integrate face algorithms with body-based algorithms to improve ranking of correct candidates
Automated Crowd Flow Estimation Enhanced by Crowdsourcing

Javier Ribera\textsuperscript{1}, Khalid Tahboub\textsuperscript{2} and Edward J. Delp\textsuperscript{2}

\textsuperscript{1}Universitat Politècnica de Catalunya, Barcelona, Spain
\textsuperscript{2}Purdue University, West Lafayette, IN, U.S.A.

Motivation

- Crowd flow is the process of estimating the number of people moving through a region in a given period of time.
- How robust is the automatic analysis of crowd flow?
- Can we identify uncertainties and utilize human intelligence to help machines perform better?
- How effective is crowdsourcing in enhancing the accuracy of crowd flow estimation?

Crowd Flow Estimation

- The number of people crossing the desired region of the image is proportional to the foreground pixels in the region.

\[ \frac{V_N}{S_N} = \frac{S_N}{C} \]

- The weighted accumulation of foreground pixels related to moving objects inside the region and during the period of N frames.
- The number of people that have crossed the region during a period of N frames.
- A scaling factor representing the number of foreground pixels per person

Estimation of Crowdedness

- Large crowd density results in a large number of occlusions which means a reduced number of foreground pixels per person.
- The level of crowdedness should be estimated in order to scale the foreground pixel count properly.
- Crowded scene presents a fine texture, while a sparse scene has a coarse texture.
- To estimate the level of crowdedness, we use a 16-D texture feature based on the Gray Level Co-occurrence Matrix (GLCM).
- Four GLCM matrices are created, each matrix represents a different direction for the adjacent pixels: right, top-right, top and top-left.
- For each matrix we extract four scalar features: energy, entropy, homogeneity and contrast to get a 16-D vector.

A region through which people passing are counted (Tripwire) and a Region of Interest surrounding it (ROI) are both selected.

Training process generates feature vectors to represent the texture of each level of crowdedness. It also determines the scaling factor for each level (number of foreground pixels per moving person).

Step 1: Extract the texture feature vector and classify it into one of the levels of crowdedness. Classification is based on the nearest reference vector.

Step 2: Find the count of foreground pixels related to moving objects and scale it properly using the scaling factor for this specific level of crowdedness.

Enhancement Using Crowdsourcing

- To enhance the classification performance, we use crowdsourcing to reduce uncertainty.
- The distance between a texture feature vector and the nearest reference vector, \( d_1 \), might be very comparable to the distance to the second nearest neighbor, \( d_2 \).
- We define “uncertainty” as the ratio \( d_1/d_2 \) to represent the uncertainty of the classification decision.
- We use crowdsourcing and ask members of the crowd to classify frames whenever “uncertainty” is greater than a predefined threshold:

\[ \mu = \frac{d_1}{d_2} > \alpha \]

- The new feature vectors are incorporated into the classifier to reduce future uncertainty

Robustness against Compression and Contrast Change

Degradation of video quality can significantly impact the automatic analysis.

Degradation of video quality is a result of compression and contrast change. H.264 compression was implemented using FFmpeg Constant Rate Factor (CRF), CRF was set to 33. Contrast was boosted to 1:68 using FFmpeg eq2 filter.

Results – Crowdsourcing Enhancement

This demonstrates that our proposed method manages to identify uncertainties and use the crowd in an effective way to lower the error rates.

Conclusion and Future Work

- We present enhancements to our previously developed crowd flow estimation method.
- Experimental evaluation demonstrates that our proposed method identifies uncertainties and use crowdsourcing in an effective way.
- Future work includes building a pyramid model to differentiate the accuracy of various crowd members and incorporating it in our method.

Publications


Acknowledgements

- This work is partially funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CN0001.
SUMMARY

Visual saliency is the perceptual process that makes attractive objects “stand out” from their surroundings in the low-level human visual system. A map of the “salient objects” or a saliency map is generated to indicate the locations of salient regions in a scene.

We describe four frequency domain visual saliency models based on new spectrum processing methods, i.e., the Gamma Corrected Spectrum (GCS) model, the Gamma Corrected Log Spectrum (GCLS) model, the Gaussian Filtered Spectrum (GFS) model, and the Gaussian Filtered Log Spectrum (GFLS) model.

A set of saliency map candidates are generated by inverse transform of a set of modified spectrums. An output saliency map is selected by minimizing the entropy among the set of saliency map candidates. Experimental results show that four extensions of our GCS, GCLS, GFS, and GFLS models are more accurate and efficient than some state-of-the-art saliency models in predicting eye fixation on standard image datasets.

Proposed Visual Saliency Models

- **Gamma Corrected Spectrum (GCS) Model**
  - Gamma correction is a nonlinear operation
  - A set of gamma corrections with different gamma values are utilized to modify the amplitude spectrum
  \[
  A_{GCS}(u, v, k) = (A(u, v))^\gamma_k
  \]
  \[
  S(x, y, k) = T^{-1}[(A(u, v))^\gamma_k \cdot \exp(i \cdot \mathcal{P}(u, v))]
  \]

- **Gamma Corrected Log Spectrum (GCLS) Model**
  - A set of gamma corrections with different gamma values are utilized to modify the log amplitude spectrum
  \[
  \mathcal{L}_{GCLS}(u, v, k) = (\mathcal{L}(u, v))^\gamma_k
  \]
  \[
  S(x, y, k) = T^{-1}[\exp((\mathcal{L}(u, v))^\gamma_k + i \cdot \mathcal{P}(u, v))]
  \]

- **Gaussian Filtered Spectrum (GFS) Model**
  - A set of Gaussian filters with various standard deviations are used to filter the amplitude spectrum
  \[
  A_{GFS}(u, v, k) = A(u, v) \ast GF(u, v, k)
  \]
  \[
  S(x, y, k) = T^{-1}[A_{GFS}(u, v, k) \cdot \exp(i \cdot \mathcal{P}(u, v))]
  \]

- **Gaussian Filtered Log Spectrum (GFLS) Model**
  - A set of Gaussian filters with various standard deviations are used to filter the log amplitude spectrum
  \[
  \mathcal{L}_{GFLS}(u, v, k) = \mathcal{L}(u, v) \ast GF(u, v, k)
  \]
  \[
  S(x, y, k) = T^{-1}\exp([\mathcal{L}_{GFLS}(u, v, k) + i \cdot \mathcal{P}(u, v))]
  \]

- **Naming Convention of The Extended Models (A-B-C)**
  - A represents a proposed model's abbreviation or an existing model's abbreviation
  - B represents a specific transform used for the model
  - C represents a particular color space used for the model

Eye Fixation Image Datasets

- Saliency models provide accurate prediction of human eye fixation on natural images in free viewing

Two standard image datasets

1. The Bruce and Tsotsos (BT) dataset [Bruce2009]
   - 120 color images with resolution of 681x511 pixels from indoor and outdoor scenes and the eye fixation data is based on 20 subjects

2. The Li et al.'s (Li) dataset [Li2013]
   - 235 color images with resolution of 640x480 pixels in six categories based on 19 subjects

EXPERIMENTAL RESULTS

- **Comprehensive Evaluation**: Comparing our 4 best extended saliency models with 10 existing saliency models using two standard image datasets
  - Our 4 best extended models: GCS-FT-Lab, GCLS-FT-Lab, GFS-FT-Lab, and GFLS-HFT-IRGBY
  - 10 existing models: SBVA [Itti1998], AIM [Bruce2009], FTSD [Achanta2009], PFDN [Bian2010], SR [Hou2007], PFT [Guo2008], PQFT [Guo2010], QDCT [Schauerte2012], ISDCT-Lab [Hou2012], SSA-HFT-IRGBY [Li2013]

- **Evaluation Measure**: Adopted the shuffled Area Under Curve (sAUC) score as the evaluation measure and developed an enhanced evaluation tool using the original MATLAB implementation of the existing models
  - Our 4 best extended models are better than some state-of-the-art models in predicting eye fixation

CONCLUSIONS

- We investigated low-complexity visual saliency models using spectral analysis approaches and proposed four visual saliency models based on spectrum processing methods
- Acknowledgement: This work was supported by the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) Center of the U.S. Department of Homeland Security (DHS) under Award Number 2009-ST-061-CIO00
Visual Saliency Models Based on Spectrum Processing

Bin Zhao and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA
Summary

• Investigated low-complexity bottom-up visual saliency using spectral analysis approaches
• Proposed 4 visual saliency models based on new spectrum processing methods
• A saliency map is selected by minimizing the entropy among a set of saliency map candidates
• Developed an evaluation tool for benchmarking visual saliency models
• Experimental results show that four extensions of our models are more accurate and efficient than previous saliency models in predicting eye fixation
Proposed Visual Saliency Models

- Propose new visual saliency models based on different spectrum processing methods in the frequency domain
  - Gamma Correction (non-linear operation)
    - Gamma Corrected Spectrum (GCS) Model
    - Gamma Corrected Log Spectrum (GCLS) Model
  - Gaussian Filtering (linear operation)
    - Gaussian Filtered Spectrum (GFS) Model
    - Gaussian Filtered Log Spectrum (GFLS) Model
Experiments

• Visual saliency models should be capable of providing accurate prediction of human eye fixation on natural images in free viewing.

• Use shuffled Area Under Curve (sAUC) as the fair evaluation measure to eliminate the center-bias effect in eye fixation data.

• Test the proposed and existing visual saliency models to predict eye fixation on two standard image datasets:
  1. The Bruce and Tsotsos dataset (BT)
     - 120 color images with resolution of 681x511 pixels
  2. The Li et al.'s dataset (Li)
     - 235 color images with resolution of 640x480 pixels
Comparison with Other Models

• A comprehensive evaluation of the 4 best extended saliency models by comparing with 10 other saliency models using two standard image datasets

  – SBVA(itti) [Itti1998], AIM [Bruce2009], FTSD [Achanta2009], PFDN [Bian2010], SR [Hou2007], PFT [Guo2008], PQFT [Guo2010], QDCT [Schauerte2012], IS-DCT-Lab [Hou2012], SSA-HFT-IRGBY [Li2013]

  – SBVA(itti), AIM, FTSD are spatial domain models

  – PFDN, SR, PFT, PQFT, IS-DCT-Lab, SSA-HFT-IRGBY and the 4 best extended models (GCS-FT-Lab, GCLS-FT-Lab, GFS-FT-Lab, and GFLS-HFT-IRGBY) are all frequency domain models
Comparison with Other Models

• The summary of each model
  – The rank in the same image dataset, the maximum sAUC score, and the associated Gaussian $\sigma_{opt}$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BT [6] Rank</td>
<td>13</td>
<td>8</td>
<td>14</td>
<td>7</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Max sAUC</td>
<td>0.6453</td>
<td>0.6956</td>
<td>0.5883</td>
<td>0.7015</td>
<td>0.6898</td>
<td>0.6898</td>
<td>0.6906</td>
<td>0.7041</td>
<td>0.7111</td>
<td>0.6932</td>
<td>0.7135</td>
<td>0.7138</td>
<td>0.7087</td>
<td>0.7030</td>
</tr>
<tr>
<td>$\sigma_{opt}$</td>
<td>0.030</td>
<td>0.030</td>
<td>0.040</td>
<td>0.045</td>
<td>0.040</td>
<td>0.040</td>
<td>0.040</td>
<td>0.040</td>
<td>0.040</td>
<td>0.040</td>
<td>0.045</td>
<td>0.040</td>
<td>0.045</td>
<td>0.040</td>
</tr>
<tr>
<td>Li [22] Rank</td>
<td>13</td>
<td>9</td>
<td>14</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Max sAUC</td>
<td>0.6525</td>
<td>0.6727</td>
<td>0.6186</td>
<td>0.6724</td>
<td>0.6649</td>
<td>0.6639</td>
<td>0.6742</td>
<td>0.6757</td>
<td>0.6758</td>
<td>0.6795</td>
<td>0.6805</td>
<td>0.6803</td>
<td>0.6780</td>
<td>0.6803</td>
</tr>
<tr>
<td>$\sigma_{opt}$</td>
<td>0.035</td>
<td>0.040</td>
<td>0.045</td>
<td>0.050</td>
<td>0.055</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.045</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
</tr>
</tbody>
</table>

• The 4 best extended models are generally better than previous models
  – The top 6 out of all the 14 models for the BT dataset
  – The top 5 out of all the 14 models for the Li dataset
• The best model: GCS-FT-Lab
Comparison with Other Models

The sAUC score of each model for BT (left) and Li (right) datasets
Comparison with Other Models

The average execution time of each model for BT (left) and Li (right) datasets
Beyond Big Data Analytics: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
Purdue University
ebertd@purdue.edu
What We Do: Visual Analytics

Visual Analytics\(^1\) is the science of **analytical reasoning** facilitated by **interactive visual interfaces**

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

---

1. **Illuminating the Path: The R&D Agenda for Visual Analytics**
Big Data

These forces fuel an explosion of data – a new economic asset forming the basis of opportunity.

- **In just two days**, we now generate as much data as was generated in total through 2003.
- **Over 1 billion tweets** are sent every 3 days.
- **80% of all data** is unstructured and growing at 15-times the rate of structured data.
- **5 million trade events** are clocked every day.

We are here.

---

- Sensors and devices
- Percentage of uncertain data
- Social media
- VoIP
- Enterprise data
Big Data: Solution to Global Challenges?

• Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables

• Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• Qualitative, fuzzy, and social data
  • Preferences, significance of one relationship over another; Trust

• Context
  • Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it
    into the collected data

• Spurious vs. Significant
  • Big data means more statistically significant events and correlations, but
    they may not have any relevance
  • Increases noise to signal ratio

• Big problems
  • Complex, multifaceted, multiparameter big challenges with unquantified
    dependencies
What Our Visual Analytic Solutions Offer

**We** enable users to be more effective through innovative interactive visualization, analysis, and decision making tools.

- Provide the *right information*, in the *right format* within the *right time* to solve the problem.
- Turn data deluge into a pool of relevant, actionable knowledge.
- Enable user to be more effective from planning to detection to response to recovery.
- Enable effective communication of information.

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Ebola Vs. Pandemic Influenza (H1N1)
Day 34 With Origin of Chicago, IL

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Scientific Discovery

• Flow dynamics visualization
  • Providing insights on large flow data
  • Visualization linked with simulations
  • Innovative feature visualization

• Nanohub
  • Information-assisted data analysis and visualization of nanoelectronic models

• Smart Agriculture
  • Predictive analytics and decision making for sustainability, resource management, product quality
Precision Agriculture VA
Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
- October 2011: Purdue, WL Police, Lafayette PD
VALET Layout

Calendar View

Time Series View

Time Slider

Prediction Panel

Map View

Clock View

Menus
Calendar Display

Selecting a date brings up more information

 Displays number of crimes that occurred on certain days
Crimes In A Single Day From 6PM – 9PM
Crimes During Certain Hours In A Month

All crimes in the month of September at 3PM
Select Specific Region To Display Crimes

All burglaries in the month of October in downtown Lafayette

Specific details on crime
Example: More Data Sources – Intoxication

- **PU vs. Notre Dame**
  - PU Lost: 10-38

- **Homecoming (Sat.)**
  - PU vs. Illinois
  - PU Won: 21-14

- **PU vs. Iowa**
  - PU Lost: 21-31

---

**Football season**

- Home
- Away

**Map of City Area**

**Day-of-the-Week**

- Mon, Tue, Wed, Thu, Fri, Sat, Sun

**Time of Day**

- 5pm, 6pm, 7pm, 8pm, 9pm, 10pm, 11pm, midnight, 1am, 2am, 3am, 4am, 5am, 6am

---

**PU vs. Other Teams**

- PU vs. Other
  - Won: 21-14
  - Lost: 10-38
  - Lost: 21-31
iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:

• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy:
SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper

• Used by U.S. Coast Guard
  • District 8 for events in 2014 - Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations

• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
• Ohio State Football – more successful than commercial tools
Social Media Analysis and Reporting Toolkit (SMART)
Marysville-Pilchuck High School (WA) Shooting
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14

• Prevented violence – the poor tuba player
Example From Ohio State Football 11/29/14
Example From Ohio State Football 11/29/14

Imagine a bone or something was popped out of his leg. Training staff was right away that he was going to have to be carted off.

Safety □ 31
Security □ 50
PUPD □ 11
Weather □ 5

OPD □ 67
 OPP □ 90
Example From Ohio State Football 11/29/14
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat


Follow Up Tweets

- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072
VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California San Diego
- University Washington
- Arizona State University
- Bethune-Cookman U.
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University

- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

• Federal Operating Components:
  • US Coast Guard
  • US Transportation Security Agency
  • US Citizenship and Immigration Service
  • US Federal Emergency Management Agency
  • US Customs and Border Protection
  • US CERT

• Law Enforcement
  • Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

• State Fusion Centers
  • Ohio (SAIC), Indiana (IIFC)

• Companies
  • Motorola, Kimberly Clark, Duke Energy, Bank of America
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

**IMPACT:**
- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region.
- Will allow the graffiti images to be “interpreted”
- More than 140 users and 1300 graffiti images acquired

**GARI delivered:**
- **Summer/Fall 2011:**
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- **Spring/Summer 2013:**
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

• **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

• **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

![Diagram of VASA Workbench with connections between USCC, Purdue, U Minn, and TACC nodes. The diagram illustrates the flow of information and processes related to critical infrastructure and supply chain management.]

February 2015
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• System of systems: binds together multiple simulations models from collaborators into coherent whole
  • Minnesota: food distribution model
  • Texas: simulated and historical weather (hurricanes, storms)
  • UNCC: critical infrastructure
  • Purdue: roads + interaction visual analytics tool

• Challenge:
  Combine interactive VA with complex simulation models
For Further Information

www.VisualAnalytics-CCI.org
Summary

• Powerful tool for daily use and during special events
• Realtime monitoring and email alerts
• More features than many tools on the market
• How to get SMART?
• Visit: www.vaccinetools.org
• Contact us: vaccine@purdue.edu
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Empowering the user, not replacing them

• *If you have a problem… we can develop a solution!*
Beyond Big Data Analytics: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
Purdue University
ebertd@purdue.edu
Overview

• VACCINE center
• Visual analytics
• Big data challenges
• Research approach
• Example projects
• Path forward
VACCINE - Who We Are: International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

• Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

• Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

• State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)

• Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America
Visual Analytics is the science of analytical reasoning facilitated by interactive visual interfaces.

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

1. Illuminating the Path: The R&D Agenda for Visual Analytics
Big Data

These forces fuel an explosion of data – a new economic asset forming the basis of opportunity.

**In just two days**
we now generate as much data as was generated in total through 2003.

**80% of all data**
is unstructured and growing at 15-times the rate of structured data.

**Over 1 billion tweets**
are sent every 3 days.

**5 million trade events**
are clocked every day.

We are here.

---

**Sensors and devices**

**Percentage of uncertain data**

**Social media**

**VoIP**

**Enterprise data**
Big Data: Solution to Global Challenges?

- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables

- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - Enable effective, efficient analysis, decision making, planning, and action
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• Qualitative, fuzzy, and social data
  • Preferences, significance of one relationship over another; Trust

• Context
  • Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it
    into the collected data

• Spurious vs. Significant
  • Big data means more statistically significant events and correlations, but
    they may not have any relevance
  • Increases noise to signal ratio

• Big problems
  • Complex, multifaceted, multiparameter big challenges with unquantified
    dependencies
Our Visual Analytics Process
(extended and adapted from Bertin, Normal, Tversky)

Process:
• Define problem / question
• Determine:
  • Workflow and interaction
  • Relevant data
  • Data characteristics (e.g., types, quality)
  • Abstraction/representation level
  • Natural scales
• Map to appropriate visual representation for task
• Utilize iterative refinement with user feedback from the start

Guiding Principles:
• Appropriateness principle
  • Neither more nor less
• Naturalness principle
  • Match with cognitive models
• Matching principle
  • Match the task, offer action affordances
• Principle of apprehension
  • Content should be accurately and easily perceived
What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools:

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Research topics

**Example application areas and projects**

- Path forward
Public Safety & Health

• Public safety visual analytics
  • cgSARVA
    • Coast Guard Search And Rescue Visual Analytics
  • MERGE
    • Mobile Emergency Response Guide

• Public health visual analytics
  • LAHVA
    • Linked Animal-Human Health Visual Analytics
  • RVF
    • Rift Valley Fever
    • Decision support environment for epidemic modeling and responses
  • PanVis
    • Pandemic influenza modeling and visualization tool
  • Cancer Care Engineering
Ebola Vs. Pandemic Influenza (H1N1)

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Scientific Discovery

- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization

- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models

- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality
Precision Agriculture VA
Scientific Discovery

- Illustrative Visualization
- Mechanical assemblies
- Schematics and wiring diagrams
- Illustrative flow visualization
- Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analytics
Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment
• Evaluate courses of action
• Evaluate above at both Strategic and Tactical/Operational level
Our Spatiotemporal Visual Analytics Work Contributions

- **Techniques:**
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - **Natural scale work**
  - Social medial visual analytics

- **Uses:**
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance
The Scale Problem

- **Data scale:**
  - Coarse scale analysis can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis may limit applicability because of fine resolution and can have data sparseness issues

- **User / problem scale:**
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally
Our Approach: Natural Scale Templates

- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of Naturalness\(^1\)
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- **Natural Scales:** Mappings that facilitate human perception to comprehend the data/task/problem
  - Can facilitate discovery from data
  - **Goal:** Empower the casual expert to be more effective

---

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

Past 2 Years

March + Tuesday Filter
(Past 10 Years)
Prediction Using KDE Distribution
Prediction During Certain Hours In The Day

9AM – 3PM

3PM – 7PM

9AM – 5PM (Day)

9PM – 3AM (Night)
Overview

- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system result
  - US Coast Guard
  - Social Media
- Path forward
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
- October 2011: Purdue, WL Police, Lafayette PD
VALET Layout

Calendar View

Time Slider

Prediction Panel

Time Series View

Map View

Clock View

Menus
Calendar Display

Selecting a date brings up more information

Displays number of crimes that occurred on certain days
Crimes In A Single Day From 6PM – 9PM
Crimes During Certain Hours In A Month

All crimes in the month of September at 3PM.
Select Specific Region To Display Crimes

All burglaries in the month of October in downtown Lafayette

Specific details on crime
Example: More Data Sources – Intoxication

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Football season

Home
Away

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week
iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant’s budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River:
- District 8, AIRSTA NEW ORLEANS
- District 9, AIRFAC WAUKEGAN
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

• **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

• **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: Combine interactive VA with complex simulation models
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu

December 2014
Why Use Social Media?

Marysville-Pilchuck High School (WA) Shooting
Time: around 10:40 am PST
(01:40 pm EST ), 10/24/2014
### Earliest Tweets About The Event

<table>
<thead>
<tr>
<th>Username</th>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>@_FatSackMatt</td>
<td>14-10-24 14:01:49 EST</td>
<td>cops are blocking all the entrances</td>
</tr>
<tr>
<td></td>
<td>14-10-24 13:55:45 EST</td>
<td>Cops are fucking EVERYWHERE right now</td>
</tr>
<tr>
<td></td>
<td>14-10-24 13:55:21 EST</td>
<td>Cops are fucking EVERYWHERE right now</td>
</tr>
<tr>
<td></td>
<td>14-10-24 13:55:12 EST</td>
<td>There is a shooting at mp?</td>
</tr>
<tr>
<td>@Miguel_Rocha34</td>
<td>14-10-24 13:54:06 EST</td>
<td>do you have any idea who the person with the gun was</td>
</tr>
<tr>
<td></td>
<td>14-10-24 13:52:12 EST</td>
<td>you got some serious issues that you need to seek help if you shoot a high school</td>
</tr>
<tr>
<td>@Ray_Raaaay21</td>
<td>14-10-24 13:49:11 EST</td>
<td>one of the kids in the lunchroom started shooting and everyone booked it</td>
</tr>
<tr>
<td></td>
<td>14-10-24 13:46:17 EST</td>
<td>shooting at mp</td>
</tr>
<tr>
<td>@_FatSackMatt</td>
<td>14-10-24 14:17:43 EST</td>
<td>2 people are dead</td>
</tr>
<tr>
<td></td>
<td>14-10-24 14:17:26 EST</td>
<td>Wtf is going on?? Shooting at mp?</td>
</tr>
<tr>
<td></td>
<td>14-10-24 14:14:39 EST</td>
<td>@Lake_stonerrrr a student shot several students, possible 2 students dead 6 injured</td>
</tr>
<tr>
<td></td>
<td>14-10-24 14:10:38 EST</td>
<td>Shooting at pikchuck??</td>
</tr>
<tr>
<td></td>
<td>14-10-24 14:08:26 EST</td>
<td>Something's fucking wrong !!!!! Or my mom wouldn't be crying about the MP shooting !!!</td>
</tr>
<tr>
<td></td>
<td>14-10-24 14:07:51 EST</td>
<td>About 5 mins away from the scene. Sad. Another school shooting. #MPHS</td>
</tr>
<tr>
<td></td>
<td>14-10-24 14:01:49 EST</td>
<td>@_FatSackMatt cops are blocking all the entrances</td>
</tr>
</tbody>
</table>

Some online news feed:
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 - Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools
Social Media Analysis and Reporting Toolkit (SMART)
Marysville-Pilchuck High School (WA) Shooting
December 2014

Location of the school
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston Marathon
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY 1 minute right after the incident
Anomaly Detection Superbowl Weekend
Examples From Fleet Week 10/12/14

Fleet Week Sunday 10/12/14

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More

12:56 PM - 12 Oct 2014
Ebola tweets during Fleet Week 10/12/14
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston Marathon bombing
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14

- Prevented violence – the poor tuba player
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00

Visualizations showing activity patterns before Sandy.
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
Keene, NH Pumpkin festival
10/18/14 – 10/19/14 Overview

Security classifier became prominent.

Major topics:
Keene, NH

Time range: 10/18 12:00pm – 10/19 06:00am, Security classifier)
The picture was later used by ABC news([link]).
Keene, NH
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Boston Marathon bombing
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:07:50 EDT
  EVERYONE SKIP SCHOOL TOMORROW RT “@TaZeNaToR1: IM LITERALLY GOING TO KILL SOMEONE”
  http://www.twitter.com/WeakndMLG

Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT
  Now the coast guard will be monitoring my tweets?
  http://www.twitter.com/ WeakndMLG

- WeakndMLG 10-17 17:08:07 EDT
  @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs
  http://www.twitter.com/ WeakndMLG
SMART: Network Linkage
Summary

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetools.org
- Contact us: vaccine@purdue.edu
Directions Forward, 
Keys to Success & Challenges
Seven Challenges for Proactive & Predictive Visual Analytics

1. Creating computer-human visual cognition environments
2. Integrating interactive simulation and analytics
3. Solving specific scale issues and cross-scale issues
4. Managing uncertainty and time
5. Enabling risk-based decision making environments
6. Developing the Science of Interaction for Visual Analytics
7. Defining appropriate evaluation methodologies for visual analytics
Visual Analytics: Remember…

• We need to be cognizant of parameters for visual representations
• Appropriate analysis can guide users to interesting features in the data
• Refined analysis through user interaction and their domain knowledge can help discover hidden problems
• There is no single catch-all visual representation or analysis
Keys for Success

• User and problem driven
• Balance human cognition and automated analysis and modeling
  • Often applied on-the-fly for specific components identified by the user
• Interactivity and easy interaction
  • Utilizing HPC and novel analysis approaches
• Understandability of why predicted value is what it is
• Intuitive visual cognition
• Not overloaded with features
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Empowering the user, not replacing them

• *If you have a problem... we can develop a solution!*
Big Data and Visual Analytics: Decision-Makers of the Digital Age

David S. Ebert
School of ECE
Purdue University
ebertd@purdue.edu
Visual Analytics

Visual Analytics\(^1\) is the science of analytical reasoning facilitated by interactive visual interfaces.

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

1. Illuminating the Path: The R&D Agenda for Visual Analytics
Big Data Challenges

DILBERT

CONSULTANTS SAY THREE QUINTILLION BYTES OF DATA ARE CREATED EVERY DAY.

IT COMES FROM EVERYWHERE. IT KNOWS ALL.

ACCORDING TO THE BOOK OF WIKIPEDIA, ITS NAME IS "BIG DATA."

BIG DATA LIVES IN THE CLOUD. IT KNOWS WHAT WE DO.

IN THE PAST, OUR COMPANY DID MANY EVIL THINGS.

BUT IF WE ACCEPT BIG DATA IN OUR SERVERS, WE WILL BE SAVED FROM BANKRUPTCY.

LET US PAY.

IS IT TOO LATE TO SIDE WITH EVIL?

SHHHH! IT HEARS YOU.

BY SCOTT ADAMS
Big Data

These forces fuel an explosion of data – a new economic asset forming the basis of opportunity

**In just two days**
we now generate as much data as was generated in total through 2003

**80% of all data**
is unstructured and growing at 15-times the rate of structured data

**Over 1 billion tweets**
are sent every 3 days

**5 million trade events**
are clocked every day

We are here

---

2013

2015
Big Data: Solution to Global Challenges?

- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables

- **Need:**
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• Qualitative, fuzzy, and social data
  • Preferences, significance of one relationship over another; Trust

• Context
  • Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it into the collected data

• Spurious vs. Significant
  • Big data means more statistically significant events and correlations, but they may not have any relevance
  • Increases noise to signal ratio

• Big problems
  • Complex, multifaceted, multiparameter big challenges with unquantified dependencies
VA Approaches to Big Data

• Don’t display all the data!

• Extract relevant information at the appropriate natural scales using analytics and statistical models to reduce data space so that it is mappable to the visual space

• Drive process with User interaction and expertise
What Our Visual Analytic Solutions Offer

*We* enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the **right information**, in the **right format** within the **right time** to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Research Area Overview and Example Projects

• Research applications areas
• VACCINE
• Example projects
  • Public safety and law enforcement
  • US Coast Guard
  • Social media visual analytics
Public Safety & Health

- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide

- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering
Ebola Vs. Pandemic Influenza (H1N1)

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Complex Decision Making: Advanced Decision Support Tools (e.g., Rift Valley Fever)

- Integrated simulation models
- Explore alternative courses of action in decision space and real space
Scientific Discovery

• Flow dynamics visualization
  • Providing insights on large flow data
  • Visualization linked with simulations
  • Innovative feature visualization

• Nanohub
  • Information-assisted data analysis and visualization of nanoelectronic models

• Smart Agriculture
  • Predictive analytics and decision making for sustainability, resource management, product quality
Precision Agriculture VA
Scientific Discovery

- Illustrative Visualization
- Mechanical assemblies
- Schematics and wiring diagrams
- Illustrative flow visualization
- Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analytics
Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation
Market Analyzer

Pixel-oriented Display Matrix

Geographical View

Proportional Legend View

Stacked Bar View & Time Sliders

Line Graph View
SemanticPrism: A Multi-aspect View of Large High-dimensional Data (Purdue University)

- VAST 2012 Mini Challenge 1 Award: Outstanding Integrated Analysis and Visualization
- Geo-Temporal
- Time-serial
- Pixel-based
- Semantic Zoom

Victor Yingjie Chen, Ahmad M Razip, Sungahn Ko, Cheryl Zhenyu Qian, David S.Ebert
VACCINE - Who We Are: International Team of Experts
75+ Faculty, 26 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

• Federal Operating Components:
  • US Coast Guard
  • US Transportation Security Agency
  • US Citizenship and Immigration Service
  • US Federal Emergency Management Agency
  • US Customs and Border Protection
  • US CERT

• Law Enforcement
  • Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

• State Fusion Centers
  • Ohio (SAIC), Indiana (IIFC)

• Companies
  • Motorola, Kimberly Clark, Banfield, Bank of America
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Visual Analytics Law Enforcement Toolkit (i)VALET
Example Projects: Public Safety and Law Enforcement Applications
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
• In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
• Mobile version being released to public (February 2013) for community-based policing
• Investigating correlation factors
• Analyzing time of day problems and improving accuracy of police record management system
• Novel statistical predictive model incorporated for planning
• Incorporating predictive alerts

VALET delivered:
• Spring 2011: WL, Lafayette Police
• Fall 2013: Ohio State Highway Patrol
• Spring 2014: NYPD
• Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
• October 2011: Purdue, WL Police
VALET Overview

Map View

Time Series View

Clock View

Twitter monitoring

Calendar View

Time Slider

Menus
VALET Demo
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• System of systems: binds together multiple simulations models from collaborators into coherent whole
  • Minnesota: food distribution model
  • Texas: simulated and historical weather (hurricanes, storms)
  • UNCC: critical infrastructure
  • Purdue: roads + interaction visual analytics tool

• Challenge:
  Combine interactive VA with complex simulation models
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)
cgSARVA Results – Aircraft coverage

Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.
Airstation Utility Visualization Comparison

• Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River:
- District 8, AIRSTA NEW ORLEANS
- District 9, AIRFAC WAUKEGAN
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

• Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
• Economic sector impact, local and national impact
• Impact and effectiveness of alternative mitigation strategies
Social Media: Real-time Visual Analytics
(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction (LDA, novel STL based estimation)
- Dynamically linked views
- Topics extracted shown as a dynamic word cloud
- Sources: Twitter, Facebook, Flickr
- Uses:
  - USCG Sector Ohio Valley
  - Indianapolis PD, Purdue PD
  - Fleet week Nov. 2014: USCG PAC

Warning: Many tweets can be explicit and offensive
Explosion Area in Boston

Keywords that have been used most often in the area

Tweets

October 2014
First Response (Tweet & Picture)

1 minute right after the incident
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 - Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
Social Media Analysis and Reporting Toolkit (SMART)

Hurricane Sandy
Anomaly Detection
Tweets about John Deere
Keywords (ContentLens) in the 3 Hotspots

Des Moines, Iowa  Davenport, Iowa  Monterrey, Mexico
Tweets related to ‘break down, shut down, stop working’ and ‘maintenance, repair’

@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking.

1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of.

<table>
<thead>
<tr>
<th>User ID</th>
<th>Creation Date</th>
<th>Tweets Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>630487309</td>
<td>13-06-14 15:55:47 EST</td>
<td>1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of</td>
</tr>
<tr>
<td>35458727</td>
<td>13-10-18 22:59:17 EST</td>
<td>@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking.</td>
</tr>
<tr>
<td>33628327</td>
<td>13-06-09 14:27:17 EST</td>
<td>And, for your viewing pleasure, a nicely maintained 1930s John Deere. <a href="http://t.co/Zavel5QnDi">http://t.co/Zavel5QnDi</a></td>
</tr>
<tr>
<td>50371494</td>
<td>13-06-07 10:13:50 EST</td>
<td>The Class of 2013 presents its class gift at commencement - a John Deere Gator for the Maintenance Department. <a href="http://t.co/4j1r1cFulm">http://t.co/4j1r1cFulm</a></td>
</tr>
<tr>
<td>494325743</td>
<td>13-08-16 15:07:21 EST</td>
<td>Rondorf yard maintenance! @chevy @johnndeere <a href="http://t.co/HyHyV5pTNR">http://t.co/HyHyV5pTNR</a></td>
</tr>
<tr>
<td>504120195</td>
<td>13-09-03 16:13:01 EST</td>
<td>Just saw a Clarkson maintenance worker ride a John Deere lawn tractor to the gas station. No need to bother with a gas can. #Townlife #labor #jobs Industrial Maintenance Technician: position with John Deere Water: Industrial Maintenance <a href="http://t.co/azE8mPmFk5">http://t.co/azE8mPmFk5</a></td>
</tr>
<tr>
<td>503857016</td>
<td>13-07-28 22:58:57 EST</td>
<td>That went easier than I expected. #easyout #johnndeere @ Mike S Lawn Mower Repair <a href="http://t.co/8pqlu01x7g">http://t.co/8pqlu01x7g</a></td>
</tr>
</tbody>
</table>
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Empowering the user, not replacing them

• *If you have a problem... we can develop a solution!*
For Further Information

www.VisualAnalytics-CCI.org
Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
Purdue University
ebertd@purdue.edu
Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Example projects
VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University

- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

• **Federal Operating Components:**
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

• **Law Enforcement**
  - Over 50 local and state agencies (IN, IL, OH, PA, NC, NY)

• **State Fusion Centers**
  - Ohio (SAIC), Indiana (IIFC)

• **Companies**
  - Motorola, Kimberly Clark, Duke Energy, Bank of America
Visual Analytics is the science of **analytical reasoning** facilitated by **interactive visual interfaces**.

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

---

1. *Illuminating the Path: The R&D Agenda for Visual Analytics*
Big Data: Solution to Global Challenges?

- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables

- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information  (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies
What Our Visual Analytic Solutions Offer

**We** enable users to be more effective through innovative interactive visualization, analysis, and decision making tools:

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Public Safety & Health

- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide

- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering
Ebola Vs. Pandemic Influenza (H1N1)
Day 34 With Origin of Chicago, IL

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation
Market Analyzer

Pixel-oriented Display Matrix

Geographical View

Proportional Legend View

Stacked Bar View & Time Sliders

January 2015

VACCINE
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:
• Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
• Economic sector impact, local and national impact
• Impact and effectiveness of alternative mitigation strategies
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
Our Spatiotemporal Visual Analytics Work
Contributions

• Techniques:
  • New anomaly and prediction techniques
  • New interactive correlative analysis
  • Natural scale work
  • Social medial visual analytics

• Uses:
  • Anomaly detection
  • Situational awareness
  • Predictive policing
  • Coast Guard visual analytics
    • Search and rescue
    • Risk-based resource allocation
    • Risk-based operational performance
Event/Anomaly Detection - STL

- Temporal trend analysis and prediction
- Seasonal Trend decomposition based on locally weighted regression (STL)
- Time series data viewed as the sum of multiple components: $Y = T + S + R$
  - $Y$: original time series
  - $T$: trend component
  - $S$: seasonal (daily/weekly)
  - $R$: remainder component
- Large value of $R$ indicates substantial variation
- Filter out the noise component
- If value varies by 2 SD from STL expectation, generate Alert
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Date:
- 1/1/2008
- 1/3/2008
- 1/5/2008
- 1/7/2008
- 1/9/2008
- 1/11/2008
- 1/13/2008

Respiratory Count:
- 0
- 10
- 20
- 30
- 40
- 50
- 60
Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space
Multivariate Network Analytics with Information Theory Anomaly Detection
The Scale Problem

- **Data scale:**
  - Coarse scale analysis can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis may limit applicability because of fine resolution and can have data sparseness issues

- **User / problem scale:**
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally
Our Approach: Natural Scale Templates

• Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
• Based on design principle of Naturalness¹
  • Selecting appropriate mappings natural to human cognition
• Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
• Natural Scales: Mappings that facilitate human perception to comprehend the data/task/problem
  • Can facilitate discovery from data
  • Goal: Empower the casual expert to be more effective

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

Past 2 Years

March + Tuesday Filter
(Past 10 Years)
Prediction During Certain Hours In The Day

9AM – 3PM

3PM – 7PM

9AM – 5PM (Day)

9PM – 3AM (Night)
Overview

- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system
  - US Coast Guard
  - Imaging tools
  - Social Media
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
- October 2011: Purdue, WL Police, Lafayette PD
VALET Layout

Calendar View

Time Slider

Prediction Panel

Map View

Time Series View

Clock View

Menus
Calendar Display

Selecting a date brings up more information

Displays number of crimes that occurred on certain days
All crimes in the month of September at 3PM
iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant’s budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River:
- District 8, AIRSTA NEW ORLEANS
- District 9, AIRFAC WAUKEGAN
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
Interactive Operational Performance Assessment Report (iOPAR)

- **OPAR**
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas

- **iOPAR** – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Analyze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA COMPONENT:
CRITICAL INFRASTRUCTURE
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• System of systems: binds together multiple simulations models from collaborators into coherent whole
  • Minnesota: food distribution model
  • Texas: simulated and historical weather (hurricanes, storms)
  • UNCC: critical infrastructure
  • Purdue: roads + interaction visual analytics tool

• Challenge:
  Combine interactive VA with complex simulation models
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

**IMPACT:**
- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region.
- Will allow the graffiti images to be “interpreted”
- More than 140 users and 1300 graffiti images acquired

**GARI delivered:**
- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper
• Used by U.S. Coast Guard
  • District 8 for events in 2014 - Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations
• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
• Ohio State Football – more successful than commercial tools
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY

1 minute right after the incident
Anomaly Detection Superbowl Weekend
#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

12:56 PM - 12 Oct 2014
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14

• Prevented violence – the poor tuba player
Example From Ohio State Football 11/29/14
Example From Ohio State Football 11/29/14

Imagine a bone or something was popped out of TT's leg. Training staff in the hallway said he was going to have to be carted off.
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
Keene, NH Pumpkin festival
10/18/14 – 10/19/14 Overview

Major topics:

Security classifier became prominent.
Keene, NH

Before they tear gased us...we out here, literally mobbing 🤣🤣🤣

Dominique Pascoal @DparOippa

Keene, NH
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat

• WeakndMLG 10-16 23:07:50 EDT EVERYONE SKIP SCHOOL TOMORROW RT “@TaZeNaToR1: IM LITERALLY GOING TO KILL SOMEONE” http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125203,-85.81704

Follow Up Tweets

• WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108

• WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072
Summary

• Powerful tool for daily use and during special events
• Realtime monitoring and email alerts
• More features than many tools on the market
• How to get SMART?
• Visit: www.vaccinetools.org
• Contact us: vaccine@purdue.edu
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Empowering the user, not replacing them

• *If you have a problem… we can develop a solution!*
For Further Information

www.VisualAnalytics-CCI.org
ORAM Visualization
Introduction

Risk-based Decision Making

Process that collects and organizes information about possible outcomes in an ordered structure that helps analysts make informed choices. [B. Macesker]

Performing risk analysis is a complex and challenging analytical task.

- Complex and large multivariate datasets.
- Clear understanding of the risk at hand.
- Cognitive limitations and information overload.
Related Work

Communication of Risk

- Static risk displays [Lipkus and Holland]
- Portfolio risk management [Gershonon and Eick]

Risk Analysis and Modeling

- Operational Risk Visualization [Bonafede and Marmo]
- Visualization Techniques to Present Risks and Options for Risk Mitigation [Feather et al.]
- Visual exploration and classification models for risk assessment [Migut and Worrning]
Problem Characterization: Risk Management Process

(a) Define the context
(b) Identify potential risks
(c) Assess and analyze
(d) Develop alternatives
(e) Decision-Making and Implementation
(f) Evaluate and Monitor

Main Framework of the Application

Historical Views, Operational Risk Profiles
Temporal and Spatial (heat maps, coverage maps)

Left to the analyst
Risk Terminology

• ORAM – Operational Risk Assessment Model
• 11 Missions in the Coast Guard – 10 Missions with risk values
• RIN – Risk index numbers.
• Total Risk – Mitigated Risk = Residual Risk.
Visual Analytics Environment

**Design Requirements:**

Interaction with the system.
Integration with previously used tool.
Creation of customizable scenarios.
Scenario

Questions that drive the planning strategy:

• What risks exist in the region and where they are distributed?
• Where are our resources allocated?
• What constraints exist in the system that will require a prioritization of resource use?
Identify and Analyze Potential Risk

General understanding on how the risk is distributed between the different areas and missions.

(A) District Choropleth Map for Total Risk across all missions

(B) Total Risk distribution by mission for any given district.

(C) Risk profiles across all districts and mission areas

(D) District X Risk Profiles across all mission area

(E) Total M10 Risk Values distributed among District X’s station AOR

(F) KDE heatmap showing Total M10 Risk values

Specific areas can be analyzed by time and the analyst can obtain details by demand.
Scenario

1. In a resource constrained environment, we want to use resources in the mission area that provides the greatest return on investment (large amount of total risk but very little residual risk).

- District Y has the largest total risk values, it mitigated most of the risk effectively.
- District X shows less total risk, but the amount of residual risk as well as residual to total risk ratio is the highest as encoded by the darkest red shade.
- District X can be seen as more problematic than District Y.
2. Determine the key drivers of risk within a district. What mission drives the risk?

3. Are there several big events that drive the risk, or are there many small events with smaller consequences accumulated to affect the operation?

4. Evaluation of current resource allocation.
Feedback

- The prototype components went through an iterative design refinement process with the collaboration of four Coast Guard personnel: an operation research analyst, a former Coast Guard officer, one in-field officer, and a high level officer.

- Overall they emphasized the system allowed them to answer questions derived from developing the planning strategy in a fast manner. It also allows them to observe the underlying factors that drive the risk profiles.
Summary

- The system presented assists in the cognitive processes of quantifying evidence and comparing lines of evidence, as well as adapting to new information.
- Case Study providing useful visual reference that can communicate recommendations based on risk management.
- The risk-based decision making process described serves as a blueprint for future systems dealing with risk values and resource planning.
Acknowledgments

• This work was partially funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0003. Jang’s work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A1011170).
• Thank you to Ben Maule and Joe DiRenzo from the US Coast Guard for their help.
Backup Slides
Definition of Risk

• “Potential for an unwanted outcome resulting from an incident, event, or occurrence, as determined by its likelihood and the associated consequences”

## Coast Guard Missions

<table>
<thead>
<tr>
<th>Coast Guard Mission</th>
<th>CG Mission (Short Name)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ports, Waterways, and Coastal Security</td>
<td>PWCS</td>
</tr>
<tr>
<td>Drug Interdiction</td>
<td>CD</td>
</tr>
<tr>
<td>Migrant Interdiction</td>
<td>AMIO</td>
</tr>
<tr>
<td>Living Marine Resources</td>
<td>LMR</td>
</tr>
<tr>
<td>Other Law Enforcement</td>
<td>OLE</td>
</tr>
<tr>
<td>Marine Safety</td>
<td>MARSAFE</td>
</tr>
<tr>
<td>Search and Rescue</td>
<td>SAR</td>
</tr>
<tr>
<td>Marine Environmental Protection</td>
<td>MEP</td>
</tr>
<tr>
<td>Defense Readiness</td>
<td>DR</td>
</tr>
<tr>
<td>Ice Operations</td>
<td>DOMICE</td>
</tr>
<tr>
<td>Aids to Navigation</td>
<td>ATON</td>
</tr>
</tbody>
</table>
Operational Risk Profiles

Compare risk values against districts.

Compare risk values against missions.

Iterative process with feedback from the Coast Guard.
Choropleth Maps
Heatmap for risk values
Total vs Residual comparison.
Coverage Maps for Performance Evaluation
Analyst’s Evaluation
Target to be re-identified selected

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

Find the Euclidean Distance & rank candidates

Human Detection (HoG)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

Moving Object Detection (GMM)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

The top 5 identified targets
EBOLO Analysis

Human Detection (HoG)

Moving Object Detection (GMM)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

VIPER Lab
Experimental Results

Target to be re-identified selected

Human Detection (HoG)

Moving Object Detection (GMM)

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

Find the Euclidean Distance & rank candidates

The top 5 identified targets
Experimental Results

Target to be re-identified selected

Find the Euclidean Distance & rank candidates

The top 5 identified targets

VIPER Lab
Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
Purdue University
ebertd@purdue.edu
Overview

• VACCINE center
• Visual analytics
• Big data challenges
• Example projects
VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions

• Purdue University
• Georgia Institute of Technology
• Pennsylvania State University
• Stanford University
• University of North Carolina at Charlotte
• University of California, San Diego
• University Washington
• Arizona State University
• Simon Fraser University
• University of British Columbia
• Ontario Institute of Technology
• Dalhousie University

• University of Houston, Downtown
• Virginia Tech
• Indiana University
• Florida International University
• University of Texas at Austin
• Morgan State University
• Navajo Technical College
• Oak Ridge National Laboratories
• University of Stuttgart
• University of Swansea
• Oxford University
• University of Calgary
• University of Manitoba
• Carleton University
• University of Victoria
Some of Our Engaged End-Users

• **Federal Operating Components:**
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

• **Law Enforcement**
  - Over 50 local and state agencies (IN, IL, OH, PA, NC, NY)

• **State Fusion Centers**
  - Ohio (SAIC), Indiana (IIFC)

• **Companies**
  - Motorola, Kimberly Clark, Duke Energy, Bank of America
What We Do: Visual Analytics

Visual Analytics\(^1\) is the science of \textit{analytical reasoning facilitated by interactive visual interfaces}.

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

1. Illuminating the Path: The R&D Agenda for Visual Analytics
What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools:

- Provide the right information, in the right format within the right time to solve the problem.
- Turn data deluge into a pool of relevant, actionable knowledge.
- Enable users to be more effective from planning to detection to response to recovery.
- Enable effective communication of information.

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Public Safety & Health

- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide

- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering
Ebola Vs. Pandemic Influenza (H1N1)
Day 34 With Origin of Chicago, IL

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level
Our Spatiotemporal Visual Analytics Work

Contributions

• Techniques:
  • New anomaly and prediction techniques
  • New interactive correlative analysis
  • Natural scale work
  • Social medial visual analytics

• Uses:
  • Anomaly detection
  • Situational awareness
  • Predictive policing
  • Coast Guard visual analytics
    • Search and rescue
    • Risk-based resource allocation
    • Risk-based operational performance
Event/Anomaly Detection - STL

- Temporal trend analysis and prediction
- Seasonal Trend decomposition based on locally weighted regression (STL)
- Time series data viewed as the sum of multiple components
  \[ Y = T + S + R \]
  
  \( Y \): original time series
  \( T \): trend component
  \( S \): seasonal (daily/weekly)
  \( R \): remainder component

- Large value of \( R \) indicates substantial variation
- Filter out the noise component
- If value varies by 2 SD from STL expectation, generate Alert

Graph showing seasonal, intraseasonal, and day of week components.
Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Graph shows respiratory count over dates from 1/1/2008 to 1/13/2008.
Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space
Multivariate Network Analytics with Information Theory Anomaly Detection
The Scale Problem

- **Data scale:**
  - Coarse scale analysis can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis may limit applicability because of fine resolution and can have data sparseness issues

- **User / problem scale:**
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally
Our Approach: Natural Scale Templates

- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of Naturalness
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- Natural Scales: Mappings that facilitate human perception to comprehend the data/task/problem
  - Can facilitate discovery from data
  - Goal: Empower the casual expert to be more effective

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

Past 2 Years

March + Tuesday Filter
(Past 10 Years)
Prediction During Certain Hours In The Day

- 9AM – 3PM
- 3PM – 7PM
- 9AM – 5PM (Day)
- 9PM – 3AM (Night)
Overview

- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system
  - US Coast Guard
  - Imaging tools
  - Social Media
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:

• In use to analyze crime patterns and to connect strings of activities (200+ downloads)
• Mobile version being released to public (Spring 2015) for community-based policing
• Investigating correlation factors
• Analyzing time of day problems and improving accuracy of police record management system
• Novel statistical predictive model incorporated for planning
• Incorporating predictive alerts

VALET delivered:
• Spring 2011: WL, Lafayette Police
• Fall 2013: Ohio State Highway Patrol
• Spring 2014: NYPD
• Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
• October 2011: Purdue, WL Police, Lafayette PD
Calendar Display

Selecting a date brings up more information

Displays number of crimes that occurred on certain days
Crimes During Certain Hours In A Month

All crimes in the month of September at 3PM
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis

July 2015
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy:
SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA COMPONENT: CRITICAL INFRASTRUCTURE

Weather proxy → Infrastructure proxy → Routing proxy → Supply chain proxy

USER
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**: Combine interactive VA with complex simulation models
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

IMPACT:
• Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
• Will allow the graffiti images to be “interpreted”
• More than 140 users and 1300 graffiti images acquired

GARI delivered:
• Summer/Fall 2011:
  • IPD gang detectives
  • IPD at large
  • Ind Fusion Gang Task Force
  • Gang detectives across Indiana
• Spring/Summer 2013:
  • Cook County, IL
  • INGang ownership
  • Tattoo database added
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 - Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools
Social Media Analysis and Reporting Toolkit (SMART)
Anomaly Detection Superbowl Weekend
Examples From Fleet Week 10/12/14

Fleet Week Sunday 10/12/14

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston Marathon bombing
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14

• Prevented violence – the poor tuba player
Example From Ohio State Football 11/29/14
Imagine a bone or something was popped out of TTB leg. Training staff knew right away that he was going to have to be carted off.
Example From Ohio State Football 11/29/14
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Boston Marathon bombing
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat


Follow Up Tweets

- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072
For Further Information

www.VisualAnalytics-CCI.org
Introduction

Risk-based Decision Making

Process that collects and organizes information about possible outcomes in an ordered structure that helps analysts make informed choices. [B. Macesker]

Performing risk analysis is a complex and challenging analytical task.
- Complex and large multivariate datasets.
- Clear understanding of the risk at hand.
- Cognitive limitations and information overload.
Related Work

Communication of Risk
- Static risk displays [Lipkus and Holland]
- Portfolio risk management [Gershon and Eick]

Risk Analysis and Modeling
- Operational Risk Visualization [Bonafede and Marmo]
- Visualization Techniques to Present Risks and Options for Risk Mitigation [Feather et al.]
- Visual exploration and classification models for risk assessment [Migut and Worring]
Problem Characterization: Risk Management Process

- (a) Define the context
- (b) Identify potential risks
- (c) Assess and analyze
- (d) Develop alternatives
- (e) Decision-Making and Implementation
- (f) Evaluate and Monitor

Main Framework of the Application

Historical Views, Operational Risk Profiles

Temporal and Spatial (heat maps, coverage maps)

Left to the analyst
Risk Terminology

• ORAM – Operational Risk Assessment Model
• 11 Missions in the Coast Guard – 10 Missions with risk values
• RIN – Risk index numbers.
• Total Risk – Mitigated Risk = Residual Risk.
Visual Analytics Environment

**Design Requirements:**

- Interaction with the system.
- Integration with previously used tool.
- Creation of customizable scenarios.
Scenario

Questions that drive the planning strategy:

• What risks exist in the region and where they are distributed?
• Where are our resources allocated?
• What constraints exist in the system that will require a prioritization of resource use?
General understanding on how the risk is distributed between the different areas and missions.
Scenario

1. In a resource constrained environment, we want to use resources in the mission area that provides the greatest return on investment (large amount of total risk but very little residual risk).

- District Y has the largest total risk values, it mitigated most of the risk effectively.
- District X shows less total risk, but the amount of residual risk as well as residual to total risk ratio is the highest as encoded by the darkest red shade.
- District X can be seen as more problematic than District Y.
Scenario

2. Determine the key drivers of risk within a district. What mission drives the risk?
3. Are there several big events that drive the risk, or are there many small events with smaller consequences accumulated to affect the operation?
4. Evaluation of current resource allocation.
Feedback

• The prototype components went through an iterative design refinement process with the collaboration of four Coast Guard personnel: an operation research analyst, a former Coast Guard officer, one in-field officer, and a high level officer.

• Overall they emphasized the system allowed them to answer questions derived from developing the planning strategy in a fast manner. It also allows them to observe the underlying factors that drive the risk profiles.
Summary

• The system presented assists in the cognitive processes of quantifying evidence and comparing lines of evidence, as well as adapting to new information.

• Case Study providing useful visual reference that can communicate recommendations based on risk management.

• The risk-based decision making process described serves as a blueprint for future systems dealing with risk values and resource planning.
Acknowledgments

• This work was partially funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0003. Jang’s work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A1011170).
• Thank you to Ben Maule and Joe DiRenzo from the US Coast Guard for their help.
Backup Slides
Definition of Risk

• “Potential for an unwanted outcome resulting from an incident, event, or occurrence, as determined by its likelihood and the associated consequences”

## Coast Guard Missions

<table>
<thead>
<tr>
<th>Coast Guard Mission</th>
<th>CG Mission (Short Name)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ports, Waterways, and Coastal Security</td>
<td>PWCS</td>
</tr>
<tr>
<td>Drug Interdiction</td>
<td>CD</td>
</tr>
<tr>
<td>Migrant Interdiction</td>
<td>AMIO</td>
</tr>
<tr>
<td>Living Marine Resources</td>
<td>LMR</td>
</tr>
<tr>
<td>Other Law Enforcement</td>
<td>OLE</td>
</tr>
<tr>
<td>Marine Safety</td>
<td>MARSAFE</td>
</tr>
<tr>
<td>Search and Rescue</td>
<td>SAR</td>
</tr>
<tr>
<td>Marine Environmental Protection</td>
<td>MEP</td>
</tr>
<tr>
<td>Defense Readiness</td>
<td>DR</td>
</tr>
<tr>
<td>Ice Operations</td>
<td>DOMICE</td>
</tr>
<tr>
<td>Aids to Navigation</td>
<td>ATON</td>
</tr>
</tbody>
</table>
Operational Risk Profiles

Compare risk values against districts.

Compare risk values against missions.

Iterative process with feedback from the Coast Guard.
Choropleth Maps
Heatmap for risk values

Total vs Residual comparison.
Coverage Maps for Performance Evaluation
Analyst’s Evaluation
Target to be re-identified selected

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

Find the Euclidean Distance & rank candidates

Human Detection (HoG)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

Find the Euclidean Distance & rank candidates

Moving Object Detection (GMM)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

Find the Euclidean Distance & rank candidates

The top 5 identified targets
EBOLO Analysis

Human Detection (HoG)

Moving Object Detection (GMM)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

VIPER Lab
Experimental Results

Target to be re-identified selected

Human Detection (HoG)

Moving Object Detection (GMM)

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

Find the Euclidean Distance & rank candidates

The top 5 identified targets
Experimental Results

Target to be re-identified selected

Find the Euclidean Distance & rank candidates

The top 5 identified targets

VIPER Lab
Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
Purdue University
ebertd@purdue.edu

CVADA
Center for Visual Analytics and Data Analytics
A U.S. Department of Homeland Security Center of Excellence
December 2014
Overview

• VACCINE center
• Visual analytics
• Big data challenges
• Example projects
VACCINE - Who We Are: International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University

- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

• Federal Operating Components:
  • US Coast Guard
  • US Transportation Security Agency
  • US Citizenship and Immigration Service
  • US Federal Emergency Management Agency
  • US Customs and Border Protection
  • US CERT

• Law Enforcement
  • Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

• State Fusion Centers
  • Ohio (SAIC), Indiana (IIFC)

• Companies
  • Motorola, Kimberly Clark, Duke Energy, Bank of America
What We Do: Visual Analytics

Visual Analytics\(^1\) is the science of **analytical reasoning** facilitated by **interactive visual interfaces**

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

---

1. Illuminating the Path: The R&D Agenda for Visual Analytics
Big Data

These forces fuel an explosion of data – a new economic asset forming the basis of opportunity.

**In just two days**
we now generate as much data as was generated in total through 2003

**Over 1 billion tweets**
are sent every 3 days

80% of all data is unstructured and growing at 15-times the rate of structured data

5 million trade events are clocked every day

We are here

2013

2015
Big Data: Solution to Global Challenges?

- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables

- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information  
    (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust

- Context
  - Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it into the collected data

- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio

- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies
What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Overview

• VACCINE center
• Visual analytics
• Big data challenges
• Research topics

• Example application areas and projects
Public Safety & Health

- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide

- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering
Ebola Vs. Pandemic Influenza (H1N1)

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Business Visual Analytics

• Risk-based decision making and resource allocation
  • Coast guard operational risk assessment model
  • Helping to prioritize efforts to minimize risk

• Competitive Intelligence
  • Visual analytics system for business intelligence

• EconVIS
  • Visual analytics in various economic problems
  • Improving decision making and identifying key motivations in knowledge creation
Market Analyzer

Pixel-oriented Display Matrix

Geographical View

Proportional Legend View

Stacked Bar View & Time Sliders
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment
• Evaluate courses of action
• Evaluate above at both Strategic and Tactical/Operational level
Our Spatiotemporal Visual Analytics Work Contributions

**Techniques:**
- New anomaly and prediction techniques
- New interactive correlative analysis
- **Natural scale work**
- Social medial visual analytics

**Uses:**
- Anomaly detection
- Situational awareness
- Predictive policing
- Coast Guard visual analytics
  - Search and rescue
  - Risk-based resource allocation
  - Risk-based operational performance
Event/Anomaly Detection - STL

- Temporal trend analysis and prediction
  - Seasonal Trend decomposition based on locally weighted regression (STL)
  - Time series data viewed as the sum of multiple components:
    \[ Y = T + S + R \]
    - \( Y \): original time series
    - \( T \): trend component
    - \( S \): seasonal (daily/weekly)
    - \( R \): remainder component
- Large value of \( R \) indicates substantial variation
- Filter out the noise component
- If value varies by 2 SD from STL expectation, generate alert
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Date

Respiratory Count
Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space
Multivariate Network Analytics with Information Theory Anomaly Detection
The Scale Problem

• **Data scale:**
  • *Coarse scale analysis* can have too much variance or contain combinations of multiple signals reducing effectiveness
  • *Fine scale analysis* may limit applicability because of fine resolution and can have data sparseness issues

• **User / problem scale:**
  • Users perform analysis at scales that are natural to them and the task/problem
  • Characteristics might not be well articulated or available digitally
Our Approach: Natural Scale Templates

- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of **Naturalness**¹
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- **Natural Scales**: Mappings that facilitate human perception to comprehend the data/task/problem
  - Can facilitate discovery from data
  - Goal: **Empower the casual expert to be more effective**

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

Past 2 Years

March + Tuesday Filter
(Past 10 Years)
Prediction Using KDE Distribution
Prediction During Certain Hours In The Day

9AM – 3PM

3PM – 7PM

9AM – 5PM (Day)

9PM – 3AM (Night)
Overview

• VACCINE center
• Big Data challenges
• Visual analytics
• Research topics
• Example projects
  • VALET system
  • US Coast Guard
  • Imaging tools
  • Social Media
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
• In use to analyze crime patterns and to connect strings of activities (200+ downloads)
• Mobile version being released to public (Spring 2015) for community-based policing
• Investigating correlation factors
• Analyzing time of day problems and improving accuracy of police record management system
• Novel statistical predictive model incorporated for planning
• Incorporating predictive alerts

VALET delivered:
• Spring 2011: WL, Lafayette Police
• Fall 2013: Ohio State Highway Patrol
• Spring 2014: NYPD
• Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
• October 2011: Purdue, WL Police, Lafayette PD
VALET Layout

- Map View
- Time Series View
- Calendar View
- Prediction Panel
- Time Slider Clock View
- Menus

The image shows a graphical interface with various views and panels labeled accordingly.
Calendar Display

Displays number of crimes that occurred on certain days

Selecting a date brings up more information
Crimes In A Single Day From 6PM – 9PM
Crimes During Certain Hours In A Month

All crimes in the month of September at 3PM
Select Specific Region To Display Crimes

All burglaries in the month of October in downtown Lafayette
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis

December 2014
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River:
- District 9, AIRFAC WAUKEGAN
- District 8, AIRSTA NEW ORLEANS
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:
• Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
• Economic sector impact, local and national impact
• Impact and effectiveness of alternative mitigation strategies
Interactive Operational Performance Assessment Report (iOPAR)

• OPAR
  • Standard report within the Coast Guard Business Intelligence (CGBI) system
  • Displays resource use and performance by core CG mission areas

• iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  • Analyze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

• Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

• End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

---

**Diagram:**
- **VASA Workbench** connected to:
  - **S1: Disasters**
  - **S2: Critical infrastructure**
  - **S3: Supply Chain & Re-routing**

---

[Diagram of VASA Workbench connections to S1, S2, and S3 nodes]
VASA COMPONENT: CRITICAL INFRASTRUCTURE

- Weather
- Critical Infrastructure
- Routing
- Supply Chain

Network connections between components:

- Weather proxy
- Infrastructure proxy
- Routing proxy
- Supply chain proxy

User interaction point:

USER
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**: Combine interactive VA with complex simulation models
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

**IMPACT:**
- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be “interpreted”
- More than 140 users and 1300 graffiti images acquired

**GARI delivered:**
- **Summer/Fall 2011:**
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- **Spring/Summer 2013:**
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
Camera/Image Forensics

← Original “Girls”

Altered “Girls” →
**EBOLO**

- Target to be re-identified selected
- Candidate (resize, extract horizontal strip (8))
- Extract 8 features / strip (4 HSV & 4 texture) 64-D vector
- Find the Euclidean Distance & rank candidates
- Human Detection (HoG)
- Moving Object Detection (GMM)
  - Candidate (resize, extract horizontal strip (8))
  - Extract 8 features / strip (4 HSV & 4 texture)
  - Extract 8 features / strip (4 HSV & 4 texture) 64-D vector
- The top 5 identified targets
EBOLO Analysis

- Human Detection (HoG)
- Moving Object Detection (GMM)
  - Candidate (resize, extract horizontal strip (8))
  - Extract 8 features / strip (4 HSV & 4 texture)
  - Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

VIPER Lab
Experimental Results

- Target to be re-identified selected
- Human Detection (HoG)
- Moving Object Detection (GMM)

Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

Find the Euclidean Distance & rank candidates

1  2  3  4  5

The top 5 identified targets

VIPER Lab
Experimental Results

Target to be re-identified selected

Find the Euclidean Distance & rank candidates

The top 5 identified targets

VIPER Lab
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper
• Used by U.S. Coast Guard
  • District 8 for events in 2014 - Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations
• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
• Ohio State Football – more successful than commercial tools
Location of the school
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston Marathon
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY

1 minute right after the incident
Anomaly Detection Superbowl Weekend
#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More

12:56 PM - 12 Oct 2014
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14

- Prevented violence – the poor tuba player
Example From Ohio State Football 11/29/14
Example From Ohio State Football 11/29/14
Example From Ohio State Football 11/29/14
Where Can It Be Helpful?

During large-scale planned events, e.g.,
- Boston Marathon bombing
- Superbowl 2014 in NJ
- Riverboat Festival in Louisville October 2014
- Fleet Week 2014 in San Francisco

During recurring events
- Purdue football games – the poor tuba player

During unexpected events
- Pensacola tornado
- Keene, NH Pumpkin Festival

Detecting daily issues
- Teen threatening violence – Louisville, KY
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00

Supermarket  Park  Shelter
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
Keene, NH Pumpkin festival
10/18/14 – 10/19/14 Overview

Major topics:

Security classifier became prominent.
Keene, NH
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Boston Marathon bombing
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat

• WeakndMLG 10-16 23:07:50 EDT EVERYONE SKIP SCHOOL TOMORROW RT “@TaZeNaToR1: IM LITERALLY GOING TO KILL SOMEONE”
http://www.twitter.com/WeakndMLG
https://www.google.com/maps/place/38.125203,-85.81704

Follow Up Tweets

• WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets?
http://www.twitter.com/WeakndMLG
https://www.google.com/maps/place/38.125221,-85.817108

• WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs
http://www.twitter.com/WeakndMLG
https://www.google.com/maps/place/38.125206,-85.817072
Summary

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetoools.org
- Contact us: vaccine@purdue.edu
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Empowering the user, not replacing them

• *If you have a problem… we can develop a solution!*
For Further Information

www.VisualAnalytics-CCI.org
VACCINE: Advanced Interactive Visual Analytic Environments

David Ebert
Purdue University
VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

- **Federal Operating Components:**
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

- **Law Enforcement**
  - Over 50 local and state agencies (IN, IL, OH, PA, NC, NY)

- **State Fusion Centers**
  - Ohio (SAIC), Indiana (IIFC)

- **Companies**
  - Motorola, Kimberly Clark, Duke Energy, Bank of America
Big Data: Solution to Global Challenges?

• Big Data is often defined along three dimensions:
  • **Volume** – size
  • **Velocity** – rate of input, update, change
  • **Variety** – different types, sources, variables

• Need:
  • Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  • **Enable effective, efficient analysis, decision making, planning, and action**
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it
    into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but
    they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified
    dependencies
What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools:

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Scientific Discovery

- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations

- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models

- Cancer biomarker
  - Visual analysis suite for exploring multiple samples of data from cancer care engineering
Scientific Discovery

- Illustrative Visualization
- Mechanical assemblies
- Schematics and wiring diagrams
- Illustrative flow visualization
- Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analytics
Flow Visual Analytics

- Inspired by work of Leonardo da Vinci
- Allow understanding of flow processes and effects
- Enable interactive analysis of massive unstructured data through functional encoding, analysis and rendering
- Enable Interactive correlative feature analysis
Flow Analytics

- Inspired by work of Leonardo da Vinci
- Main feature conveyed with a few simple strokes
- Easy to understand 3D turbulent flow
Illustrative Flow Analytics

- Two examples
Flow - Photographic Inspired Techniques
Flow Analytic Environment
Volume Illustration: Overview

- Utilize illustration principles and techniques
Interactive Volume Illustration Motifs: Anatomical Illustration
Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk

- Competitive Intelligence
  - Visual analytics system for business intelligence

- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation
Market Analyzer

Pixel-oriented Display Matrix

Geographical View

Proportional Legend View

Stacked Bar View & Time Sliders

Line Graph View
Public Safety & Health

- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide

- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool

- Cancer Care Engineering
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA COMPONENT:
CRITICAL INFRASTRUCTURE

Weather

Critical Infrastructure

Routing

Supply Chain

Weather proxy

Infrastructure proxy

Routing proxy

Supply chain proxy

USER
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: Combine interactive VA with complex simulation models
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Our Spatiotemporal Visual Analytics Work

Contributions

- **Techniques:**
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - **Natural scale work**
  - Social medial visual analytics

- **Uses:**
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance
Event/Anomaly Detection - STL

- Temporal trend analysis and prediction
- Seasonal Trend decomposition based on locally weighted regression (STL)
- Time series data viewed as the sum:
  \[ Y = T + S + R \]

  \( Y \) : original time series
  \( T \) : trend component
  \( S \) : seasonal (daily/weekly)
  \( R \) : remainder component

- Large value of \( R \) indicates substantial variation
- Filter out the noise component
- If value varies by 2 SD from STL expectation, generate Alert
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
- October 2011: Purdue, WL Police, Lafayette PD
VALET Layout

- Calendar View
- Time Series View
- Prediction Panel
- Map View
- Time Slider
- Clock View
- Menus
Calendar Display

Selecting a date brings up more information

Displays number of crimes that occurred on certain days
Crimes During Certain Hours In A Month

All crimes in the month of September at 3PM
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment
• Evaluate courses of action
• Evaluate above at both Strategic and Tactical/Operational level
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant’s budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper
• Used by U.S. Coast Guard
  • District 8 for events in 2014 - Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations
• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
• Ohio State Football – more successful than commercial tools
Social Media Analysis and Reporting Toolkit (SMART)
Where Can It Be Helpful?

- Understanding Customers
- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY 1 minute right after the incident
Anomaly Detection Superbowl Weekend
Examples From Fleet Week

Fleet Week Sunday 10/12/14

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

@TWilsonTV
San Francisco, CA

Reply Retweet Favorite More

12:56 PM - 12 Oct 2014
Where Can It Be Helpful?

• **During large-scale planned events, e.g.,**
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

• **During recurring events**
  - Purdue football games – the poor tuba player

• **During unexpected events**
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

• **Detecting daily issues**
  - Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Boston Marathon bombing
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:07:50 EDT
  "EVERYONE SKIP SCHOOL TOMORROW RT "@TaZeNaToR1: IM LITERALLY GOING TO KILL SOMEONE"
  http://www.twitter.com/WeakndMLG
  https://www.google.com/maps/place/38.125203,-85.81704

Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT
  "Now the coast guard will be monitoring my tweets?"
  http://www.twitter.com/WeakndMLG
  https://www.google.com/maps/place/38.125221,-85.817108

- WeakndMLG 10-17 17:08:07 EDT
  "@TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs"
  http://www.twitter.com/WeakndMLG
  https://www.google.com/maps/place/38.125206,-85.817072
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Empowering the user, not replacing them

• *If you have a problem… we can develop a solution!*
For Further Information

www.VisualAnalytics-CCI.org
VACCINE: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
Purdue University
ebertd@purdue.edu
VACCINE - Who We Are: International Team of Experts
75+ Faculty, 29 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California San Diego
- University of Maryland
- University Washington
- Arizona State University
- Bethune-Cookman U.
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
What We Do: Visual Analytics

Visual Analytics\textsuperscript{1} is the science of \textit{analytical reasoning} facilitated by \textit{interactive visual interfaces}

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

\textsuperscript{1}. Illuminating the Path: The R&D Agenda for Visual Analytics
What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools.

• Provide the right information, in the right format within the right time to solve the problem.
• Turn data deluge into a pool of relevant, actionable knowledge.
• Enable user to be more effective from planning to detection to response to recovery.
• Enable effective communication of information.

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Ebola Vs. Pandemic Influenza (H1N1)
Day 34 With Origin of Chicago, IL

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Scientific Discovery

- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization

- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models

- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality
Precision Agriculture VA
Business Visual Analytics

• Risk-based decision making and resource allocation
  • Coast guard operational risk assessment model
  • Helping to prioritize efforts to minimize risk

• Competitive Intelligence
  • Visual analytics system for business intelligence

• EconVIS
  • Visual analytics in various economic problems
  • Improving decision making and identifying key motivations in knowledge creation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment
• Evaluate courses of action
• Evaluate above at both Strategic and Tactical/Operational level
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: Combine interactive VA with complex simulation models
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
- October 2011: Purdue, WL Police, Lafayette PD
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper

• Used by U.S. Coast Guard
  • District 8 for events in 2014 - Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations

• Purdue Police for home football games Fall 2014

• Indianapolis Metropolitan Police for special events

• US CBP for investigations Fall 2014 (Boston, AMOC)

• Ohio State Football – more successful than commercial tools
Social Media Analysis and Reporting Toolkit (SMART)
Marysville-Pilchuck High School (WA) Shooting
Location of the school
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston Marathon bombing
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14

• Prevented violence – the poor tuba player
Example From Ohio State Football 11/29/14
Example From Ohio State Football 11/29/14
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Boston Marathon bombing
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:07:50 EDT
  EVERYONE SKIP SCHOOL TOMORROW RT “@TaZeNaToR1: IM LITERALLY GOING TO KILL SOMEONE”  http://www.twitter.com/WeakndMLG
  https://www.google.com/maps/place/38.125203,-85.81704

Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT
  Now the coast guard will be monitoring my tweets?  http://www.twitter.com/WeakndMLG
  https://www.google.com/maps/place/38.125221,-85.817108

- WeakndMLG 10-17 17:08:07 EDT
  @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs  http://www.twitter.com/WeakndMLG
  https://www.google.com/maps/place/38.125206,-85.817072
What can we learn from this?

- Tracking Competitors
- Customer Satisfaction
- Employee Satisfaction
- Early Warning System
John Deere
Ten Dominant Topics Related to John Deere
Keywords (Contentlens) in the 3 Hotspots

Des Moines, Iowa  
Davenport, Iowa  
Monterrey, Mexico
Hashtags Related to John Deere

• #johndeere, #nothingrunslikeadeere
• #tractor, #tractortuesday
• #farm, #farming, #farmlife
• #jobs, #work, #harvest
• #green, #country, #summer
• #diesel, #dieselgenerator, #jasonaldean
• #genset, #wayouthere
Con Expo
Tweets related to John Deere skidder

Buster Beans @cinthfbbf - 22h
Just got a new John Deere skidder today pic.twitter.com/c8WiMnXVTV

John Deere @JohnDeere - Apr 24
If a lion is the king of the jungle, a John Deere Skidder is the king of the forest. pic.twitter.com/afZ76bXVV

Service Manuals @EbookManuals - Apr 4
#download John Deere 848G Skidder Technical Manual Download TM12424-1 pic.twitter.com/5KChgF

J Lowe @jennyfaithl - Apr 22
Firestone 16 ply Skidder Tires on john deere wheels 7/8ths holes #ebaymobile

Want Ad Digest @Wantaddigest - Apr 16
John Deere 540 B Cable Skidder LOWVILLE NY: 23x26 tire ring chain, diamond chains, runs very good, needs l... bit.ly/1ePVzwM

Burge Mcmanus @manualdownloads - Apr 13
Tweets related to ‘maintain, maintenance, repair’

<table>
<thead>
<tr>
<th>User ID</th>
<th>Creation Date</th>
<th>Tweets Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>933628327</td>
<td>13-06-09 14:27:17 EST</td>
<td>And, for your viewing pleasure, a nicely maintained 1930s John Deere. <a href="http://t.co/ZaveI5QnDi">Link</a></td>
</tr>
<tr>
<td>50371494</td>
<td>13-06-07 10:13:50 EST</td>
<td>The Class of 2013 presents its class gift at commencement – a John Deere Gator for the Maintenance Department. <a href="http://t.co/iju1rbc5U4">Link</a></td>
</tr>
<tr>
<td>194325743</td>
<td>13-08-16 15:07:21 EST</td>
<td>Rondorf yard maintenance! #chevy #johndeere <a href="http://t.co/Hyhly5TNR">Link</a></td>
</tr>
<tr>
<td>404120195</td>
<td>13-09-03 16:13:01 EST</td>
<td>Just saw a Clarkson maintenance worker ride a John Deere lawn tractor to the gas station. No need to bother with a gas can. #TownieLife</td>
</tr>
<tr>
<td>1038577016</td>
<td>13-07-28 22:38:57 EST</td>
<td>#labor #jobs Industrial Maintenance Technician: position with John Deere Water: Industrial Maintenance... <a href="http://t.co/aZEBmrPMtE">Link</a></td>
</tr>
<tr>
<td>502127999</td>
<td>13-08-07 18:47:22 EST</td>
<td>That went easier than I expected. #easyout #johndeere @ Mike s Lawn Mower Repair <a href="http://t.co/8pgJu0Uv9g">Link</a></td>
</tr>
</tbody>
</table>
'Maintain'

And, for your viewing pleasure, a nicely maintained 1930s John Deere. pic.twitter.com/Zavel5QnDI

The Class of 2013 presents its class gift at commencement - a John Deere Gator for the Maintenance Department. pic.twitter.com/ju1rchEuLm
Tweets related to ‘break down, shut down, stop working’

@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking.

1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of.

<table>
<thead>
<tr>
<th>User ID</th>
<th>Creation Date</th>
<th>Tweets Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>630487309</td>
<td>13-06-14 15:55:47 EST</td>
<td>1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of</td>
</tr>
<tr>
<td>35458727</td>
<td>13-10-18 22:59:17 EST</td>
<td>@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking</td>
</tr>
</tbody>
</table>
For Further Information

www.VisualAnalytics-CCI.org
VACCINE: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
Purdue University
ebertd@purdue.edu
VACCINE - Who We Are: International Team of Experts
75+ Faculty, 29 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California San Diego
- University of Maryland
- University Washington
- Arizona State University
- Bethune-Cookman U.
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
What We Do: Visual Analytics

Visual Analytics\(^1\) is the science of **analytical reasoning** facilitated by **interactive visual interfaces**

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

---

\(^1\) Illuminating the Path: The R&D Agenda for Visual Analytics
What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools.

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Ebola Vs. Pandemic Influenza (H1N1)
Day 34 With Origin of Chicago, IL

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Scientific Discovery

- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization

- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models

- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality
Precision Agriculture VA
Business Visual Analytics

• Risk-based decision making and resource allocation
  • Coast guard operational risk assessment model
  • Helping to prioritize efforts to minimize risk

• Competitive Intelligence
  • Visual analytics system for business intelligence

• EconVIS
  • Visual analytics in various economic problems
  • Improving decision making and identifying key motivations in knowledge creation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • Minnesota: food distribution model
  • Texas: simulated and historical weather (hurricanes, storms)
  • UNCC: critical infrastructure
  • Purdue: roads + interaction visual analytics tool

• **Challenge**: Combine interactive VA with complex simulation models
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

**Impacts:**
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

**VALET delivered:**
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

**iVALET delivered:**
- October 2011: Purdue, WL Police, Lafayette PD
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper
• Used by U.S. Coast Guard
  • District 8 for events in 2014 - Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations
• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
• Ohio State Football – more successful than commercial tools
Social Media Analysis and Reporting Toolkit (SMART)

Control Panel:
- Filter:
  - Keywords:
  - Start date:
  - End date:
- Spatial Filter
- Visualization:
  - Topic Model
  - Content Line
  - Tag Map
  - Heat Map
  - Tweet Analyzer
  - Tweet Classifier
- View:
  - Topic Model View
  - Tweet Table
- Search Places:
  - Search
- Reset Settings:
  - Reset

Map:
- Time: 2013-04-15 16:00:00 - 2013-04-15 16:56:49
- Locations:
  - Emergency
  - Security
  - Weather

Tweets:
- JFK Library is a separate incident. Electrical fire. No relation to #BosMarathon explosion.
- My heart goes out to the people in the explosion today... I'm near the location the explosion occurred. My prayers go to everyone who was hurt.
- JFK Library is a separate incident. Electrical fire. No relation to #BosMarathon explosion.
- Our hearts go out to anyone affected by the Boston Marathon bombing today.
- We're in Boston for the first time and this happened to be an explosion at the Boston Marathon. We are okay, everyone.
- #BostonMarathon bomb attack. Prayers out to those hurt.
- Boston Police is doing a coordinated defusal of a third bomb right by the finish line.
- Glad to be safe and that this is an unlikely terror. People freaking out everywhere and I'm glad I was no close there was. aigay.
- Marathon explosion caused probablities: 10% gas main, 10% foreign criminal, 80% drone strike.
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  • Boston Marathon bombing
  • Superbowl 2014 in NJ
  • Riverboat Festival in Louisville October 2014
  • Fleet Week 2014 in San Francisco

• During recurring events
  • Purdue football games – the poor tuba player

• During unexpected events
  • Pensacola tornado
  • Keene, NH Pumpkin Festival

• Detecting daily issues
  • Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14
Example From Ohio State Football 11/29/14
Imagine a bone or something was popped out of TTU leg. Training staff huge hit away that he was going to have to be carted off.
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:07:50 EDT EVERYONE SKIP SCHOOL TOMORROW RT “@TaZeNaToR1: IM LITERALLY GOING TO KILL SOMEONE”

Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets?

- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs
May 2015

What can we learn from this?

• Tracking Competitors
• Customer Satisfaction
• Employee Satisfaction
• Early Warning System
John Deere
Ten Dominant Topics Related to John Deere
Keywords (Contentlens) in the 3 Hotspots

Des Moines, Iowa

Davenport, Iowa

Monterrey, Mexico
Hashtags Related to John Deere

- #johndeere, #nothingrunslikeadeere
- #tractor, #tractortuesday
- #farm, #farming, #farmlife
- #jobs, #work, #harvest
- #green, #country, #summer
- #diesel, #dieselgenerator, #jasonaldean
- #genset, #wayouthere
Con Expo

May 2015
Tweets related to John Deere skidder

Buster Beans @cinthiabfj - 22h
Just got a new John Deere skidder today pic.twitter.com/c8WIMnXVTV

John Deere @JohnDeere - Apr 24
If a lion is the king of the jungle, a John Deere Skidder is the king of the forest. pic.twitter.com/afZ76bXvVT

Service Manuals @EbookManuals - Apr 4
#download John Deere 848G Skidder Technical Manual Download TM15622A drivr.it/5KChgF @EbookManuals

J Lowe @jennyfaithlowe - Apr 22
Firestone 16 ply Skidder Tires on john deere wheels 7/8ths holes #ebaymobile bit.ly/1nn9J24

Want Ad Digest @Wantaddigest - Apr 16
John Deere 540 B Cable Skidder LOWVILLE NY: 23x/x26 tire ring chains, runs very good, needs diamond chains... bit.ly/1ePVzwM

Burge Mcm anus @manualdownloads - Apr 13
John Deere 640D Skidder. 648D Operation and Tests Manual drivr.it/5NyshS
## Tweets related to ‘maintain, maintenance, repair’

<table>
<thead>
<tr>
<th>User ID</th>
<th>Creation Date</th>
<th>Tweets Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>933628327</td>
<td>13-09-09 14:27:17 EST</td>
<td>And, for your viewing pleasure, a nicely maintained 1930s John Deere. <a href="http://t.co/Zavel5QnDI">http://t.co/Zavel5QnDI</a></td>
</tr>
<tr>
<td>50371494</td>
<td>13-06-07 10:13:50 EST</td>
<td>The Class of 2013 presents its class gift at commencement – a John Deere Gator for the Maintenance Department. <a href="http://t.co/ju1r6h5Ulm">http://t.co/ju1r6h5Ulm</a></td>
</tr>
<tr>
<td>194325743</td>
<td>13-08-16 15:07:21 EST</td>
<td>Rondorf yard maintenance! #chevy #johndeere <a href="http://t.co/Hyhy5pTNR">http://t.co/Hyhy5pTNR</a></td>
</tr>
<tr>
<td>404120195</td>
<td>13-09-03 16:13:01 EST</td>
<td>Just saw a clarkson maintenance worker ride a John Deere lawn tractor to the gas station. No need to bother with a gas can. #Townieliife</td>
</tr>
<tr>
<td>1038577016</td>
<td>13-07-28 22:58:57 EST</td>
<td>#labor #jobs Industrial Maintenance Technician: position with John Deere Water: Industrial Maintenance... <a href="http://t.co/azEBmrPMtE">http://t.co/azEBmrPMtE</a></td>
</tr>
<tr>
<td>502127999</td>
<td>13-08-07 18:47:22 EST</td>
<td>That went easier than I expected. #easyout #johndeere @ Mike s Lawn Mower Repair <a href="http://t.co/8pgiu0Vu7g">http://t.co/8pgiu0Vu7g</a></td>
</tr>
</tbody>
</table>
‘Maintain”

And, for your viewing pleasure, a nicely maintained 1930s John Deere. pic.twitter.com/Zavel5QnD1

The Class of 2013 presents its class gift at commencement - a John Deere Gator for the Maintenance Department. pic.twitter.com/ju1rchEuLm
Tweets related to ‘break down, shut down, stop working’

@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking.

1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of.

<table>
<thead>
<tr>
<th>User ID</th>
<th>Creation Date</th>
<th>Tweets Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>630487309</td>
<td>13-06-14 15:55:47 EST</td>
<td>1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of</td>
</tr>
<tr>
<td>35458727</td>
<td>13-10-18 22:59:17 EST</td>
<td>@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking</td>
</tr>
</tbody>
</table>
VACCINE and Visual Analytics Solutions

David Ebert, Kaethe Beck
Purdue University
Who We Are: International Team of Experts
75+ Faculty, 26 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
**We** enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research
Visual Analytic Solutions: What We Offer

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information
- **Innovative Fielded Solutions**: We provide innovative visual analytic and scalable solutions to the extended homeland security community
- **People and Partnerships**
  - Interdisciplinary world-leading team of researchers and students
  
  "cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"
  
  VADM Robert Parker, 2012 MRS Keynote Address
Public Safety & Health

• Public safety visual analytics
  • cgSARVA
    • Coast Guard Search And Rescue Visual Analytics
  • MERGE
    • Mobile Emergency Response Guide

• Public health visual analytics
  • LAHVA
    • Linked Animal-Human Health Visual Analytics
  • RVF
    • Rift Valley Fever
    • Decision support environment for epidemic modeling and responses
  • PanVis
    • Pandemic influenza modeling and visualization tool
  • Cancer Care Engineering
Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Multi-Visualization, Multivariate VA for Risk-Based Decision Making
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

• Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
• Economic sector impact, local and national impact
• Impact and effectiveness of alternative mitigation strategies
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)
Interactive Operational Performance Assessment Report (iOPAR)

- OPAR
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas
- iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Anaylze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
VACCINE Spatiotemporal VA

- Coast Guard visual analytics
  - Search and rescue
  - Risk-based resource allocation
  - Risk-based operational performance
- Anomaly detection
- Situational awareness
- Predictive policing
Visual Analytics Law Enforcement Toolkit (i)VALET
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:

• In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
• Mobile version being released to public (February 2013) for community-based policing
• Investigating correlation factors
• Analyzing time of day problems and improving accuracy of police record management system
• Novel statistical predictive model incorporated for planning
• Incorporating predictive alerts

VALET delivered:
• Spring 2011: WL, Lafayette Police
iVALET delivered:
• October 2011: Purdue, WL Police
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

• **Purpose**: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• **Collaborating Institution(s)**:
Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

• **End-User(s)**: Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• System of systems: binds together multiple simulations models from collaborators into coherent whole
  • Minnesota: food distribution model
  • Texas: simulated and historical weather (hurricanes, storms)
  • UNCC: critical infrastructure
  • Purdue: roads + interaction visual analytics tool

• Challenge:
  combine interactive VA with complex simulation models
System Interface
Complex Decision Making: Advanced Decision Support Tools (e.g., Rift Valley Fever)
Social Media: Real-time Twitter Visual Analytics
(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction using LDA and novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
Social Media Visual Analytics: SMART
Anomaly Detection
Explosion Area in Boston

Keywords that have been used most often

Tweets

Topic Analysis

VACCINE
First Response (Tweet & Picture)

1 minute right after the incident

[Map and photograph of a street scene with people and a fire in the background]
Police Patrolling The River

Sinead Kennedy
@SineadD_Kennedy

Police patrolling the Charles River after the #bostonmarathon
pic.twitter.com/yuzq4OZ9mB

2:16 PM · 16 Apr 2013 from Boston, MA
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00

Supermarket
Park
Shelter
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
For Further Information

www.VisualAnalytics-CCI.org
VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)

- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America
What Our Visual Analytic Solutions Offer

**We** enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the *right information*, in the *right format* within the *right time* to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Public Safety & Health

- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide

- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering
Ebola Vs. Pandemic Influenza (H1N1)

Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Visual Analytics Uses for Public Safety

• Risk visualization and analysis
• Predictive analytics
• Uncertain decision making
• Alternative evaluation and consequence investigation
• Trend analysis, clustering, anomaly detection
• Multisource, multimedia massive data integration & analysis
• Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
Example Tools
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

**IMPACT:**
- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be “interpreted”
- More than 140 users and 2000 graffiti and tattoo images acquired

**GARI delivered:**
- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
- Spring 2014
  - Public reporting version
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

• **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

• **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: Combine interactive VA with complex simulation models
VASA COMPONENT: WEATHER

**NOAA**: National Oceanic and Atmospheric Administration

**ADCIRC**: Advanced Circulation

**Historic data** (Irene, Sandy, etc)

Prepares **event datasets** from server

Visualizes **hurricane cone** over time using slider

Generates **inputs** to downstream components

MODEL

PROXY
VASA COMPONENT: CRITICAL INFRASTRUCTURE

Weather

Critical Infrastructure

Routing

Supply Chain

Weather proxy

Infrastructure proxy

Routing proxy

Supply chain proxy

USER
VASA COMPONENT: CRITICAL INFRASTRUCTURE

**Vu** environment with submodels

**Simplified connectivity graph** of important structures

E.g. Electric grids, telecom networks, gas distribution

**Example**: show impact of hurricanes on restaurants
VASA COMPONENT: ROUTING
VASA COMPONENT: ROUTING

Maintains **road network** for critical infrastructure

Inputs: barriers and closures

Outputs: new transport routes

Simulation engine: **ArcGIS Server**

Approximates disabled routes and facilities

Visualizes disabled routes and facilities

MODEL

PROXY
VASA COMPONENT: SUPPLY CHAIN

Weather

Critical Infrastructure

Routing

Supply Chain

network

network

network

network

Weather proxy

Infrastructure proxy

Routing proxy

Supply chain proxy

VASA WORKBENCH

USER
Supply chains depend on **business** and **goods**

Our models: poultry firm + fast food = farm to restaurant

Discrete event simulation on chain in geolocated facilities

Accepts external inputs (**weather** and **roads**)

Supports **road closures**, **powerless stores** and **flooding** (polygons)

Optimizes distribution and even redistributes products

**Food contamination** also modeled and visualized
Example: U.S. Hurricane Season

34-knot winds batter the coast; critical infrastructure proxy estimates impacted restaurant

Hurricane Irene hits North Carolina on August 27, 2011
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

**Impacts:**
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
- October 2011: Purdue, WL Police, Lafayette PD
VALET Layout

Calendar View

Time Series View

Prediction Panel

Map View

Clock View

Menus

Time Slider Clock View

Prediction Panel
Calendar Display

Selecting a date brings up more information

Displays number of crimes that occurred on certain days
Crimes In A Single Day From 6PM – 9PM
All crimes in the month of September at 3PM.
Select Specific Region To Display Crimes

All burglaries in the month of October in downtown Lafayette

Specific details on crime
Example: More Data Sources – Intoxication

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Football season
- Home
- Away

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis

December 2014
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

• Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx

• Economic sector impact, local and national impact

• Impact and effectiveness of alternative mitigation strategies
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
cgSARVA Results – Aircraft coverage

Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
cgSARVA District 8
Mississippi River

December 2014
D8 SAR Cases:
Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport

Aggregate of SAR Cases 2006-2013
D8 Hazard To Navigation Cases:
Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport

Aggregate of cases 2006-2013
D8 MS Cases:
Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport

Aggregate of MS Cases 2006-2013. Red indicates cases which status is “open”
SAR Cases - Overview

- All SAR cases near the Mississippi River in District 8.
- 10/2/2006-9/29/2013
SAR Case Distribution in D8

Aggregate of SAR Cases 2006-2013 and respective word cloud:

SAR Cases don’t follow direct seasonal pattern:
SAR Data Around The Southeast Portion Of MS River

- Three D8 stations
  - STA GRAND ISLE
  - STA NEW ORLEANS
  - STA VENICE
- 10/2/2006--9/29/2013
MS Cases-Overview

- Marine Safety
- MS taken from MISLE column: “Case Distress Class”
- 10/2/2006-9/29/2013
- MS Cases include the following:
  - Alleged Violation of Law/Regulation
  - Loss of Life/Injury
  - Equipment Failure
  - Fire
  - Allision
  - Explosion
  - Grounding
  - Hazardous Vessel Operation
  - Capsize
  - Hazard to Navigation
  - Collision
  - Force Majeure Claim
  - Breakaway
  - Maritime Explosive Ordnance Incidents
  - Irregular Navigation Incident
  - Waterway Restriction/Closure
  - Near-miss Situation
  - Special Operation
  - Heavy Weather Damage
  - Bridge Closures
  - Vsl Inspection/PSC
  - Cargo/Container Inspection
  - Facility Inspection
  - Towing Vsl Exam
  - Cargo Transfer Monitor
D8 MS Cases:
Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport

Aggregate of MS Cases 2006-2013. Red indicates cases which status is “open”
MS Case Distribution (in D8)

Aggregate of MS Cases 2006-2013 and respective word cloud:

MS Cases don’t follow seasonal patterns:
D8 Hazard To Navigation Cases: Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport

Aggregate of cases 2006-2013
Hazard to Navigation Case Distribution
Utility Visualization

- Boat Coverage Map
  - STA GRAND ISLE
  - STA VENICE

Covered Portion of Mississippi River
Utility Visualization

- Rescue21 Towers in District 8

Covered Portion of Mississippi River
Utility Visualization

- FixedWing Coverage Map
  - AIRSTA MOBILE
  - AIRSTA CORPUS CHRISTI

Covered Portion of Mississippi River
Utility Visualization

- Rotary Wing Coverage Map by AirSTA

District 8, AIRSTA NEW ORLEANS

District 9, AIRFAC WAUKEGAN

Covered Portion of Mississippi River

December 2014
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University

ebertd@purdue.edu
Why Use Social Media?

Marysville-Pilchuck High School (WA) Shooting
Time: around 10:40 am PST
(01:40 pm EST ), 10/24/2014
### Earliest Tweets About The Event

<table>
<thead>
<tr>
<th>Time</th>
<th>User</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-10-24</td>
<td>@FatSackMatt</td>
<td>Cops are fucking EVERYWHERE right now</td>
</tr>
<tr>
<td>13:55:45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td>@FatSackMatt</td>
<td>Cops are fucking EVWRYWHERE right now</td>
</tr>
<tr>
<td>13:55:21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>There is a shooting at mp?</td>
</tr>
<tr>
<td>13:55:12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td>@Miguel_Rocha34</td>
<td>do you have any idea who the person with the gun was</td>
</tr>
<tr>
<td>13:54:06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>you got some serious issues that you need to seek help</td>
</tr>
<tr>
<td>13:52:12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td>@Ray_Raaaay21</td>
<td>one of the kids in the lunchroom started shooting and everyone booked it</td>
</tr>
<tr>
<td>13:49:11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>shooting at mp</td>
</tr>
<tr>
<td>13:46:17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>2 people are dead...</td>
</tr>
<tr>
<td>14:17:43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>Wtf is going on?? Shooting at mp?</td>
</tr>
<tr>
<td>14:17:26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>Shooting at Marysville Pilchuck. Praying for the kids and the families.</td>
</tr>
<tr>
<td>14:16:37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td>@Lake_stonerrr</td>
<td>a student shot several students, possible 2 students dead 6 injured</td>
</tr>
<tr>
<td>14:14:39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>Shooting at pikchuck??</td>
</tr>
<tr>
<td>14:10:38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>Something's fucking wrong !!! Or my mom wouldn't be</td>
</tr>
<tr>
<td>14:08:26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>About 5 mins away from the scene. Sad. Another school</td>
</tr>
<tr>
<td>14:07:51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-10-24</td>
<td></td>
<td>@FatSackMatt cops are blocking all the entrances</td>
</tr>
<tr>
<td>14:01:49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some online news feed:
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper
• Used by U.S. Coast Guard
  • District 8 for events in 2014 - Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations
• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
• Ohio State Football – more successful than commercial tools
Social Media Analysis and Reporting Toolkit (SMART)
Marysville-Pilchuck High School (WA) Shooting
Scene at parent reunification center after shooting at #MPHS

Location of the school

TwitterStream

Safety

Security

Weather

14-10-24
16:47:14 EST
Scene at parent reunification center after shooting at #MPHS http://t.co/AgYpohRgVO

Marysville-Pilchuck High School

User ID: 23845798

name: joshtrujillo

friends: 2146
followers: 6674
location: Seattle, WA
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Anomaly Detection Superbowl Weekend
#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More

12:56 PM - 12 Oct 2014
Ebola tweets during Fleet Week 10/12/14
Where Can It Be Helpful?

- **During large-scale planned events, e.g.,**
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- **During recurring events**
  - Purdue football games – the poor tuba player

- **During unexpected events**
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- **Detecting daily issues**
  - Teen threatening violence – Louisville, KY
Where Can It Be Helpful?

• During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

• During recurring events
  - Purdue football games – the poor tuba player

• During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

• Detecting daily issues
  - Teen threatening violence – Louisville, KY
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
Keene, NH Pumpkin festival
10/18/14 – 10/19/14 Overview

Major topics:

Security classifier became prominent.
Keene, NH
Time range: 10/18 12:00pm – 10/19 06:00am, Security classifier
The picture was later used by ABC news(link).
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat


Follow Up Tweets

- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072
SMART: Network Linkage
Summary

• Powerful tool for daily use and during special events
• Realtime monitoring and email alerts
• More features than many tools on the market
• How to get SMART?
• Visit: www.vaccinetools.org
• Contact us: vaccine@purdue.edu
Conclusions

• Visual Analytics is a powerful tool for harnessing big data
• Empowering the user, not replacing them

• If you have a problem…
  we can develop a solution!
Visual Analytics for Security Applications

VACCINE, Purdue University
David S. Ebert

December 18, 2014
Team Profile

- Purdue – VASA workbench and visualization
- UNC Charlotte – critical infrastructure
- U. Minn. (NCFPD) - supply chain
- U. Texas (TACC) – disaster simulation

- U. Konstanz
- U. Stuttgart
- Fraunhofer IGD
- Siemens
- German utilities
Society is under threat from many sources...
When DISASTER strikes...

HOW CAN WE PREPARE?
WHEN REAL-WORLD EXERCISES ARE COSTLY AND DANGEROUS

simulation

IS THE ANSWER
Distributed component-based framework for computational steering of systems-of-systems simulations for societal infrastructure based on a visual analytics approach
VASA Main Idea:
Pipeline Of Asynchronous Simulations
(System of System)
VASA Workbench

Interactive **desktop** tool for a distributed system

Control **distributed simulations** using REST API

Visual analytics **dashboard** w/ multiple coordinated views

**Simulation proxy** provides real-time response

Configure + steer + explore (simulation models)
VASA COMPONENT: WEATHER

Weather

Critical Infrastructure

Routing

Supply Chain

Weather proxy

Infrastructure proxy

Routing proxy

Supply chain proxy

USER
VASA COMPONENT:
CRITICAL INFRASTRUCTURE
VASA COMPONENT: CRITICAL INFRASTRUCTURE

*Vu* environment with submodels

**Example**: show impact of hurricanes on restaurants

**E.g.** Electric grids, telecom networks, gas distribution

Simplified *connectivity graph* of important structures

**PROXY**
VASA COMPONENT: ROUTING
VASA COMPONENT: ROUTING

Maintains **road network** for critical infrastructure

Simulation engine: **ArcGIS Server**

**Inputs:** barriers and closures

**Outputs:** new transport routes

Approximates disabled routes and facilities

Visualizes disabled routes and facilities

MODEL

PROXY
VASA COMPONENT: SUPPLY CHAIN

Weather proxy → Infrastructure proxy → Routing proxy → Supply chain proxy

Weather → Critical Infrastructure → Routing → Supply Chain

User
VASA COMPONENT: SUPPLY CHAIN

Supply chains depend on **business** and **goods**

**Our models**: poultry firm + fast food = farm to restaurant

**Discrete event** simulation on chain in geolocated facilities

Accepts external inputs (**weather** and **roads**)

Supports **road closures, powerless stores** and **flooding** (polygons)

**Optimizes** distribution and even redistributes products

**Food contamination** also modeled and visualized
VASA

Weather

Critical Infrastructure

Routing

Supply Chain

Weather proxy

Infrastructure proxy

Routing proxy

Supply chain proxy

USER

V A S A W O R K B E N C H
Hurricane Irene hits North Carolina on August 27, 2011.

34-knot winds batter the coast; critical infrastructure proxy estimates impacted restaurant.

Hurricane Irene hits North Carolina on August 27, 2011.
Transition: Initial Partners

- FEMA – regional FEMA feedback:
  - Unprecedented work
  - Whole community approach
- U.S. Coast Guard
  - Emergency preparedness interest
  - Dire need with no apparent solution available
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• System of systems: binds together multiple simulations models from collaborators into coherent whole
  • Minnesota: food distribution model
  • Texas: simulated and historical weather (hurricanes, storms)
  • UNCC: critical infrastructure
  • Purdue: roads + interaction visual analytics tool

• Challenge: Combine interactive VA with complex simulation models
Contact

- ebertd@purdue.edu
- www.VisualAnalytics-CCI.org
Visual Analytics Solutions for Data-Driven Decision Making

David Ebert
Purdue University

Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina Charlotte
- University of Washington
- Arizona State University
- Oak Ridge National Lab
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
VACCINE: What We Do –
Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

• Provide the right information, in the right format within the right time to solve the problem
• Turn data deluge into a pool of relevant, actionable knowledge
• Enable user to be more effective from planning to detection to response to recovery
• Enable effective communication of information

Approach: Partner-driven solutions and research

Visual Analytics Uses for Public Safety

• Risk visualization and analysis
• Predictive analytics
• Uncertain decision making
• Data-driven policy analysis
• Alternative evaluation and consequence investigation
• Trend analysis, clustering, anomaly detection
• Multisource, multimedia massive data integration & analysis
• Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment
• Evaluate courses of action
• Evaluate above at both Strategic and Tactical/Operational level

U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant's budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy:
SAR Cases November 2011 NJ/NYC Area

Example: D8 Western Rivers and Coastal Operations - SAR Cases
Filtered to region between Station Lake Charles and Station Gulfport
Aggregate of SAR Cases 2006-2013
Example: D8: Maritime Safety Cases
Filtered to region between Station Lake Charles and Station Gulfport

Aggregate of MS Cases 2006-2013: Red indicates status is “open”

Example: D8 Hazards to Navigation
Filtered to region between Station Lake Charles and Station Gulfport

Aggregate of cases 2006-2013
Response Efficiency – Potential Future Assets

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)

cgSARVA Results – Aircraft coverage

Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River:
District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

**Impacts and Accomplishments:**
- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems:** binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota:** food distribution model
  - **Texas:** simulated and historical weather (hurricanes, storms)
  - **UNCC:** critical infrastructure
  - **Purdue:** roads + interaction visual analytics tool
- **Challenge:** Combine interactive VA with complex simulation models
Social Media: Real-time Visual Analytics
(Purdue, Stuttgart, Penn St.)

• Anomalous topic extraction (LDA, novel STL based estimation)
• Dynamically linked views
• Topics extracted shown as a dynamic word cloud
• Sources: Twitter, facebook, flickr
• Uses:
  • USCG Sector Ohio Valley
  • Indianapolis PD, Purdue PD
  • Fleet week Nov. 2014: USCG PAC

Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)

Hurricane Sandy

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
Coast Guard False Distress Call Analysis Using SMART

1. Overview of the problem

<table>
<thead>
<tr>
<th>Case</th>
<th>Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1:22 AM on 5 January 2014</td>
<td>Buxton, NC</td>
</tr>
<tr>
<td>2</td>
<td>5-6:00 PM on 11 August 2014</td>
<td>Annapolis, MD</td>
</tr>
</tbody>
</table>

2. Data

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Time Range</th>
<th>Data Source</th>
<th>Geo Bound</th>
<th>After filtering / total size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geo Twitter</td>
<td>01/2013 – present</td>
<td>VACCINE Twitter DB</td>
<td>US</td>
<td>9,335 / 1,484,449,638</td>
</tr>
<tr>
<td>Non-geo Twitter</td>
<td>01/2014 – present</td>
<td>VACCINE Twitter DB plus Twitter.com</td>
<td>None</td>
<td>7,422 / 959,291,538</td>
</tr>
<tr>
<td>Public Facebook</td>
<td>08/2014 – present</td>
<td>VACCINE DB</td>
<td>None</td>
<td>238 / 3,183,363</td>
</tr>
</tbody>
</table>

3. Process

We monitored social media traffic, filtered based on carefully defined keyword lists, restricted to detailed area and time ranges, created classifiers to categorize messages of different topics and narrowed to suspicious messages and users.

Results – Buxton Case (Jan. 5th, 2014)

Searched tweets posted 12/2013-1/2014 within a radius of 60 miles, covering LOB (Line of Bearing) provided – Only tweets related to emergency ferries found.
Results – Illegal Immigration and False Call

- Target: Near So. California coastline in 2013 and 2014
- Most relevant tweets posted in Los Angeles and San Diego about 17 illegal immigrants found on a panga boat on April 4, 2014
- None related to hoax calls

For Further Information

www.VisualAnalytics-CCI.org
Visual Analytics Solutions for Data-Driven Decision Making

David Ebert
Purdue University
Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

- **Federal Operating Components:**
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

- **Law Enforcement**
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

- **State Fusion Centers**
  - Ohio (SAIC), Indiana (IIFC)

- **Companies**
  - Motorola, Kimberly Clark, Banfield, Bank of America
VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• Qualitative, fuzzy, and social data
  • Preferences, significance of one relationship over another; Trust

• Context
  • Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it into the collected data

• Spurious vs. Significant
  • Big data means more statistically significant events and correlations, but they may not have any relevance
  • Increases noise to signal ratio

• Big problems
  • Complex, multifaceted, multiparameter big challenges with unquantified dependencies
Visual Analytics Uses for Public Safety

• Risk visualization and analysis
• Predictive analytics
• Uncertain decision making
• Data-driven policy analysis
• Alternative evaluation and consequence investigation
• Trend analysis, clustering, anomaly detection
• Multisource, multimedia massive data integration & analysis
• Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Visual Analytics Law Enforcement Toolkit (i)VALET
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

**Impacts:**
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
Example for Crime Analytics: VALET
Example: Drunkenness / Public Intoxication

- Home vs. Notre Dame
  - PU Lost: 10-38
- Homecoming (Sat.)
- PU vs. Illinois
  - PU Won: 21-14
- PU vs. Iowa
  - PU Lost: 21-31

Football season
- Home
- Away

Day-of-the-Week
- Mon
- Tue
- Wed
- Thu
- Fri
- Sat
- Sun

PU vs. Notre Dame
PU Lost: 10-38
PU vs. Illinois
PU Won: 21-14
PU vs. Iowa
PU Lost: 21-31
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s):**
  Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s):** Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**: combine interactive VA with complex simulation models
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko
IMPACT:
• Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
• Will allow the graffiti images to be “interpreted”
• More than 140 users and 1300 graffiti images acquired

GARI delivered:
• Summer/Fall 2011:
  • IPD gang detectives
  • IPD at large
  • Ind Fusion Gang Task Force
  • Gang detectives across Indiana
• Spring/Summer 2013:
  • Cook County, IL
  • INGang ownership
  • Tattoo database added
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
Social Media: Real-time Twitter Visual Analytics
(Purdue, Stuttgart, Penn St.)

• Anomalous topic extraction using LDA and novel STL based remainder estimation technique
• Dynamically linked views providing options to monitor emerging / emergent twitter feeds
• Topics extracted shown as a dynamic word cloud
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)
SMART: Network Linkage
SMART Anomaly Detection
Anomaly Detection
Oil Spill in Mississippi River

2014/02/23
Non-geo Tweets On Twitter About Oil Spill Event

ProPublica @ProPublica
An oil spill just shut down 65 miles of the Miss. River. As production surges, how do we ship oil safely? propub.ca/1dxUL6g @Bloomberg

Bloomberg News
Mississippi Oil Spill Highlights Risk of U.S. Oil Boom
A barge crash that spilled enough oil to temporarily shut a stretch of the Mississippi River highlights the transportation risks of the U.S. energy boom just as regulators respond to several rail...

Guardian Environment @guardianeco
Mississippi River closed at New Orleans after barge crash prompts oil spill
gu.com/p/3n2qphf

Motherboard @motherboard
The Mississippi already a massive toxic cesspool every other day of the year
bit.ly/1fVvUpU

Marcy McGowan @marcylearnin
Oil Spill Shuts Down Sixty-Five Mile Stretch of Mississippi River! Americans Against the Tea Party aattp.org/oil-spill-shut...

EDF-Louisiana @EDF_Louisiana
@oilspill update: Lover Mississippi River reopens after oil spill near New Orleans. Watch the video bit.ly/MZf7ST

Sierra Club @sierriclub
The Mississippi River Was a Toxic Mess Long Before this Weekend’s Oil Spill
sc.org/Mississippi (via @motherboard)

Motherboard
The Mississippi River Was a Toxic Mess Long Before the Oil Spill
Everyone is outraged about a barge spilling oil into the Mississippi—perhaps everyone forgot that it’s a massive toxic cesspool every other day of the year, too.

National Wildlife @WWF
A barge has crashed into a tugboat, spilling 31,500 gallons of oil into the Mississippi River. Read more bit.ly/Mm2Dd

Margaritha Norrega @margareth
31,500 gallons of light crude oil spilled into the Mississippi River weathert.com/news/science/e...

National Wildlife @WWF
New Oil Spill Floods Mississippi River bit.ly/gqzz01

inhabitat @inhabitat
It is unclear when the ships stuck on the Mississippi River will be able to move as crews clean up a new oil spill bit.ly/dp64w

inhabitat
Louisiana Oil Spill Shuts Down New Orleans Port and Parts of the...
The Port of New Orleans and surrounding areas were closed Sunday after a barge collided with a tow boat, causing oil to spill into the Mississippi River. The boats were located between Baton Rouge...
Explosion Area in Boston

Keywords that have been used most often in the area

Tweets

Text:
Oh my god what just happened
Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What's going on?
Multiple people are injured near the Boston Marathon finish line after two explosions.
The #BostonMarathon has been stopped.
Two bombs just went off on Boylston
BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finish line this afternoon.
What the FUCK was that?
Back in Sept., @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind.
Literally what the fuck get me out of here
@DTenenbaum my office right next to it
@FRANCESCalO I figured 3 people would get the joke
This is crazy I seen that blow up #BostonMarathon
can someone tell me what that explosion was? #boston #bostonmarathon
First Response (Tweet & Picture)

1 minute right after the incident
Police Patrolling The River

Sinead Kennedy
@SineadD_Kennedy

Police patrolling the Charles River after the #bostonmarathon pic.twitter.com/yuzq4OZ9mB
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

**IMPACTS:**

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- **Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization**
- Informed Commandant’s budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
ORAM Visualization
Component Overview: Identification of Potential Risks

General understanding on how the risk is distributed between the different areas and missions.

(A) District Choropleth Map for Total Risk across all missions
(B) Risk profiles across all districts and mission areas
(C) District X Risk Profiles across all mission area
Component Overview: Analysis and Assessment

Specific areas can be analyzed by time and details on demand.

(D) Total M10 Risk Values distributed among District X’s station AOR

(E) KDE heatmap showing Total M10 Risk values
Interactive Operational Performance Assessment Report (iOPAR)

- **OPAR**
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas

- **iOPAR** – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Analyze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies
For Further Information

www.VisualAnalytics-CCI.org
One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00
Visual Analytics Solutions for Data-Driven Decision Making

David Ebert
Purdue University
Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions

• Purdue University
• Georgia Institute of Technology
• Pennsylvania State University
• Stanford University
• University of North Carolina at Charlotte
• University of Washington
• Arizona State University
• Simon Fraser University
• University of British Columbia
• Justice Institute of British Columbia
• Ontario Institute of Technology
• Dalhousie University

• University of Houston, Downtown
• Virginia Tech
• Indiana University
• Florida International University
• University of Texas at Austin
• Morgan State University
• Navajo Technical College
• University of Stuttgart
• University of Swansea
• Oxford University
• University of Calgary
• University of Manitoba
• Carleton University
• University of Victoria
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

- Law Enforcement
  - Over 40 local and state agencies (GA, IN, IL, OH, PA, NC, NY)

- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)

- Companies
  - Motorola, Kimberly Clark, Banfield, Bank of America
VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research
Visual Analytic Solutions: What We Offer

• **Improved Effectiveness:** We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information

• **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community

• **People and Partnerships**
  - Interdisciplinary world-leading team of researchers and students

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
cgSARVA Results – Aircraft coverage

Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

• **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

• **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

---

Diagram:

- VASA Workbench
- S1: Disasters
- S2: Critical Infrastructure
- S3: Supply Chain & Re-routing
- Purdue
- UNCC
- U Minn
- TACC
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration

- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool

- **Challenge**: Combine interactive VA with complex simulation models
Social Media: Real-time Visual Analytics
(Purdue, Stuttgart, Penn St.)

• Anomalous topic extraction (LDA, novel STL based estimation)
• Dynamically linked views
• Topics extracted shown as a dynamic word cloud
• Sources: Twitter, facebook, flickr
• Uses:
  • USCG Sector Ohio Valley
  • Indianapolis PD, Purdue PD
  • Fleet week Nov. 2014: USCG PAC
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)
SMART: Network Linkage
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko

Documents in

Knowledge out
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

**IMPACT:**

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be “interpreted”
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- **Summer/Fall 2011:**
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana

- **Spring/Summer 2013:**
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
For Further Information

www.VisualAnalytics-CCI.org
Visual Analytics Solutions for Data-Driven Decision Making

David Ebert
Purdue University
We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

**Impacts:**
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level
Social Media: Real-time Twitter Visual Analytics
(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction using LDA and novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)
SMART: Network Linkage
SMART Anomaly Detection
Anomaly Detection
Oil Spill in Mississippi River

2014/02/23
<table>
<thead>
<tr>
<th>User ID</th>
<th>Creation Date</th>
<th>Tweets Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>229758274</td>
<td>14-02-23 17:13:32 CST</td>
<td>Oil spill in the Mississippi River ??</td>
</tr>
<tr>
<td>2341900254</td>
<td>14-02-23 17:23:07 CST</td>
<td>When ppl spill tea to me I be like <a href="http://example.com">Link</a></td>
</tr>
<tr>
<td>180897996</td>
<td>14-02-23 19:10:48 CST</td>
<td>WTF!!! because NOLA needs more oil in its water? RT @friedbiscuit: Oil spill closes part of Miss. River in... <a href="http://example.com">Link</a></td>
</tr>
<tr>
<td>235383494</td>
<td>14-02-23 19:21:46 CST</td>
<td>I like my little sister now. A year ago I hated the bitch. #fatalispill ?</td>
</tr>
<tr>
<td>632289479</td>
<td>14-02-23 20:20:03 CST</td>
<td>I'm still waiting for @FranklyDimes review of the HOA episode. I know he will be spilling all the tea later...</td>
</tr>
<tr>
<td>20611227</td>
<td>14-02-23 22:08:24 CST</td>
<td>Aerial photos show stalled traffic on Mississippi River after barge oil spill near Vacherie <a href="http://example.com">Link</a></td>
</tr>
<tr>
<td>624332153</td>
<td>14-02-24 22:38:54 CST</td>
<td>Amazing pics RT @NOLAnews: Aerial photos of stalled traffic on Miss. River. Barge oil spill near Vacherie... <a href="http://example.com">Link</a></td>
</tr>
</tbody>
</table>

**Additional Notes:**
- [Link](http://example.com) for image and additional resources.
Non-geo Tweets On Twitter About Oil Spill Event
Explosion Area in Boston

Keywords that have been used most often in the area

Tweets

Text

Oh my god what just happened
Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What’s going on?

Two bombs just went off on boylston

BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finish line this afternoon.

What the FUCK was that

Back in Sept, @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind.

Literally what the fuck get me out of here

@OTenenbaum my office right next to it

@FRANCESCAIO I figured 3 people would get the joke


This is crazy i seen that blow up #bostonmarathon

can someone tell me what that explosion was!? #boston #bostonmarathon

VACCINE
First Response (Tweet & Picture)

1 minute right after the incident
Police Patrolling The River

Police patrolling the Charles River after the #bostonmarathon pic.twitter.com/yuzq4OZ9mB
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC:
10/29, 8:00 PM
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant’s budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: 
SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
cgSARVA 2.0 New Features (1/2)

• Rotary-Wing Coverage
  • The new shapes depict the maximum range from land that can be achieved if the aircraft returns to the nearest point of land possible

• Offshore and Inland Filtering

• Siting Plans
  • Shows the current inventory of boats at the station

• Reliability Coverage
  • Indicates the probability of the boats at the given station being able to respond to that point on the water
cgSARVA 2.0 New Features (2/2)

• Rescue 21
  • Visualize the VHF Receiving Range of the R-21 Towers and allows the user to turn on/off each tower

• Patrol Boats
  • Shows the home port for the different cutters, and displays the level of effort that is defined by the percentage of presence the group of patrol boats provide to a Sector relative to a number of hours parameter
cgSARVA Results – Aircraft coverage

Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.
cgSARVA Results – Offshore and Inland Filtering

Before filtering. Year 2010

Filtered cases using 5 NM inland and 50 NM offshore. Year 2010
cgSARVA Results – Rescue 21
cgSARVA Results – Patrol Boat

Showing homeport of Atlantic Beach and its 3 cutters.

The choropleth map represents the aggregate number of hours for all the boats in the Sector.
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
ORAM Visualization
Component Overview: Identification of Potential Risks

General understanding on how the risk is distributed between the different areas and missions.

(A) District Choropleth Map for Total Risk across all missions

(B) Risk profiles across all districts and mission areas

(C) District X Risk Profiles across all mission area
Component Overview: Analysis and Assessment

Specific areas can be analyzed by time and details on demand.

(D) Total M10 Risk Values distributed among District X’s station AOR

(E) KDE heatmap showing Total M10 Risk values
Interactive Operational Performance Assessment Report (iOPAR)

• OPAR
  • Standard report within the Coast Guard Business Intelligence (CGBI) system
  • Displays resource use and performance by core CG mission areas

• iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  • Analyze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

• Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
• Economic sector impact, local and national impact
• Impact and effectiveness of alternative mitigation strategies
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose**: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s)**:
Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s)**: Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply

![Diagram showing collaborations between Purdue, UNCC, U Minn, and TACC]
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: combine interactive VA with complex simulation models
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko
IMPACT:
• Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
• Will allow the graffiti images to be “interpreted”
• More than 140 users and 1300 graffiti images acquired

GARI delivered:
• Summer/Fall 2011:
  • IPD gang detectives
  • IPD at large
  • Ind Fusion Gang Task Force
  • Gang detectives across Indiana
• Spring/Summer 2013:
  • Cook County, IL
  • INGang ownership
  • Tattoo database added
For Further Information

www.VisualAnalytics-CCI.org
Visual Analytics Solutions for Data-Driven Decision Making

David Ebert
Purdue University
Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

- **Federal Operating Components:**
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

- **Law Enforcement**
  - Over 40 local and state agencies (GA, IN, IL, OH, PA, NC, NY)

- **State Fusion Centers**
  - Ohio (SAIC), Indiana (IIFC)

- **Companies**
  - Motorola, Kimberly Clark, Banfield, Bank of America
VACCINE: What We Do –
Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

• Provide the right information, in the right format within the right time to solve the problem
• Turn data deluge into a pool of relevant, actionable knowledge
• Enable user to be more effective from planning to detection to response to recovery
• Enable effective communication of information

Approach: Partner-driven solutions and research
Visual Analytic Solutions: What We Offer

- **Improved Effectiveness:** We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information

- **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community

- **People and Partnerships**

  Interdisciplinary world leading team of researchers and students

  “cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

  VADM Robert Parker, 2012 MRS Keynote Address
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

• **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

• **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: Combine interactive VA with complex simulation models
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

**IMPACT:**
- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be “interpreted”
- More than 140 users and 2000 graffiti and tattoo images acquired

**GARI delivered:**
- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
- Spring 2014
  - Public reporting version
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:
- October 2011: Purdue, WL Police
Example for Crime Analytics: VALET

Map View

Time Series View

Menus

Clock View

Twitter monitoring

Time Slider

Calendar View

Map View
Example: Drunkenness / Public Intoxication

Football season
- PU vs. Notre Dame
  PU Lost: 10-38
- Homecoming (Sat.)
- PU vs. Illinois
  PU Won: 21-14
- PU vs. Iowa
  PU Lost: 21-31

Day-of-the-Week
- Home
- Away

[Map with data points and statistics]
Social Media: Real-time Visual Analytics
(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction (LDA, novel STL based estimation)
- Dynamically linked views
- Topics extracted shown as a dynamic word cloud
- Sources: Twitter, facebook, flickr
- Uses:
  - USCG Sector Ohio Valley
  - Indianapolis PD, Purdue PD
  - Fleet week Nov. 2014: USCG PAC

Grand Prix Weekend - Purdue University
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
Social Media Analysis and Reporting Toolkit (SMART)
SMART: Network Linkage
For Further Information

www.VisualAnalytics-CCI.org
Visual Analytics Solutions for Spatially Data-Driven Decision Making

David Ebert
Purdue University
Visual Analytics

Visual Analytics¹ is the science of **analytical reasoning** facilitated by **interactive visual interfaces**

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

1. *Illuminating the Path: The R&D Agenda for Visual Analytics*
VACCINE - Who We Are: International Team of Experts
75+ Faculty, 27 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University

- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)

- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America
VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research
Visual Analytic Solutions: What We Offer

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information

- **Innovative Fielded Solutions**: We provide innovative visual analytic and scalable solutions to the extended homeland security community

- **People and Partnerships**
  - Interdisciplinary world-leading team of researchers and students

“cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…”

VADM Robert Parker, 2012 MRS Keynote Address
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• Qualitative, fuzzy, and social data
  • Preferences, significance of one relationship over another; Trust

• Context
  • Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it into the collected data

• Spurious vs. Significant
  • Big data means more statistically significant events and correlations, but they may not have any relevance
  • Increases noise to signal ratio

• Big problems
  • Complex, multifaceted, multiparameter big challenges with unquantified dependencies
Visual Analytics Uses for Public Safety

• Risk visualization and analysis
• Predictive analytics
• Uncertain decision making
• Data-driven policy analysis
• Alternative evaluation and consequence investigation
• Trend analysis, clustering, anomaly detection
• Multisource, multimedia massive data integration & analysis
• Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment
• Evaluate courses of action
• Evaluate above at both Strategic and Tactical/Operational level
Our Spatiotemporal Visual Analytics Work

• Techniques:
  • New anomaly and prediction techniques
  • New interactive correlative analysis
  • Natural scale analytics
  • Social medial visual analytics

• Uses:
  • Anomaly detection
  • Situational awareness
  • Predictive policing
  • Coast Guard visual analytics
    • Search and rescue
    • Risk-based resource allocation
    • Risk-based operational performance
Event/Anomaly Detection - STL

- Temporal trend analysis and prediction
  - Seasonal Trend decomposition based on locally weighted regression (STL)
  - Time series data viewed as the sum of components:
    \[ Y = T + S + R \]
    
    \( Y \): original time series
    \( T \): trend component
    \( S \): seasonal (daily/weekly)
    \( R \): remainder component

- Large value of \( R \) indicates substantial variation
- Filter out the noise component
- If value varies by 2 SD from STL expectation, generate Alert
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Respiratory Count

Date

Visual Correlative Analytics: Value of Visual Correlative Exploration?

- Understanding the relationships between spatial and temporal trends
  - Acts as precursor to creating predictive models from data

- Provides insights into the workings of real world environments

- Different forces that determine nature of relationships
  - E.g., one event precursor to another
  - Requires exploration of leads/lags between datasets
  - Determine if one dataset is related to another at a different temporal location

- Aids analysts and decision makers in hypothesis generation
Multivariate Correlative Predictive Analytics

**Two Key Features:**
- Automatic correlation computation against lead/lags
- Temporal and spatial windowing

Enables detection of periodic properties among different data variables
- Allows correlations at multiple spatial and temporal granularities
- Utilizes Pearson's Correlation in current implementation

*Intoxication vs. Noise Complaints*
Spatial Windowing: Drug Abuse Violations vs. Burglaries

• Over geospatial hotspots
The Scale Problem

• Data scale:
  • Coarse scale analysis is quick but can have too much variance or contain combinations of multiple signals reducing effectiveness
  • Fine scale analysis is fast but may limit applicability because of fine resolution and can have data sparseness issues

• User/problem scale
  • Users perform analysis at scales that are natural to them and the task/problem
  • Characteristics might not be well articulated or available digitally
Natural Scale Templates

• Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues

• Some important questions:
  • How do we facilitate appropriate analysis for the casual expert?
  • How can casual expert improve performance of analytical results?
  • How do we effectively communicate the result, accuracy, and certainty?
Overall Historical Distribution:
Spatial and Temporal Variations
Note areas of sparse data

Past 2 Years

March + Tuesday Filter
(Past 10 Years)
Prediction Using KDE Distribution

> 89%
> 78%
> 67%
> 56%
> 45%
> 34%
> 23%
> 12%
> 0%
Prediction During Certain Hours In The Day

9AM – 3PM

3PM – 7PM

9AM – 5PM (Day)

9PM – 3AM (Night)
Summary of Prediction Features

- Novel prediction techniques based on seasonal, day of week, and time of day patterns
- Determination of right spatial and temporal level for reliable predictions
- Correlation of multiple types of crime with lag/lead indicators included
- Location aware updating of relevant information
Visual Analytics Law Enforcement Toolkit (i)VALET
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

**Impacts:**

- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

**VALET delivered:**

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

**iVALET delivered:**

- October 2011: Purdue, WL Police, Lafayette PD
VALET Layout

Calendar View

Time Series View

Map View

Time Slider

Prediction Panel

Clock View

Menus
Calendar Display

Selecting a date brings up more information

Displays number of crimes that occurred on certain days
Crimes In A Single Day From 6PM – 9PM
Crimes During Certain Hours In A Month

All crimes in the month of September at 3PM
Select Specific Region To Display Crimes

All burglaries in the month of October in downtown Lafayette
Example: More Data Sources – Intoxication

- PU vs. Notre Dame: Home, PU Lost: 10-38
- Homecoming (Sat.): PU vs. Illinois:
  - PU Won: 21-14
- PU vs. Iowa: Away, PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame: Home, PU Lost: 10-38
Homecoming (Sat.): PU vs. Illinois:
- PU Won: 21-14
PU vs. Iowa: Away, PU Lost: 21-31

Football season:
- Home
- Away
iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
U.S. Coast Guard COAST/ SARVA (cgSARVA)
Partners: USCG LANT 7, USCG HQ 771, USCG D9, USCG D5, RDC

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony to Congress
• Key component of USCG D9 reallocation plan for 2011-12
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s):** Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s):**
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: Combine interactive VA with complex simulation models
Social Media Analytics and Reporting Tool

David S. Ebert, Purdue University
ebertd@purdue.edu
Why Use Social Media?

Marysville-Pilchuck High School (WA) Shooting
Time: around 10:40 am PST
(01:40 pm EST ), 10/24/2014
# Earliest Tweets About The Event

<table>
<thead>
<tr>
<th>ID</th>
<th>Date/Time</th>
<th>Tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>601648236</td>
<td>14-10-24 14:01:49 EST</td>
<td>@_FatSackMatt cops are blocking all the entrances</td>
</tr>
<tr>
<td>461749842</td>
<td>14-10-24 13:55:45 EST</td>
<td>Cops are fucking EVERYWHERE right now</td>
</tr>
<tr>
<td>461749842</td>
<td>14-10-24 13:55:21 EST</td>
<td>Cops are fucking EVERYWHERE right now</td>
</tr>
<tr>
<td>707981743</td>
<td>14-10-24 13:55:12 EST</td>
<td>There is a shooting at mp?</td>
</tr>
<tr>
<td>818397241</td>
<td>14-10-24 13:54:06 EST</td>
<td>@Miguel_Rocha34 do you have any idea who the person with the gun was</td>
</tr>
<tr>
<td>818397241</td>
<td>14-10-24 13:52:12 EST</td>
<td>you got some serious issues that you need to seek help if you shoot up a fucking high school</td>
</tr>
<tr>
<td>601648236</td>
<td>14-10-24 13:49:11 EST</td>
<td>@Ray_Raaaay21 one of the kids in the lunchroom started shooting and everyone booked it</td>
</tr>
<tr>
<td>601648236</td>
<td>14-10-24 13:46:17 EST</td>
<td>shooting at mp</td>
</tr>
<tr>
<td>707981743</td>
<td>14-10-24 14:17:43 EST</td>
<td>2 people are dead</td>
</tr>
<tr>
<td>335778595</td>
<td>14-10-24 14:17:25 EST</td>
<td>Wtf is going on?? Shooting at mp?</td>
</tr>
<tr>
<td>183421522</td>
<td>14-10-24 14:14:39 EST</td>
<td>@Lake_stonerr a student shot several students, possible 2 students dead 6 injured</td>
</tr>
<tr>
<td>109944382</td>
<td>14-10-24 14:10:38 EST</td>
<td>Shooting at pikchuck??</td>
</tr>
<tr>
<td>599963823</td>
<td>14-10-24 14:08:26 EST</td>
<td>Something’s fucking wrong !!!!! Or my mom wouldn’t be crying about the MP shooting !!!</td>
</tr>
<tr>
<td>130018947</td>
<td>14-10-24 14:07:51 EST</td>
<td>About 5 mins away from the scene. Sad. Another school shooting. MPHS</td>
</tr>
</tbody>
</table>

Some online news feed:
Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

• Used at Boy Scouts of America Jamboree 2013
  • Detected tweet about sniper
• Used by U.S. Coast Guard
  • District 8 for events in 2014 -Detected 3 gang related activities to date
  • PAC for Fleet Week October 2014
  • LANT for SAR hoax call investigations
• Purdue Police for home football games Fall 2014
• Indianapolis Metropolitan Police for special events
• US CBP for investigations Fall 2014 (Boston, AMOC)
Location of the school
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY

1 minute right after the incident
Anomaly Detection Superbowl Weekend
Examples From Fleet Week

Fleet Week Sunday 10/12/14

Tiffany Wilson
@TWilsnTV

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Retweets 6  Favorites 4

12:56 PM - 12 Oct 2014
Ebola tweets during Fleet Week 10/12/14
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Example From Purdue Football 10/12/14

• Prevented violence – the poor tuba player
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco

- During recurring events
  - Purdue football games – the poor tuba player

- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival

- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Evacuation order: 10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC: 10/29, 8:00 PM

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00
Security classifier became prominent.
Keene, NH
Time range: 10/18 12:00pm – 10/19 06:00am, Security classifier)
The picture was later used by ABC news([link]).
Keene, NH
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY
Louisville, KY Oct 16 – School Shooting Threat

Follow Up Tweets


Follow Up Tweets

- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072
SMART: Network Linkage
Conclusions

• Powerful tool for daily use and during special events
• Realtime monitoring and email alerts
• More features than many tools on the market
• How to get SMART?
• Visit: www.vaccinetools.org
• Contact us: vaccine@purdue.edu
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

**IMPACT:**
- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region.
- Will allow the graffiti images to be “interpreted”.
- More than 140 users and 1300 graffiti images acquired.

**GARI delivered:**
- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
For Further Information

www.VisualAnalytics-CCI.org
Visual Analytics Solutions for The Real World

David Ebert
Purdue University
Who We Are: International Team of Experts
75+ Faculty, 26 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT

- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)

- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)

- Companies
  - Motorola, Kimberly Clark, Banfield, Bank of America
VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research
Why VA For Big Data?
(Or What Big Data Analytics Can’t Do)
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• Qualitative, fuzzy, and social data
  • Preferences, significance of one relationship over another; Trust

• Context
  • Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it
    into the collected data

• Spurious vs. Significant
  • Big data means more statistically significant events and correlations, but
    they may not have any relevance
  • Increases noise to signal ratio

• Big problems
  • Complex, multifaceted, multiparameter big challenges with unquantified
    dependencies
Visual Analytics Uses for Public Safety

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation
Visual Analytics Law Enforcement Toolkit (i)VALET
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:

- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
Example for Crime Analytics: VALET
Example: Drunkenness / Public Intoxication

- PU vs. Notre Dame: Home, PU Lost: 10-38
- Homecoming (Sat.): PU vs. Illinois, PU Won: 21-14
- PU vs. Iowa: Away, PU Lost: 21-31
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

- **Purpose:** Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

- **Collaborating Institution(s):**
  Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities

- **End-User(s):** Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

• Support decision-making for extreme weather and natural disaster scenarios
  • Combine real and simulation data
  • Allow “what-if” exploration

• **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  • **Minnesota**: food distribution model
  • **Texas**: simulated and historical weather (hurricanes, storms)
  • **UNCC**: critical infrastructure
  • **Purdue**: roads + interaction visual analytics tool

• **Challenge**: combine interactive VA with complex simulation models
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko
Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)

IMPACT:
• Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
• Will allow the graffiti images to be “interpreted”
• More than 140 users and 1300 graffiti images acquired

GARI delivered:
• Summer/Fall 2011:
  • IPD gang detectives
  • IPD at large
  • Ind Fusion Gang Task Force
  • Gang detectives across Indiana
• Spring/Summer 2013:
  • Cook County, IL
  • INGang ownership
  • Tattoo database added
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
Social Media: Real-time Twitter Visual Analytics
(Purdue, Stuttgart, Penn St.)

• Anomalous topic extraction using LDA and novel STL based remainder estimation technique
• Dynamically linked views providing options to monitor emerging / emergent twitter feeds
• Topics extracted shown as a dynamic word cloud
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)
Social Media Analysis and Reporting Toolkit (SMART)
SMART: Network Linkage
SMART Anomaly Detection
Anomaly Detection
Oil Spill in Mississippi River

2014/02/23
Non-geo Tweets On Twitter About Oil Spill Event

ProPublica @ProPublica 3h
An oil spill just shut down 65 miles of the Miss. River. As production surges, how do we ship oil safely? propub.ca/1uUu69G @Bloomberg

Sierra Club @sierracub 9h
The Mississippi River Was a Toxic Mess Long Before this Weekend’s Oil Spill. sc.org/Mississippi (via @motherboard)

Motherboard
The Mississippi River Was a Toxic Mess Long Before the Oil Spill
Everyone is outraged about a barge spilling oil into the Mississippi—perhaps everyone forgot that it’s a massive toxic cesspool every other day of the year, too.

Guardian Environment @guardianeco Feb 24
Mississippi river closed at New Orleans after barge crash prompts oil spill gu.com/p/Sm2qPhY

National Wildlife @NMWF 12h
A barge has crashed into a tugboat, spilling 31,500 gallons of oil into the Mississippi River. Read more bit.ly/1nrZ20j

Margarette Norrega @margarette 13h
31,500 gallons of light crude oil spilled into the Mississippi River weather.com/news/science...

Habitat @inhabitat Feb 24
It is unclear when ships stuck on the Mississippi River will be able to move as crews clean up a new oil spill bit.ly/Hdp44V

Marcy McGowan @marcymclean 9h
Oil Spill Shuts Down Sixty-Five Mile Stretch of Mississippi River. Americans Against the Tea Party aatp.org/oil-spill-shut...

EDF-Louisiana @EDF_Louisiana 9h
#oilspill update: Lower Mississippi River reopens after oil spill near New Orleans. Watch the video bit.ly/M2z75T

Louisiana Oil Spill Shuts Down New Orleans Port and Parts of the... 
The Port of New Orleans and surrounding areas were closed Sunday after a barge collided with a tow boat, causing oil to spill into the Mississippi River. The boats were located between Baton Rouge...
Explosion Area in Boston

Keywords that have been used most often:
- explosion
- marathon
- people
- happened
- finish
- bombs
- boston
- heard
- crazy
- nine

Text:
- "Oh my god what just happened. Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What's going on?"
- "Multiple people are injured near the Boston Marathon finish line after two explosions. The #BostonMarathon has been stopped."
- "BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finish line this afternoon."
- "What the FUCK was that. Back in Sept., @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind."
- "Literally what the fuck get me out of here."
- "@DTenenbaum my office right next to it."
First Response (Tweet & Picture)

1 minute right after the incident
Police Patrolling The River
Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order: 10/28, 10:30 AM

Hurricane Sandy’s Arrival at NYC: 10/29, 8:00 PM
U.S. Coast Guard Search and Rescue VA (cgSARVA) Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant’s budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007
USCG Port Closure Economic Impact VA
Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

• Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
• Economic sector impact, local and national impact
• Impact and effectiveness of alternative mitigation strategies
For Further Information

www.VisualAnalytics-CCI.org