Segmenting, Grouping and Tracking Vehicles in LIDAR Data

By

Benjamin Coifman, PhD
Associate Professor
The Ohio State University
Department of Civil, Environmental, and Geodetic Engineering
Department of Electrical and Computer Engineering
Hitchcock Hall 470
2070 Neil Ave, Columbus, OH 43210
E-mail: Coifman.1@OSU.edu

and

Mo Wu
The Ohio State University
Department of Electrical and Computer Engineering

and

Keith Redmill, PhD
Assistant Professor- Clinical
The Ohio State University
Department of Electrical and Computer Engineering

and

Douglas A. Thornton
Cyber Embedded Engineer
Battelle Memorial Institute

DISCLAIMER

1. Introduction

This report presents the methodology and results from a study to extract empirical microscopic vehicular interactions from an instrumented probe vehicle. The contributions of this work are twofold: first, the general method and approach to seek a cost-effective balance between automation and manual data reduction that transcends the specific application. Second, the resulting empirical data set is intended to help advance traffic flow theory in general and car following models in particular.[1]

Like many data reduction problems this work seeks to find a balance between automated and manual processing. All too often in problems like these the data reduction approach is chosen a priori to be either strictly automated or strictly manual. Usually manual approaches are inexpensive to develop but the recurring labor costs are expensive. Automated systems are usually inexpensive to run but only perform well when conditions meet expectations. In general the marginal cost becomes progressively more expensive for each unit increase in performance from an automated system. A robust automated system can prove to be quite expensive and may still yield a non-negligible error rate. So to ensure top quality data even the best automated system would need a human in the loop to validate the results. Herein lies the key methodological insight of our work: if the human is already in the loop to validate the data, provided care has been taken to develop the right time-efficient tools for the user, the marginal costs to have this human to actively clean the results should be small compared to the savings that can be realized in the automated system. If done right, a "pretty good" automated system to do the majority of the processing followed by supplemental manual cleaning (i.e., over and above simple validation) can produce a high quality data set that is beyond the capabilities of a "superior" automated system on its own while only encumbering a fraction of the labor costs from a "purely manual" approach.[2] The generic approach to seek a cost effective balance between the power of the automated system and demands on the *human in the loop* transcends the particular application to instrumented probe vehicle data.

This generic methodological approach allows us to overcome many of the constraints that have limited previous efforts in the specific area of collecting empirical microscopic vehicle interaction data. Historically automatic data extraction has not provided the precision necessary to advance traffic flow theory[3], while the labor demands of manual data extraction have limited past studies to small scales[4].

The data for this study come from a probe vehicle that is instrumented with positioning sensors for localization and ranging sensors to perceive the ambient vehicles and measure inter-vehicle relationships over time and space as the probe travels through the traffic stream. Specifically, this work develops the process of extracting vehicle trajectories from data collected

---

[1] As of publication the data have been posted to Coifman (2016).

[2] Since the automated system used in our data reduction is unique to the specific raw data set and this raw data set is of finite size, it does not make sense to find the optimally efficient balance. So throughout this paper we use "pretty good" to clearly denote that while there is almost certainly a "better" approach, the chosen automated approach is good enough to ensure that indeed the subsequent manual cleaning is far less demanding than a purely manual approach.

[3] For example, the unrealistic relationships in the NGSIM data set discussed in Section 1.1 that arose from the automated processing with only cursory manual validation, or the tracking errors exhibited by a more sophisticated vehicle tracking algorithm in Fig. 6 of Coifman et al. (1998).

[4] For example, Treiterer and Myers (1974) took several years to manually track roughly 70 vehicles over 4 minutes and 3.3 miles.

by three perception sensors, consisting of a pair of forward and rearward facing LIDAR sensors that each scan an arc of 180° in a plane parallel to the roadway and a single forward facing radar sensor that has greater range but only a 12° field of view. The automated processing consists of segmenting vehicle and non-vehicle returns in the LIDAR data, grouping the vehicle returns into distinct vehicles, tracking vehicles over time, and reporting the vehicle tracks in world coordinates. The radar sensor already tracks vehicles and only reports tracked targets, eliminating the post-processing based tracking for the radar data. In either case numerous errors remain in the automatically tracked target data that need to be cleaned by the human reviewer. The study data were collected over roughly two hours on a freeway during a typical weekday afternoon rush hour that includes recurring congestion. The probe vehicle's tour consists of two round trips over a 28 mi route along urban freeways. This tour passes through four separate major freeway interchanges involving: I-70, I-71, I-270, I-670, and SR-315; as well as many more interchanges with arterial roadways.

To place the work in context, the bold curve in Fig. 1a shows a hypothetical example of a trajectory from a conventional probe vehicle. It includes the speed and acceleration of the conventional probe vehicle, but no information about the environmental stimuli that the vehicle responded to, e.g., the unobserved trajectories of the other vehicles that are shown with faint curves. With the addition of the perception sensors on the probe, as per the current study, it is now possible to also collect the trajectories of the leading and following vehicles too, giving rise to the three tracked trajectories in the lane of travel, shown with bold curves in Fig. 1c. Thereby yielding a lengthy sample of the instrumented probe vehicle's response to its leader and concurrent shorter samples of many different followers behind the probe vehicle. This plot only shows the probe vehicle's lane of travel, the perception sensors are also used to track vehicles in the adjacent lanes, typically yielding at least two trajectories per immediately adjacent lane and often several trajectories in further lanes (but the far lanes are often occluded by vehicles in the intervening lanes, so coverage of the far lanes is lower than the immediately adjacent lanes).

This data collection is in response to a longstanding need for empirical microscopic traffic flow data, e.g., Haight (1963) observed that most of the literature on traffic flow theory was conceived in purely mathematical terms, with observational studies limited to supporting particular theories. While exceptions exist both then and now, most contemporary traffic flow theory is still built upon models that ultimately have purely mathematical origins, e.g., hydrodynamic models (Lighthill and Whitham, 1955; Richards, 1956) and car following models (Chandler et al., 1958; Gazis et al., 1961). Even though modern models are much more sophisticated than those when Haight made his observations over 50 years ago, the field remains limited by the quantity and quality of empirical traffic data. Plausible but incorrect hypotheses perpetuate in the absence of accurate empirical microscopic data for model development. While there have been a handful of true empirical microscopic freeway data sets collected, a great need remains for more data in general and larger varieties of conditions in particular. This need remains because the collection and reduction are extremely expensive, which in turn constrains the length of roadway monitored, the location observed, and duration of the study.

In the past the specific study location had to be chosen before any microscopic data were collected or analyzed. In many cases the choice of the study site was simply determined by the proximity of a convenient high vantage point (e.g., a building) to observe the roadway. Given the fact that much remains unknown about microscopic vehicle interactions it is impossible to know a priori whether a specific stretch of roadway contains the entire process of interest. For example, most of the prior studies were collected within the queue, strictly upstream of an active

bottleneck.[5] As a result, these data sets preclude study of the actual bottleneck mechanism that gave rise to the queue formation in the first place. In contrast, the data for this study comes from a corridor that is many miles long- including both upstream and downstream of several bottlenecks. The extended distance allows for subsequent microscopic analysis to select the study segment post hoc while also allowing the exploration of any presently unknown influencing factors that may prove to extend beyond the initially selected segment. For many phenomenon of interest it is quite likely that the study segment spans the microscopic vehicle interactions over several miles of freeway; e.g., Cassidy and Bertini (1999), Coifman and Kim (2011) and Kim and Coifman (2013) provide evidence that the nucleating bottleneck might occur at a location different than where macroscopic data show queuing, as has dominated the focus of previous bottleneck studies. Seemingly innocuous events like driver relaxation a mile downstream of an on-ramp could trigger huge delays upstream of the ramp.

As noted by Hurdle and Datta, 1983; Hall and Agyemang-Duah, 1991, and Cassidy and Bertini, 1999, upon finding an apparent bottleneck, it is still necessary to extend surveillance beyond the segment to make sure there are no further factors further downstream that contribute to congestion and queue formation. With the extended corridor used in this study the *look beyond* is already included in the data. Of course the extended surveillance distance comes at the cost of limited temporal coverage of any given location on a given pass, but this detraction is countered by two passes in the current data extraction (and ultimately over a hundred passes in the complete data set, as discussed in Section 4). Needless to say, there remains a need for more microscopic data in general, and in particular, from a corridor that extends many miles and comes from a much longer time span so that spatial interactions can be revealed and variability can be measured.

The remainder of this section reviews the history of empirical freeway traffic flow studies to place the work in context. Section 2 presents the methodology, starting with a review of the instrumentation on the probe vehicle and the route taken by the probe vehicle, then continuing into the process of extracting information from the raw data. Section 3 presents the results. The report closes in Section 4 with conclusions, a discussion of the limitations.

1.1. A brief history of empirical freeway traffic flow studies

Traffic flow theory is approaching the limits of conventional detector data. Most of what we know about traffic flow ultimately comes from macroscopic data collected by point detectors, which inherently has limited temporal and spatial resolution. Loop detectors (and other point detectors) are capable of monitoring traffic at fixed locations, but point detectors can only monitor signals and waves in the traffic stream that propagate past the detectors, e.g., the bold portions of the individual trajectories in Fig. 1b as the vehicles cross two short regions representing different detector stations. Conventionally the point detector data are aggregated to average speed, flow and occupancy over 30 sec sampling periods (or longer) by operating agencies to monitor the network in real time; thus, discarding the limited information of any given individual vehicle as it passed the detector. These agencies simply do not need the fine precision that is required for traffic flow studies. As dictated by the Nyquist sampling criterion

---

[5] Bottlenecks are the locations where vehicular flow on a road network is most constrained, triggering the queues that cause the delays throughout the network. A bottleneck is said to be "active" when it is limiting traffic flow. Queuing traffic is merely a symptom of an active bottleneck; any demand above the bottleneck capacity is delayed and forms a queue that grows upstream from the bottleneck location. On freeways these queues can stretch for miles and impact trips that do not even pass the bottleneck location.

one can only resolve features in the data that last at least twice as long as the sampling period. Given the fact that one headway is on the order of a few seconds, it should be clear that the current conventional sampling periods on the order of 30 sec have insufficient resolution to understand the microscopic details of traffic flow. There have been several recent advances into microscopic phenomena using novel data processing on the individual vehicle actuations from dual loop detectors. This so called Single Vehicle Passage (SVP) methodology (Coifman, 2014) has shown that on average signals propagate through longer vehicles faster than shorter vehicles (Coifman, 2015) and that car following behavior also depends on the relative speed to the adjacent lane (Ponnu and Coifman, 2015). However, these loop detector based techniques cannot provide insight into the nuances of any specific vehicle's behavior, and any given individual's behavior can differ greatly from the center of the distribution that is measured by the SVP methodology.

One of the biggest challenges facing the traffic flow community is the lack of sufficiently accurate empirical, microscopic, in situ data for the detailed analysis of vehicle interactions to understand the nuances of traffic flow leading to empirically based model development, calibration, and validation. The truth of the matter is that collecting such data is very challenging, and as a result, empirical advances of microscopic traffic flow theory (including car following behavior) are very slow in coming. By the early 2000's the traffic flow theory community recognized the need for accurate empirical microscopic data, which lead to the collection of the Next Generation Simulation (NGSIM) data sets (Kovvali et al., 2007). There are two NGSIM freeway data sets: I-80 with 5,678 vehicles collected over 45 min during the congested evening commute across 0.33 mi of freeway containing a single on-ramp; and US-101 with 6,101 vehicles collected over 45 min during the congested morning commute across 0.42 mi of freeway containing one on-ramp and one off-ramp. Since the NGSIM data were released ten years ago the number of empirical microscopic traffic studies has exploded and the NGSIM data now form the basis for the vast majority of recent advances in empirical microscopic traffic flow theory. While the NGSIM data sets are considerably larger than any microscopic data sets that came before, the research community has complained that there are only two locations, each with: less than an hour of data, collected on a single day, only spanning short distance, and the surveillance region was strictly within the queue upstream of an active bottleneck. Although NGSIM delivered on the quantity of data, there is a growing minority of researchers who have found unrealistic relationships in the NGSIM data and now question the accuracy of the NGSIM trajectories, e.g., Duret et al. (2008); Hamdar and Mahmassani (2008); Thiemann et al. (2008); Punzo et al. (2011); Montanino and Punzo (2015).

The data reduction used in the NGSIM study is not new. The process of extracting information from orthorectified imagery (video, movies, or high frame rate photography) recorded from a high vantage point was fairly common in the 1960's and 70's, relying on labor-intensive data reduction techniques, e.g., Forbes and Simpson (1968); Treiterer and Myers (1974). None of these early data sets are known to remain. Turner-Fairbank Highway Research Center (TFHRC) collected data at one frame per second from 18 locations in 1983 and used microcomputers to expedite the data reduction process (Smith, 1985; Smith and Mark, 1985). The data were released in 1985 and were distributed by request on 9-track magnetic tape. Because the data sets were difficult to access they have seen little circulation.[6] Unfortunately,

---

[6] A portion of the TFHRC data set has been preserved and as of this writing it is available from the TRB Committee on Traffic Flow Theory and Characteristics, at http://tft.ceng.usf.edu/docs.htm

most of the TFHRC locations exhibited little or no queuing, thus limiting the benefits for detailed study of congested traffic flow. The TFHRC data sets also share some of the detractions of NGSIM: small distance monitored, short duration of time, and small number of ramps at each the study location. Other recent examples using trajectory extraction from orthorectified imagery include Becker (1989) to demonstrate computer based trajectory extraction from aerial video, Coifman et al. (1998) to manually generate ground truth data to validate a video image processing based vehicle tracker, and Xin et al. (2008) to study vehicular crashes. Meanwhile, the various limitations of the NGSIM study have led to several smaller efforts to extract vehicle trajectories using similar high vantage point filming techniques (e.g., Daamen et al., 2010; Knoop et al., 2012; Marczak et al., 2014a; Marczak et al., 2014b).

Of course not all empirical microscopic traffic data are derived from orthorectified imagery. Instrumented probe vehicles have been used to study aspects of microscopic traffic flow since the 1950's, e.g., Chandler et al. (1958), Herman and Potts, (1959), Brackstone et al. (2009), and Schorr et al. (2016). In these studies the subjects typically drive a dedicated study vehicle for several hours and the studies primarily focus strictly on the driver and vehicle, without accounting for external factors that influence driver behavior, e.g., proximity to ramps or bottlenecks. The study sets are typically small and they are rarely shared with the research community at large. There is a notable group of recent studies that focus on naturalistic driving (ND) and typically employ instrumented vehicles to monitor driver behavior in situ to study safety and the precursors for accidents, e.g., Barickman and Goodman (1999), Lee et al. (2004), Dingus et al. (2006), Regan et al. (2013), Blatt et al. (2014), Eenink et al. (2014). These ND studies typically have test subjects drive an instrumented vehicle for a period of weeks to months. The ND studies are opportunistic, collecting data from drivers wherever they happen to travel, thus limiting the number of observations of a given location. While most ND studies contemplated the safety issues, of the major studies reported in the literature only Dingus et al. (2006) had extensive measurement of distance to vehicles in the neighboring lanes and that represented only 10% of the data collected. Given the great detail of personal information involved, there are very strict guidelines to limit access to the data sets, they literally remain behind locked doors in secure facilities. This necessary security severely limits what can be studied and often creates insurmountable barriers to independent validation of any research findings. Nonetheless, some of the researchers affiliated with the ND studies have leveraged the data for microscopic car following studies, e.g., Chong et al. (2013) and Sangster et al. (2013).

2. Methodology

This study uses data from front and rearward horizontal scanning LIDAR sensors mounted on an instrumented probe vehicle (or henceforth, simply LIDAR data) and a single forward mounted radar sensor to extract ambient vehicle trajectories from the sensor data. In each case the sensor data are processed using the approach laid out in Section 1, specifically a good but imperfect automated tracking process followed by manual cleaning to catch and correct the inevitable errors from the automated processing to provide accurate vehicle trajectory data.

The two perception sensors have different characteristics with the LIDAR providing a wide angle near field of view and the radar providing a narrower angle far field of view. The raw LIDAR data report the distance to a target at half degree increments over a 180° arc. The returns should have very low measurement uncertainty, at each scan point there is either a valid distance measurement accurate to a few cm or no return is reported. So it is necessary to develop the

algorithms to segment and track vehicles within the LIDAR point cloud. With LIDAR data like this, most of the tracking errors arise when: classifying a given return as either a vehicle or non-vehicle target, clustering vehicle returns into distinct vehicles, and recognizing when some of the target vehicle is not observed. On the other hand the radar sensor is a black box, it only reports tracked targets, effectively providing the automated processing in real time. In either case the automated processing exhibits vehicle classification and tracking errors that are subsequently addressed in the manual cleaning process.

The suite of sensors on the instrumented probe vehicle is a unique combination and the collected data sets are finite in size, so the specific details of the automated processing is not likely transferrable, but in addition to explicitly having a manual correction step some of the concepts employed will be of interest to the community at large. Section 2.1 reviews the instrumentation on the probe vehicle and the route taken. Section 2.2 presents the positioning and speed measurement for the probe vehicle itself. Then Section 2.3 and 2.4 respectively present the automated LIDAR and radar processing. While it is impossible to know a priori which target is correctly tracked, a human can quickly assess the situation with the aid of the concurrent video and correctly clean the automated tracking results. So rather than attempt to teach the computer to handle all of the unusual exceptions, Section 2.5 develops a suite of tools to allow a human to quickly review and clean the results from the automated processing.

2.1. The instrumented probe vehicle and tour route

Over ten years ago our group began instrumenting a probe vehicle with numerous positioning and ranging sensors to monitor the ambient traffic around it. The instrumented vehicle is equipped with positioning sensors (DGPS and inertial navigation) to track its location and ranging sensors (LIDAR and Radar) to monitor nearby vehicles in the ambient traffic. Fig. 2 shows the fully instrumented probe vehicle. Out of these sensors, the present work uses the forward and rearward LIDAR that scan 180° in a plane parallel to the ground at 37 Hz and returns the distance to the closest object at 0.5° increments. The range of the LIDAR sensors is approximately 80 m, with a resolution of 0.25 cm. The forward facing radar sensor automatically tracks targets over a narrower field of view (roughly 12° angle of coverage), its range extends to 120 m. The radar angle of coverage spans 3 lanes at 40 m and over 6 lanes at 80 m. Thus, the combination of the two data sources provides a much more complete view ahead of the probe vehicle than either taken alone. Unfortunately there was no radar sensor on the rear. For data cleaning and validation this work also uses the front and rear facing cameras that each capture a video stream with a resolution of 320 * 240 pixels at a frequency of 10 Hz.

When first equipping the instrumented probe vehicle it was envisioned that extracting vehicle trajectories from the horizontal scanning LIDAR data would be far simpler than video based systems. While the LIDAR virtually eliminates positioning errors that arise in image processing, as it turned out the challenges of automatically detecting, segmenting and tracking other vehicles in the horizontal scanning LIDAR data were far greater than anticipated, e.g., the top down view in Fig. 3b. While a few pilot studies successfully demonstrated the feasibility of automatically tracking other vehicles in the LIDAR data (e.g., Gao and Coifman, 2006; Xuan and Coifman, 2012) they did not produce a sufficiently robust algorithm for doing so large scale given the magnitude of the grouping and segmentation challenges. Hence, the current study that couples automatic tracking with manual cleaning. In the mean time the vertical scanning LIDAR data from the side sensors shown in Fig. 2 were used for several studies (Lee and Coifman, 2012, 2015; Thornton et al., 2014) and even the conventional GPS data from the probe vehicle was

used in several studies (Coifman and Krishnamurthy, 2007; Tong et al., 2009; Coifman and Kim, 2011).

While the current study only uses a single day of data, collected on September 9, 2009, the instrumented probe vehicle collected data on a 28 mi route (14 mi each way) along I-71 in Columbus, Ohio during morning and evening rush hours on various weekdays from 2004 to 2011. Fig. 4 shows the route that spans four major freeway interchanges, several dozen ramps, and a number of recurring bottlenecks. Each tour consisted of 2 or 3 passes, i.e., complete round trips, over a roughly 2 hr period, yielding multiple observations at a given location in the given day. The existing instrumented probe vehicle data from the freeway include: 350 tours (700 hrs), from 18,500 mi of travel by the probe vehicle in the corridor collected over the span of six years. Interestingly, it turns out that a portion of the route along southbound I-71 was also the subject of Treiterer and Myers (1974).

2.2. Establishing a reference lane and probe vehicle positioning

The instrumented probe vehicle is equipped with a DGPS sensor, a high accuracy yaw gyroscope and an OBD vehicle interface providing data on the current speed and other vehicular information (the latter two data sources allow for dead reckoning). The probe vehicle position data come from a fusion of the DGPS and dead reckoning reports[7], these data are then processed to derive the probe vehicle's speed, v, and acceleration, a.

The pre-defined route traverses three different freeways: SR-315, I-70, and I-71, as shown in Fig. 4. To achieve lane level accuracy over this route without the aid of a preexisting map that is accurate to the lane level, this work instead uses multiple passes to deduce the lane location. That is to say, while the lane configuration is known, the exact coordinates of the lanes are initially unknown. Fig. 5a shows an example of the vehicle trajectory in two dimensions from a single pass, but even when the lane of travel is known, lane change maneuvers and transient positioning errors may bring the probe vehicle's location out of a specific lane. So using a method similar to Xuan and Coifman (2012) the probe vehicle's positioning data from 40 tours (80 passes) are combined[8] on a single plane (Fig. 5b). Then the centerline of the probe vehicle's dominant lane of travel at each directional location along the route is found by taking the median laterally across of all passes at a given longitudinal location (Fig. 5c) and finally the adjacent lanes are identified by shifting an integer number of lane widths (12 ft), as shown in Fig. 5d. Distance along the route is calculated from the cumulative distance traveled along the dominant lane at each location, averaged across all passes out of the 80 that actually traveled within that lane at the given location. This distance along the road is then used even when the dominant lane of travel differs from probe vehicle's lane of travel in a particular pass. Obviously there will be small differences in the actual distance due to road curvature (both across lanes, and even within lane depending on the lateral position within the lane), so the lane configuration is also recorded in two dimensions (northing and easting) in case more precision is needed. In any event, this reference distance is only used for position information as a common reference across different passes. It is not used to measure speed or acceleration, which are measured directly from the recorded position and speed data for the given pass.

---

[7] The details of the localization data and the data fusion process will be reported in forthcoming publications.
[8] This task is the only one in this report that uses more than a single day of data, and the multiple days are only used in this step to control for the possibility of any transient out of lane positioning when establishing the reference lane.

There are several known mandatory lane change maneuvers along the tour and there is no single lane that persists throughout the route. So this work creates a hypothetical continuous reference lane in each direction from which all other positions are measured and this lane is called lane X[9], i.e., the longitudinal reference distance noted above is assigned to lane X and lateral distance across the road is also specified relative to the reference lane (both in meters and in terms of the discrete number of lanes). For naming purposes all vehicular lanes of travel are specified relative to lane X with values increasing from the passing lane on the left to the slow lane on the right, e.g., lane XIII is three lanes to the right of lane X. This reference persists even when lane X does not physically exist at a given location, e.g., the probe vehicle might have to take a mandatory lane change maneuver from lane X to lane XI before the physical lane X diverges, but the lane naming scheme does not change after the diverge and the instrumented probe vehicle will remain in lane XI until undertaking its next lane change maneuver. The complete lane configuration is recorded as a function of longitudinal distance along the route.

2.3. Automated detection and tracking of ambient vehicles in the LIDAR data

The automated LIDAR processing is broken into three steps: finding vehicles within a single LIDAR scan, tracking vehicles across successive LIDAR scans, and associating vehicles between the front and rear whenever there is an overtaking of/by the instrumented probe vehicle.

2.3.1. Processing each LIDAR scan individually

The first stage of processing the LIDAR data is to take each LIDAR scan individually to group returns into clusters and then classify each cluster as being "vehicle" or "non-vehicle". The key distinction is that vehicle clusters are known to be vehicles while the non-vehicle clusters could be objects outside the right-of-way, non-vehicle returns when the probe vehicle pitches forward due to braking and the front LIDAR sensor scans the road surface instead of any vehicles, and un-trackable vehicles that are partially occluded. This work takes a three-step process to group the LIDAR returns into clusters, as follows: (1) grouping angularly adjacent points into discrete clusters, (2) assigning clusters to discrete lanes, and (3) consolidating clusters.

The clustering is done in the original LIDAR coordinate system that moves with the probe vehicle. Fig. 6a shows an example of a single LIDAR scan where the x and y values are relative to the probe vehicle with (0,0) at the given LIDAR sensor unit. There are 361 returns in each LIDAR scan, corresponding to each 0.5° increment between -90° and 90°, with each return containing the range to the target (up to 80 m) or an indication that there was *no return*. The "no return" occurs either because there was no target before 80 m or if there was an intervening target the return was either too weak or too noisy to be resolved by the LIDAR sensor. The grouping algorithm steps from one angle to the next in a single LIDAR scan and tests whether the scans from the two successive angles are close enough to be grouped together into a cluster. Vehicle objects[10] will generally appear either as: a longitudinal line for a vehicle immediately adjacent to the probe vehicle (e.g., the cluster at 0<y<4 m in the right lane of Fig. 6b); an "L"

---

[9] Roman numerals are used to denote lanes relative to the reference lane to avoid potential confusion with any local lane numbering, as will be the case in subsequent work that adds concurrent loop detector stations, and the reference lane number is arbitrarily set to X so that positive numbers IX to I can be used for lanes to the left of X.

[10] The term "object" is used to describe a physical thing that is indeed a discrete object. An object may or may not be seen in a LIDAR scan. If an object is correctly identified there will be a unique cluster of LIDAR returns (usually spanning successive angles) that corresponds to the object.

shaped corner for a vehicle ahead/behind of the probe vehicle in an adjacent lane (e.g., the cluster at 5<y<10 m in the right lane of Fig. 6b); or a lateral line for a vehicle immediately ahead/behind of the probe in the same lane (e.g., the cluster at y=8 m in the center lane of Fig. 6b).

Given the probe vehicle's longitudinal distance along the route the current lane information from Section 2.2 is retrieved, projected into the probe vehicle's coordinate system and the centerlines of the lanes are found in the LIDAR coordinate system, e.g., as shown with dashed curves in Fig. 6b. At this stage all resolvable clusters that are within the demarcated roadway are retained as being on-road and assigned to a given lane while the off-road clusters are discarded as non-vehicle returns.

While the clustering seeks to achieve a one-to-one matching between objects in the world and clusters in the given scan, sometimes a single object will be over-segmented into multiple clusters (e.g., because portions of a high-clearance vehicle are not visible as a contiguous object in the LIDAR scanning plane) or several different objects will be over-grouped into a single cluster (e.g., because two separate vehicles are traveling very close together). The processing is tuned to over-segment long vehicles at this stage, which is a trade-off to minimize over-grouping vehicle clusters.[11] The rectangular bounding boxes in Fig. 6b show the identified on-road clusters for a typical scan at the conclusion of this clustering process (in this case the raw LIDAR data are shown with dark points in Fig. 6a and lighter points in Fig. 6b). Many of the automated clustering errors will be caught in the later stages of the automated processing while the subsequent manual cleaning discussed in Section 2.5 is employed to deal with the more complex scenarios that may confound the automated processing.

2.3.2. Processing across successive LIDAR scans

After each LIDAR scan is processed the discrete targets in successive scans are associated with one another to form a vehicle track. While the shape and appearance of the clusters associated with a given target vehicle might change from one scan to the next, the position of the *nearest corner* of the bounding boxes from a given vehicle's clusters in successive scans should evolve slowly. One key issue is that if a return at a given scan angle is too weak the LIDAR sensors will report "no return" even if there is an object closer than the maximum range of 80 m. This situation typically arises for targets further than 40 m away from the sensor and as a result, far target vehicles can *flicker* in and out of the LIDAR returns from one scan to the next. So the tracking process also tolerates a target disappearing for a short period of time. Whenever a tracked target disappears for longer than 1 sec the vehicle track is closed and is assigned a unique *target ID* number. Additionally, the target's location in global coordinates and cumulative distance of the target in each LIDAR scan is also calculated from the combination of the relative location to the probe vehicle and the probe vehicle's location in the global coordinate system. Finally, in the event that a target vehicle is even with the LIDAR sensor (longitudinal distance 0 m), there is a good chance that the nearest corner is unobserved, with only the nearest side of the target seen. The visible portion of the target is retained, but the position is marked as an "occluded end" to clearly indicate that the exact longitudinal location is unknown and that the target might simultaneously appear in the other LIDAR.

---

[11] This preference to over-segmenting vehicles is chosen because it is a lot easier to manually join over-segmented clusters than it is to manually segment over-grouped clusters.

2.3.3. Associating targets between the front and rear LIDAR sensors

Up to this point the front and rear LIDAR data are stored and processed separately. This step associates vehicle objects that are seen in both the front and rear LIDAR scans. Whenever the instrumented probe vehicle overtakes a shorter vehicle, as that target disappears from the front LIDAR scans it should soon appear in the rear LIDAR scans (or vice versa for a vehicle overtaking the probe vehicle). So this step seeks to identify and group those tracks belonging to a given vehicle that were seen separately in the two LIDAR views. During these overtaking maneuvers if the target vehicle is longer than the probe vehicle it should briefly be visible in both the front and rear LIDAR scans, or if it is shorter it should briefly disappear from both. In any event, except for very short vehicles (e.g., motorcycles) there is not enough room for more than one target vehicle per lane to *disappear* in the *unseen area* next to the probe vehicle. So when a target in an immediately adjacent lane disappears completely next to the probe vehicle due to overtaking, it will be associated with the next emergence on the same side in either LIDAR. Recall that the prior step specifically marked any target that ends (starts) with an occluded cluster next to the probe vehicle, and if so, except for very rare lane change events there has to be a corresponding trajectory that starts (ends) with an occluded cluster on the same side of the probe vehicle. The time of the occluded ends of the two corresponding trajectories usually are very close; thus, allowing the automated process to group same-target trajectories with occluded ends and link them together. In the event that a long enough period of time elapses the automated processing will not associate the two occluded ends, in which case the two observations of the same vehicle will retain their separate, distinct tracks and target IDs at this stage. In fact a target vehicle can bounce back and forth between the two LIDAR sensors when it is traveling at a speed close to that of the probe vehicle; as such, each new appearance is recursively associated with the previous observations of the given target vehicle. This process ends once the chain of linked tracks starts and ends without an occluded end. All but one of the target IDs associated with this chain of observations are discarded and the remaining target ID is assigned to all of the clusters that were associated in the chain.

2.4. Automated detection and tracking of ambient vehicles in the radar data

As noted previously, the radar already tracks targets within its coordinate system and only reports the tracked targets. Still there are non-vehicle targets reported by the radar sensor that need to be eliminated and over-segmented vehicles that need to be combined. Furthermore, the radar targets must still be associated with their respective lanes and it is necessary to establish the world coordinates of the tracked vehicles. This processing follows the same steps laid out in Section 2.3 for the LIDAR data, with the exceptions that there is no need to group clusters or track the targets as reported by the radar sensor, and instead of associating front and rear targets this task associates the radar targets with the forward LIDAR targets.

2.5. Manually cleaning the automatically tracked vehicles

Even if the automated system is expected to be highly accurate all results need to be manually validated to ensure the data quality necessary for developing traffic flow theory. Since the human already has to be in the loop, this work seeks to leverage the discerning abilities of that human reviewer to quickly make any necessary corrections and reduce the required complexity of the automated processing. So instead of employing a "superior" automated system, having the human clean up the data allows the use of a "pretty good" automated system to do the majority of the processing followed by supplemental manual cleaning. Thereby finding a cost

effective balance between the power of the automated system and demands on the *human in the loop* to yield a high quality data set that is beyond the capabilities of a "superior" automated system while only encumbering a fraction of the labor costs from a "purely manual" approach. The details of these steps are summarized as follows.

The instrumented probe vehicle is equipped with 10 Hz video cameras facing forward and rear, concurrent with the two LIDAR sensors. The video feeds are time stamped with the same clock used for the LIDAR, radar, and localization sensors. To facilitate comparison between the LIDAR and video, it is necessary to project the LIDAR data into the video plane using a homography transformation that is calculated using conventional approaches. With the LIDAR data projected into the video image plane for both the front and rear views the human user can quickly and accurately assess the automated tracking errors. In most cases the human can correct the problems (e.g., grouping over-segmented clusters on a frame by frame basis), at other times the human can at least exclude any problematic areas (e.g., when the LIDAR scanning plane strikes the road surface).

The design of a graphical user interface (GUI) tool is key to enabling efficient manual validation by a human reviewer. Fig. 7 is an example of the main GUI display at a typical time step. Fig. 7a shows a top down view of the two concurrent LIDAR scans combined into a single view, with the forward LIDAR at the top, rear LIDAR at the bottom and the probe vehicle at the center of the image. The LIDAR sensors are each located at the origin of their respective portion of the plot (note the two sets of labels on the vertical axis) with the gap between the two regions corresponding to the distance between the front and rear LIDAR sensors on the probe vehicle. The raw LIDAR returns are shown with points and numbered bounding boxes are shown around the various targets being tracked. To the trained eye many of the automated tracking errors become evident in this plane, without referring to the video. Fig. 7b-c show the concurrent front and rear video with the LIDAR targets superimposed. Fig. 7e-g show separately the measured trajectories over time, in this case with lane XI to the left, lane XII the probe vehicle's lane of travel, and lane XIII to the right (the bold curve in each plot is the probe vehicle's trajectory, shown only for reference as a dashed curve in lanes XI & XIII). By simultaneously reviewing all of these plots the human reviewer can spot inconsistent trajectories on the right hand set of plots, incorrectly grouped clusters on the left hand set of plots, and the correspondence between the LIDAR and video in the middle plots. Note that the clusters and trajectories in Fig. 7 illustrate the ambient vehicle data near the end of the manual validation, so few of the automated tracking errors remain in this example. Not shown in the figure are the actual navigation and data cleaning controls that are implemented with various virtual buttons below the graphical display. The three most significant tools for data cleaning involve: (i) eliminating any remaining non-vehicle returns, (ii) merging clusters from over-segmented vehicles and splitting clusters from over-grouped vehicles, and (iii) connecting multiple partial trajectories that come from a single vehicle. The navigation tools include the option to move forward and backwards in time, *playing back* the concurrent temporal evolution in all seven panels of Fig. 7.

While the functionality of the user interface continues to evolve, our current estimate is that it should take roughly 8 person-hrs to do the manual cleaning for each 2 hr tour. It is also worth noting that although this work sought to maintain a given target ID across short disruptions lasting up to 1 sec, there was little effort to maintain a unique ID across longer gaps. So a vehicle that is visible within the video field of view may receive several target ID's if it goes unseen by the perception sensors for an extended period. Likewise, there are a few distinct

vehicles in the validation video that appear to enter and leave the camera's field of view during periods of stop and go traffic. No effort was made to associate vehicles across such large gaps.

3. Results

As noted earlier, the current work extracted and manually validated roughly two hours of data from the ambient traffic as the instrumented probe vehicle made two passes over the 28 mile round trip on I-71 in Columbus, Ohio, (Fig. 4) during the evening peak on a typical day (September 9, 2009 in this case). The particular tour includes both recurring congestion and unrestricted conditions. Figure 8 shows the resulting speed spacing relationships for the instrumented probe vehicle on each of the four passes, two northbound and two southbound. Almost all of the congested traffic is seen when the instrumented probe vehicle is traveling away from the central business district (CBD), as would be expected during the evening rush hour.

Fig. 9 shows a portion of the first pass on I-71 northbound through two closely spaced freeway interchanges. Traffic moves from bottom to top and the schematic in Fig. 9d shows that traffic enters I-71 in the bottom right of the figure via a connector ramp from I-70 westbound and exits I-71 on the top left of the figure via a connector ramp to I-670 westbound. The three time space diagrams show the ambient vehicles in the probe vehicle's lane of travel, XI, and the two immediately adjacent lanes. Like Fig. 7, the probe vehicle's trajectory is shown with a bold solid line in the lane of travel and a bold dashed line in the other two lanes for reference. Ambient lane change maneuvers are highlighted with triangles pointing the direction of the maneuver (the empty triangles denote the exited lane and solid triangles denote the entered lane). In general the lane changing vehicles are moving to the left in this section as several of the vehicles entering from I-70 proceed to a subsequent exit to I-670, including two vehicles that cross two lanes in rapid succession. Also note the slow wave passing in lane XII.

Fig. 10 picks up roughly 0.5 km downstream of Fig. 9 using the same notation, with the lane numbering, time and distance consistent to the same global reference in both figures. The schematic in Fig. 10d shows that there is a connector ramp to I-670 eastbound on the right and an entrance ramp from Broad St. on the left. The probe vehicle starts in lane XI and changes to lane XII within the segment. The segment has rapidly changing traffic patterns, at the upstream end lane XIII on the right primarily has vehicles heading to the I-670 connector ramp and the lane changing patterns show a general movement to the right, with additional vehicles changing lanes to reach the exit to I-670, including one vehicle that crosses two lanes in rapid succession. This vehicle entered the segment immediately behind the instrumented probe vehicle then changes lane to the right as the probe vehicle encounters a slow wave, that vehicle then rapidly approaches a queue of nearly stopped traffic right before making the second lane change maneuver with a very small spacing to the last vehicle in the queue. Then at the head of the queue in lane XII (roughly 7,280 m) from the concurrent video it is clear that there is a box truck almost stopped in the middle of the road. Although beyond the range of the LIDAR, just before this truck disappears from the rear view camera it is apparent that the truck moves across lane XIII to get to the I-670 connector ramp. As the probe vehicle approaches the I-670 connector ramp, lane XIII is occluded by the stopped vehicles in lane XII and once past the queue, very few vehicles remain in lane XIII because most have taken the ramp to I-670 E. Meanwhile, several vehicles take advantage of the large gap ahead of the stopped truck and the fact that lane XIII is now nearly empty. In short order four vehicles (including the probe vehicle itself) move from lane XI to lane XII, in part to avoid an upcoming queue in lane XI arising from the combination

of the left-side on-ramp located at the very end of the link and a recurring bottleneck further downstream.

At the large vertical range shown in Fig. 9-10 it is difficult to see the subtle vehicle interactions. Fig. 11 shows 1.5 min of data during free flow conditions. This time the ambient target vehicle locations are shown in the probe vehicle's coordinate system to facilitate the display of the inter-vehicle relationships. The relative longitudinal distance is set to zero at the rear of the instrumented probe vehicle with the rear view LIDAR data below and the front view sensor data above, offset by the probe vehicle's length. In this case the probe vehicle starts in the center of three lanes, moves to the left lane at time 0 in the plot, overtakes a slow moving semi-trailer truck (veh 21) and SUV (veh 22), and then returns to the original lane roughly 30 sec later, as shown with vertical thin dashed lines. For this example the tracked vehicles from center lane are labeled 2• and the vehicles in the passing lane are labeled 1•. The LIDAR vehicle positions are shown with bold curves and the radar positions with thinner curves. The probe vehicle's speed is shown with a bold dashed curve, relative to the same vertical scale. The instrumented probe vehicle is initially behind veh 21[12], changes lanes at 0 sec and then accelerates to match the speed of its new leader, veh 11, that is roughly 80 m ahead, while the new follower, veh 12, slows down and creates a larger spacing in response to the instrumented probe vehicle's entrance. At about 5 sec the probe starts to overtake the semi-trailer truck. Note that the forward sensors see the rear of a given target while the rear LIDAR sees the front of a target, so the trajectory initially from the rear of the semi-trailer truck seemingly shifts to the right as its front subsequently emerges behind the probe vehicle. It is possible to measure the length most vehicles that overtake or are overtaken by the probe using the reported data, in which case the coordinates of the front (rear) bumper can be extrapolated from the observed rear (front) bumper location, thereby eliminating these visual discontinuities. The probe then passes veh 22 ahead of the truck before decelerating and returning to the center lane roughly 30 sec after the first lane change maneuver. Just before undertaking this return maneuver veh 25 enters the center lane from the right, roughly 30 m ahead of the probe, and the probe initially slows to increase the spacing to its leader. Ordinarily the sensors can only see one vehicle ahead and behind in the self-lane, but upon returning to the center lane at 30 sec the self-lane veh 21 and 23 remain intermittently visible because the probe vehicle is on a curve and so the closest vehicles do not completely occlude the further vehicles.

4. Discussion and conclusions

This research has developed a process for extracting and cleaning empirical microscopic vehicle relationships by tracking ambient vehicles observed from an instrumented probe vehicle traveling through the freeway traffic stream. In this case the vehicle is equipped with front and rear horizontal scanning LIDAR sensors with near-rage wide angle coverage and a forward facing radar with far-range narrow angle coverage. Given the inherent difficulty collecting empirical microscopic data there are only a few existing data sets for the research community, so this work provides a much-needed empirical data set for further traffic flow theory developments (as of publication the data have been posted to Coifman, 2016).

---

[12] The wobbles in veh 21 trajectory starting at (-30 sec, 40 m) reflect the fact that the radar jumped from one feature to another on veh 21 as it was tracking the vehicle. Although not done in this example, the wobble can be cleaned using a filter to follow the upstream end of the peaks.

Key to this extraction is a very powerful approach that can be transferred to many other data reduction problems. Rather than choosing between an expensive "superior" automated processor with cursory manual validation or labor-intensive manual data processing, recognizing that (i) each successive marginal gain in automated processing becomes progressively more expensive, (ii) the human must still be in the loop to validate the data, (iii) with the right user interface the marginal cost is small to design the system from the start to leverage the discerning abilities of that human reviewer to clean the uncommon complex events that stymie the automated processor, it is possible to reduce the complexity of the automated processing without sacrificing data quality. If done right, a "pretty good" automated system to do the majority of the processing followed by supplemental manual cleaning (i.e., over and above simple validation) can produce a high quality data set that is beyond the capabilities of a "superior" automated system while only encumbering a fraction of the labor costs from a "purely manual" approach.

The present work is an important milestone in a larger instrumented probe vehicle study that began over ten years ago. It is proof that the automated data extraction combined with manual cleaning is an effective means for generating empirical microscopic vehicle interaction data with sufficient precision to be used to advance traffic flow theory, e.g., for model development and calibration. Still, this work is only the first step in a larger data extraction; this pilot study only extracted the vehicle trajectories from the front and rear scanning LIDAR sensors and forward facing radar sensor for a single 2 hr tour. As noted in Section 2.1Y, the overall data collection in the corridor consists of over 300 similar tours collected over a period spanning 2004 to 2011. As these data are extracted and cleaned up they will be shared with the research community, starting with the 2 hr from this report. We anticipate that the data will reveal behavioral phenomena and capture previously unknown influencing factors, e.g., Coifman et al. (2003), Cassidy and Rudjanakanoknad (2005), Laval and Daganzo (2006), Chung et al. (2007), Wang and Coifman (2008), Duret et al. (2010), and Xuan and Coifman (2012).

4.1. Important points and limitations of the current extraction

One of the objectives of this study is to provide microscopic data on vehicle interactions to foster the development of traffic flow theories and as of publication the data have been posted to Coifman (2016). In this context, it is important for consumers to understand how to interpret the data correctly. First off, when working with the extracted vehicle trajectory data if there is no recorded lead vehicle in the data, it does not mean there is no lead vehicle in the real world. When this situation arises in the final extracted data it simply indicates that the lead vehicle was not seen by the perception sensors. In short, the "no leader" data does not provide any information on how the follower responds to the unseen leader. Or to put it another way, the vehicle trajectory data are most valuable when both a leader and a follower are observed in a given lane.

Since the probability that a vehicle goes unseen increases with the distance from the sensor, another issue to keep in mind while working with these data is the nature of the sensors and the potential for systematic sampling bias. For example, the strength of a LIDAR return varies inversely with the distance to the target. If a given return is not strong enough the LIDAR sensor disregards the reading even if the target is within the 80 m range and instead reports "no return" at that scan angle. Thus, it is more likely that distant vehicles go *unseen* due to poor returns compared to closer vehicles. This fact will impact subsequent uses of the data, for example, when plotting speed-spacing from the probe vehicle to the ambient vehicles as per Fig. 8. At low speeds the spacing to ambient vehicles is small enough that few vehicles will go

unseen in the LIDAR data, but at higher speeds the range of true spacings becomes large enough that a disproportionate number of distant vehicles could go unseen, so at a given speed the distribution of recorded spacings is skewed lower than the distribution of true spacings. This bias is less in the forward facing radar, but the radar sensor's angle of view is much narrower than the LIDAR and even though the radar's rage is much larger it is still finite. So consumers of the data are warned to treat periods with no lead vehicle with caution. This caution extends to averaging, e.g., when calculating an empirical speed spacing curve the greater the distance to a lead vehicle the more likely it will not have a return and thus, at a given speed the smaller spacings are likely to be over represented in the recorded data. In some situations it may be possible to compensate for this issue, e.g., often times the vehicle trajectories will *flicker* in and out, allowing for quick identification of periods when there is an unseen leader. Alternatively, for the most precise applications, the user can refer to the concurrent video to identify periods when there is a vehicle that is not recorded in the LIDAR.

References

Barickman, F., Goodman, M. (1999) Micro Das: In-Vehicle Portable Data Acquisition System, Transportation Research Record, Transportation Research Board, Issue 1689, pp 1-8.

Becker, U.J. (1989) Automatic traffic data collection using aerial video images, *Proc. Second Intl. Conf. on Road Traffic Monitoring*, London, UK, 7-9 February 1989, IEE, pp79-83.

Blatt, A., Pierowicz, J., Flanigan, M., Lin, P., Kourtellis, A., Jovanis, P., Jenness, J., Wilaby, M., Campbell, J., Richard, C., Good, D., Czar, N., Hoover, M. (2014) (2014) *Naturalistic Driving Study: Field Data Collection*, SHRP 2 Report, 173p.

Brackstone, M., Waterson, B., McDonald, M. (2009) Determinants of following headway in congested traffic, *Transportation Research Part F: Traffic Psychology and Behaviour*, Volume 12, Issue 2, pp 131-142.

Cassidy, M.J., Bertini, R. L., (1999). Some traffic features at freeway bottlenecks. *Transportation Research: Part B* 33(1), 25-42.

Cassidy, M.J., Rudjanakanoknad, J. (2005) Increasing the Capacity of an Isolated Merge by Metering Its On-Ramp, *Transportation Research: Part B*, vol. 39, no. 10, pp. 896-913.

Chandler, R.E., Herman, R., Montroll, E.,W. (1958) "Traffic dynamics: studies in car following," *Operation Research*, Vol. 6, No. 2, pp. 165-184.

Chong, L., Abbas, M., Medina Flintsch, A., Higgs, B. (2013) A rule-based neural network approach to model driver naturalistic behavior in traffic, *Transportation Research Part C: Emerging Technologies*, Volume 32, pp 207-223.

Chung, K., Rudjanakanoknad, J., Cassidy, M. J. (2007) Relation Between Traffic Density and Capacity Drop at Three Freeway Bottlenecks, *Transportation Research: Part B*, vol. 41, no. 1, pp. 82-95.

Coifman, B. (2014) Revisiting the Empirical Fundamental Relationship, *Transportation Research- Part B*, Vol 68, pp 173-184.

Coifman, B. (2015) Empirical Flow-Density and Speed-Spacing Relationships: Evidence of Vehicle Length Dependency, *Transportation Research- Part B,* Volume 78, pp 54-65.

Coifman, B. (2016) *Data Sets*, http://www.ece.osu.edu/~coifman/documents/

Coifman, B., Beymer, D., McLauchlan, P., Malik, J. (1998) A Real-Time Computer Vision System for Vehicle Tracking and Traffic Surveillance, *Transportation Research: Part C,* vol 6, no 4, pp 271-288.

Coifman, B., Kim, S. (2011). Extended Bottlenecks, the Fundamental Relationship and Capacity Drop on Freeways, *Transportation Research: Part A*. vol. 45, no. 9, pp. 980-991.

Coifman, B., Krishnamurthy, S., and Wang, X. (2003) Lane Change Maneuvers Consuming Freeway Capacity, *Proc. of the Traffic and Granular Flow 2003 Conference*, October 3, 2003, Delft, Netherlands, pp. 3-14.

Coifman, B., Krishnamurthy, S. (2007) Vehicle Reidentification and Travel Time Measurement Across Freeway Junctions Using the Existing Detector Infrastructure, *Transportation Research- Part C,* Vol 15, No 3, pp 135-153.

Daamen, W., Loot, M., Hoogendoorn, S. (2010) Empirical Analysis of Merging Behavior at Freeway On-Ramp, *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2188, pp 108-118.

Dingus, T., Klauer, S., Neale, V., Petersen, A., Lee, S., Sudweeks, J., Perez, M., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z., Jermeland, J., Knipling, R. (2006) *The 100-Car Naturalistic Driving Study, Phase II - Results of the 100-Car Field Experiment*, Virginia Polytechnic Institute and State University, Blacksburg; National Highway Traffic Safety Administration, 422p.

Duret, A., Buisson, C., Chiabaut, N. (2008) Estimating Individual Speed-Spacing Relationship and Assessing Ability of Newell's Car-Following Model to Reproduce Trajectories, *Transportation Research Record 2088*, pp 188-197.

Duret, A., Bouffier, J., Buisson, C. (2010) Onset of Congestion due to Low Speed Merging Maneuvers within a Free-Flow Traffic Stream: Analytical Solution, *Transportation Research Record 2188*, pp. 96-107.

Eenink, R., Barnard, Y., Baumann, M., Augros, X., Utesch, F. (2014) UDRIVE: the European naturalistic driving study, *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*, Institut Francais des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR), 10p.

Forbes, T., Simpson, M. (1968) Driver-And-Vehicle Response In Freeway Deceleration Waves, *Transportation Science*, Operations Research Society of America, Volume 2, Issue 1, pp 77-104.

Gao, B., Coifman, B. (2006) Vehicle Identification and GPS Error Detection from a LIDAR Equipped Probe Vehicle, *Proc. of the 9th IEEE International Conference on Intelligent Transportation Systems*, Toronto, Canada, September 17-20.

Gazis, D. C., Herman, R., Rothery, R. W. (1961) Nonlinear follow the leader models of traffic flow, *Operations Research*, Vol. 9, No. 4, pp. 545-567.

Haight, F. (1963) The Future of Traffic Flow Theory, *Traffic Quarterly*, Volume 17, Issue 4, pp. 516-527

Hall, F.L., Agyemang-Duah, K., (1991) "Freeway Capacity Drop and the Definition of Capacity," *Transportation Research Record 1320*, pp. 91-98.

Hamdar, S., Mahmassani, H. (2008) Driver Car-Following Behavior: From Discrete Event Process to Continuous Set of Episodes, *Proc. of the 87th Annual Meeting of the Transportation Research Board*.

Herman, R., Potts, R. (1959) Single Lane Traffic Theory and Experiment, *Proc. Symposium on Theory of Traffic Flow*, Elsevier, New York, pp 147-157.

Hurdle, V.F., Datta, P. K., (1983) "Speeds and Flows on an Urban Freeway: Some Measurements and a Hypothesis," *Transportation Research Record 905*, pp. 127-137.

Kim, S., Coifman, B. (2013) Driver Relaxation Impacts on Bottleneck Activation, Capacity, and the Fundamental Relationship, *Transportation Research- Part C,* Vol 36, pp 564-580.

Knoop, V., Hoogendoorn, S., Shiomi, Y., Buisson, C. (2012) Quantifying the Number of Lane Changes in Traffic: Empirical Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2278, pp 31-41.

Kovvali, V., Alexiadis, V., Zhang, L. (2007) Video-Based Vehicle Trajectory Data Collection, *Proc. of the 86th Annual TRB Meeting*, TRB.

Laval, J.A., Daganzo, C.F. (2006) Lane-Changing in Traffic Streams, *Transportation Research: Part B*, vol. 40, no. 3, pp. 251-64.

Lee, H., Coifman, B. (2012) Side-Fire LIDAR Based Vehicle Classification, *Transportation Research Record 2308*, pp 173-183.

Lee, H., Coifman, B. (2015) Using LIDAR to Validate the Performance of Vehicle Classification Stations, *Journal of Intelligent Transportation Systems*. Vol 19, No 4, pp 355-369.

Lee, S., Olsen, E., Wierwille, W. (2004) *A Comprehensive Examination Of Naturalistic Lane-Changes*, Virginia Polytechnic Institute and State University, Blacksburg; National Highway Traffic Safety Administration, 232 p.

Lighthill, M., Whitham, G. (1955) "On Kinematic Waves II. a Theory Of Traffic Flow on Long Crowded Roads, "*Proc. Royal Society of London, Part A*, Vol. 229, No. 1178, pp 317-345.

Marczak, F., Daamen, W., Buisson, C. (2014a) Empirical Analysis Of Lane Changing Behaviour At A Freeway Weaving Section, *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*, Institut Francais des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR), 10p.

Marczak, F., Daamen, W., Buisson, C. (2014b) Empirical Analysis Of Lane Changing Behavior At A Freeway Weaving Section, *Transportation Research Board 93rd Annual Meeting*, Transportation Research Board, 2014, 14p.

Montanino, M., Punzo, V. (2015) Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns, *Transportation Research Part B*, Vol 80, pp 82-106.

Making NGSIM data usable for studies on traffic flow theory: a Multistep Method for Vehicle Trajectory Reconstruction, *Transportation Research Record: Journal of the Transportation Research Board, Issue 2390*, pp 99-111.

Ponnu, B., Coifman, B. (2015) Speed-Spacing Dependency on Relative Speed from the Adjacent Lane: New Insights for Car Following Models," *Transportation Research Part B*. Vol 82, 2015, pp 74-90.

Punzo, V., Borzacchiello, M., Ciuffo, B. (2011) On the Assessment of Vehicle Trajectory Data Accuracy and Application to the Next Generation SIMulation (NGSIM) Program Data, *Transportation Research Part C*, Vol 19, No 6, pp 1243-1262.

Regan, M., Williamson, A., Grzebieta, R., Charlton, J., Lenne, M., Watson, B., Haworth, N., Rakotonirainy, A., Woolley, J., Anderson, R., Senserrick, T., Young, K. (2013) The Australian 400-car Naturalistic Driving Study: innovation in road safety research and policy. *Australasian Road Safety Research Policing Education Conference*, 2013, Brisbane, Queensland, Australia, 13p.

Richards, P. (1956) "Shock Waves on the Highway," *Operations Research*, Vol. 4, No. 1, pp 42-51.

Sangster, J., Rakha, H., Du, J. (2013) Application of Naturalistic Driving Data to Modeling of Driver Car-Following Behavior, *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2390, pp 20-33.

Schorr, J. and Hamdar, S., Silverstein , C. (2016) Measuring the Safety Impact of Road Infrastructure Systems on Driver Behavior: Vehicle instrumentation and Real World Driving Experiment, *Journal of Intelligent Transportation Systems*, http://www.tandfonline.com/doi/abs/10.1080/15472450.2016.1198699, 11p.

Smith, S. (1985) *Freeway Data Collection for Studying Vehicle Interactions*, Technical Report FHWA/RD-85/108, FHWA, USDOT

Smith, S., Mark, E. (1985) Creation of data sets to study microscopic traffic flow in freeway bottleneck sections, *Transportation Research Record 1005*, pp 121-128.

Thiemann, C., Treiber, M., Kesting, A. (2008) Estimating acceleration and Lane-Changing Dynamics from Next Generation Simulation Trajectory Data, *Transportation Research Record 2088*, pp 90-101.

Thornton, D., Redmill, K., Coifman, B. (2014) Automated Parking Surveys from a LIDAR Equipped Vehicle, *Transportation Research- Part C,* Vol 39, pp 23-35.

Tong, D., Merry, C., Coifman, B. (2009) New Perspectives on the Use of GPS and GIS to Support a Highway Performance Study," *Transactions in GIS,* Vol 13 No 1, pp 69-85.

Treiterer, J., Myers, J. (1974) The Hysteresis Phenomenon in Traffic Flow, *Proc., Sixth International Symposium on Transportation and Traffic Theory* (D. J. Buckley, ed.), pp 13-38.

Wang, C., Coifman, B. (2008) The Effect of Lane Change Maneuvers on a Simplified Car-following Theory, *IEEE Transactions on Intelligent Transportation Systems*, Vol 9, No 3, pp 523-535.

Xin, W., Hourdos, L., Michalopoulos, P. (2008) Vehicle Trajectory Collection and Processing Methodology and Its Implementation. *Proc. of the 87th Annual TRB Meeting*, TRB.

Xuan, Y., Coifman, B. (2012) Identifying Lane Change Maneuvers with Probe Vehicle Data and an Observed Asymmetry in Driver Accommodation, *ASCE Journal of Transportation Engineering*, Vol 138, No 8, pp 1051-1061.
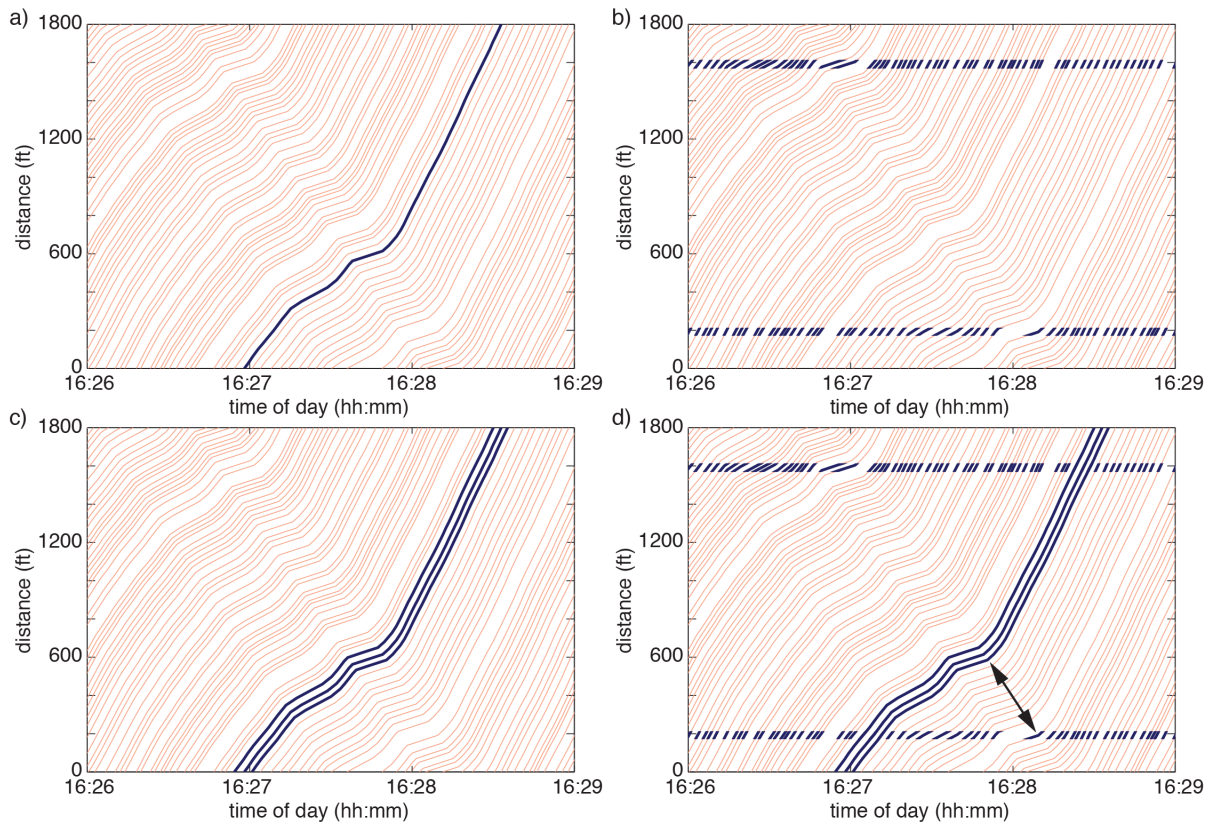
Figure 1, Hypothetical vehicle trajectories within a single lane over 1/3 mile, bold curves observed, faint curves unobserved, as seen by: (a) a conventional probe vehicle, (b) two loop detector stations, (c) ambient vehicles from the instrumented probe vehicle, (d) the combination of b and c to provide a more complete picture than either source alone.
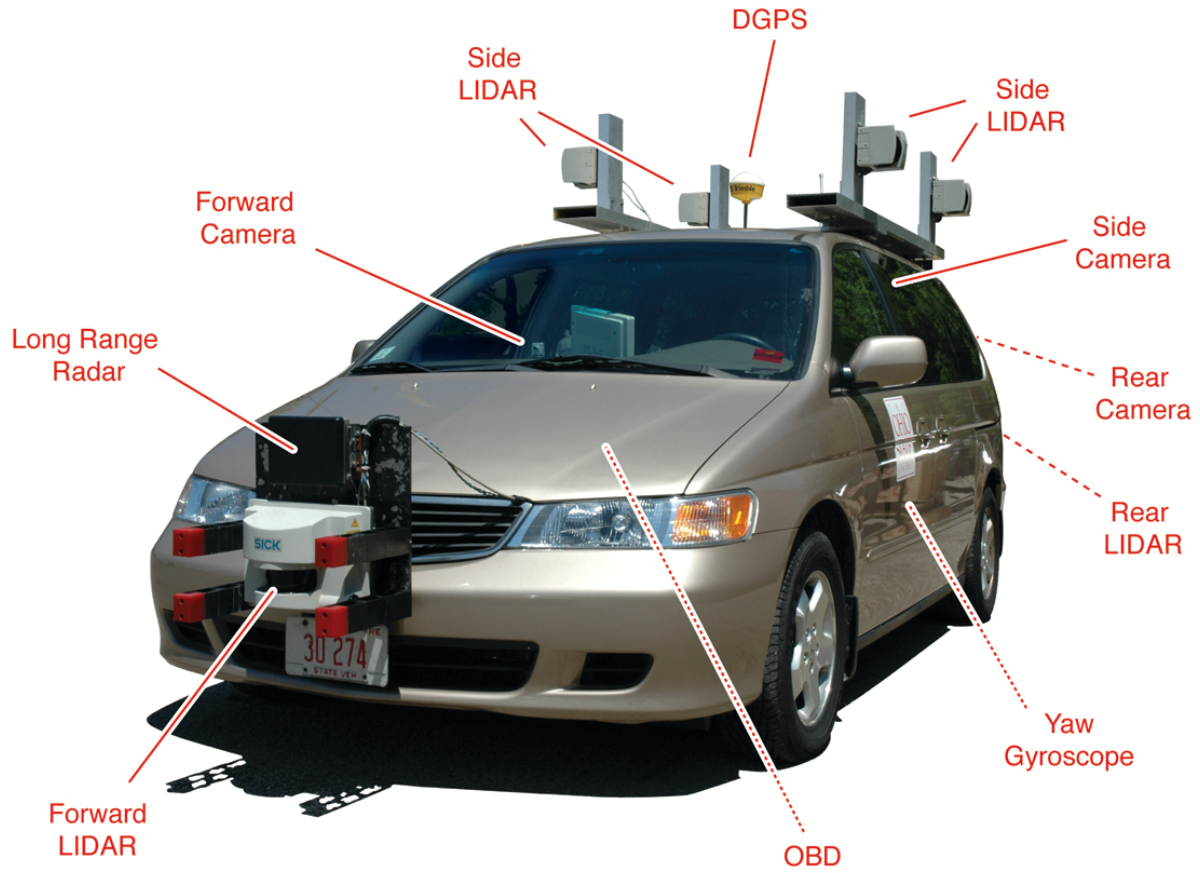
Figure 2, The instrumented probe vehicle with the various sensors highlighted, including the forward and rearward LIDAR sensors, forward radar sensor, DGPS, OBD data, and cameras for collecting validation video.
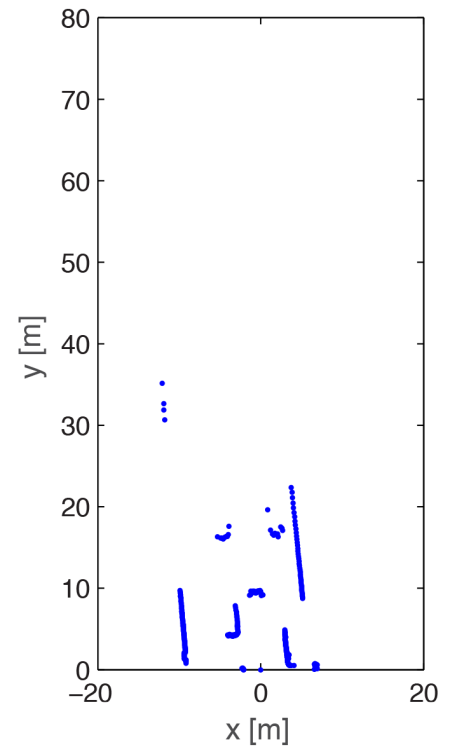
Figure 3, Forward LIDAR (a) superimposed on the corresponding video frame, and (b) as viewed in the original scanning plane parallel to the ground. The projection on to the video frame makes it easier for a person to distinguish between different vehicle and non-vehicle targets.

Figure 4, The instrumented probe vehicle made two round trips of the highlighted freeway route in Columbus, Ohio, USA over approximately 2 hr during the evening peak on September 9, 2009.
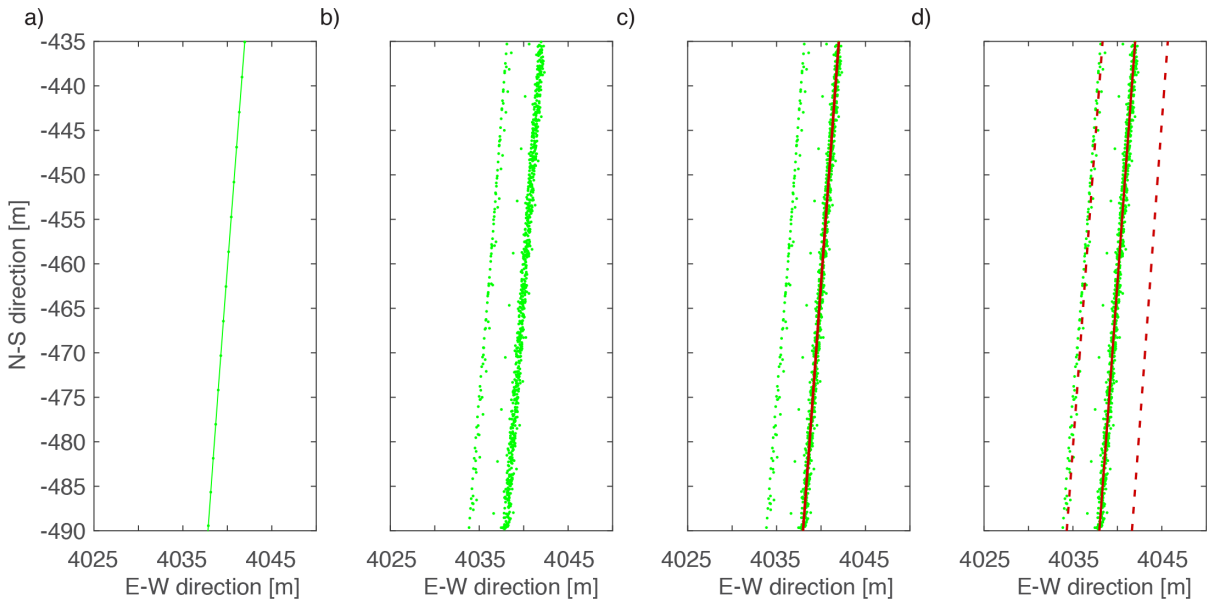
Figure 5, An example of establishing the reference lane, (a) a single directional pass of the probe vehicle, (b) overlapping 80 passes, (c) dominant lane identification, (d) calculated location of all remaining lanes.
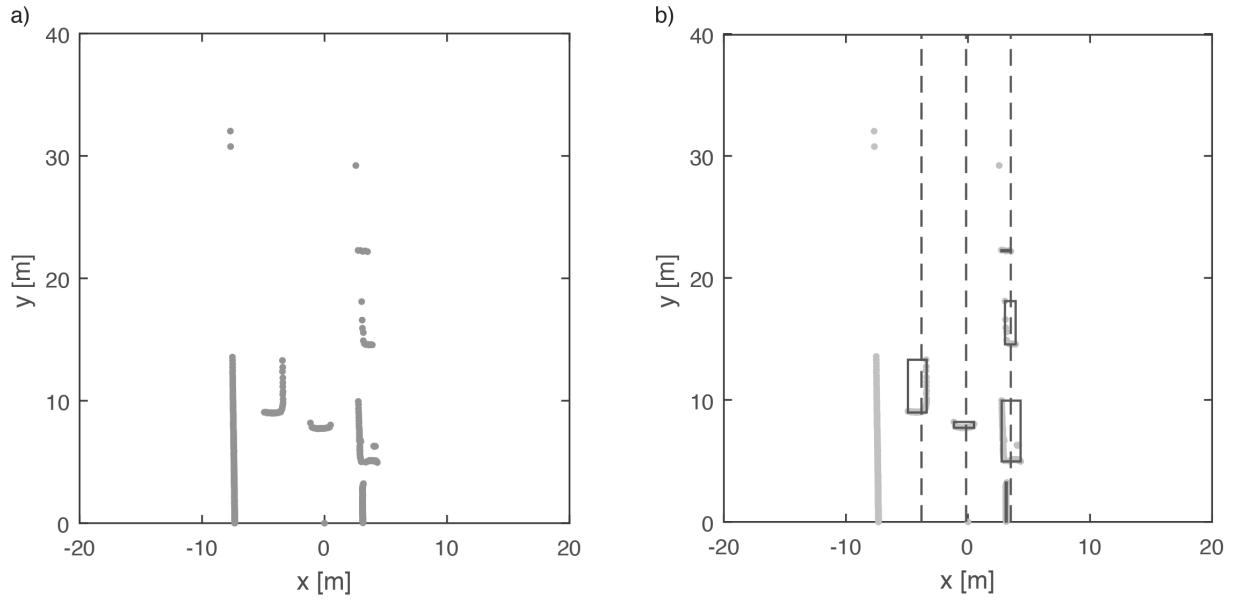
Figure 6, (a) Raw LIDAR returns in a typical scan from the front LIDAR sensor, (b) the corresponding bounding boxes for the identified on-road clusters after being assigned to specific lanes. The centerline of each lane is denoted with a dashed curve and the probe vehicle always travels in the lane passing through the origin of the LIDAR coordinate system.
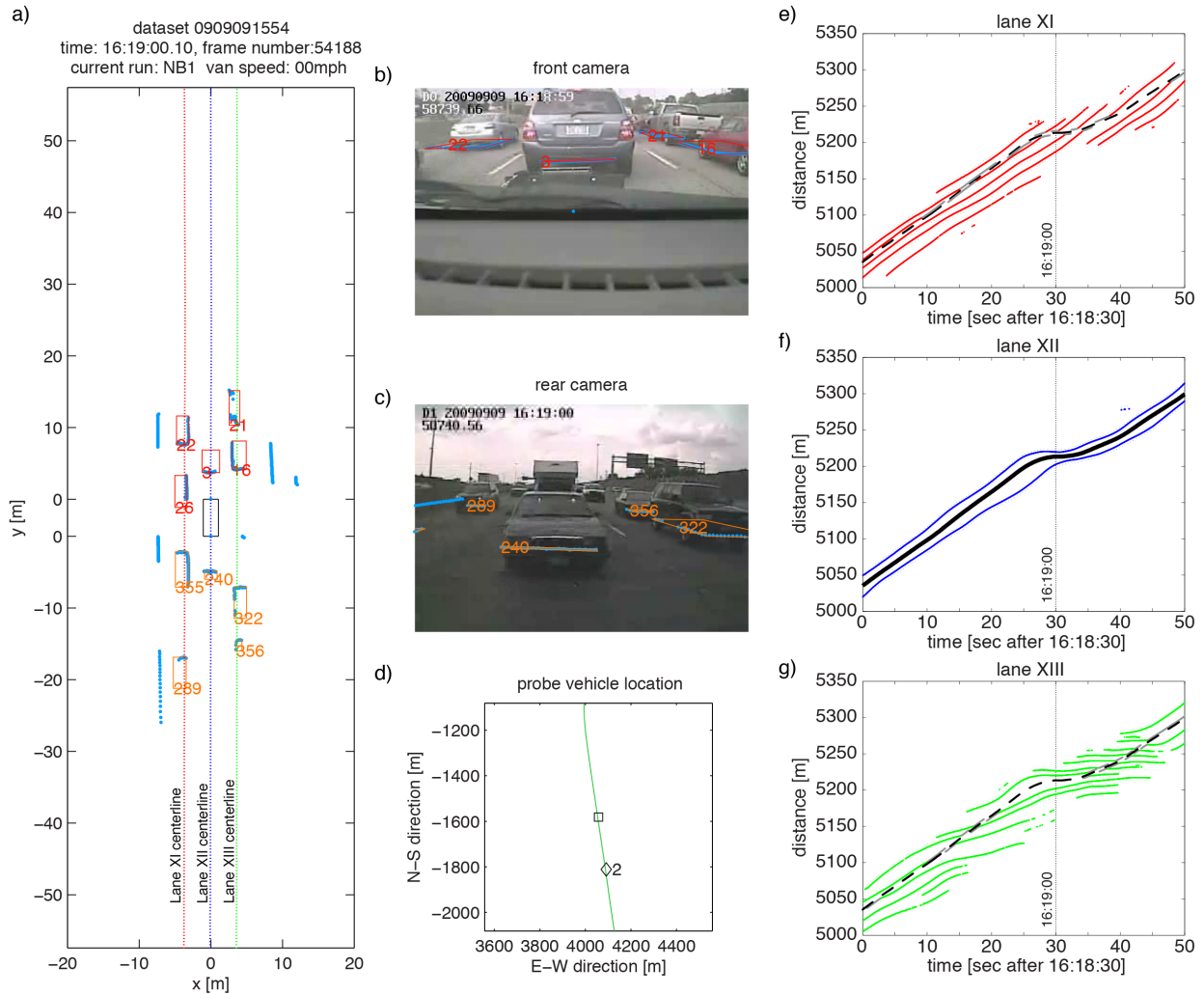
Figure 7, An example of the graphical user interface used to review the vehicle grouping, (a) front and rear LIDAR returns in the ground plane at 16:19:00.10 and numbered bounding boxes around the tracked targets. Corresponding video frames from, (b) the front camera, and (c) the rear camera. (d) The current location of the probe vehicle. The extracted vehicle trajectories from shortly before to shortly after the current instant shown in a-d, with this current instant shown by a vertical dashed line in (e) lane XI, to the left of the probe vehicle, (f) lane XII, the probe vehicle's lane of travel, and (g) lane XIII, to the right of the probe vehicle. The bold curve in each trajectory plot is the probe vehicle's trajectory, shown only for reference with a dashed line in lanes XI and XII.
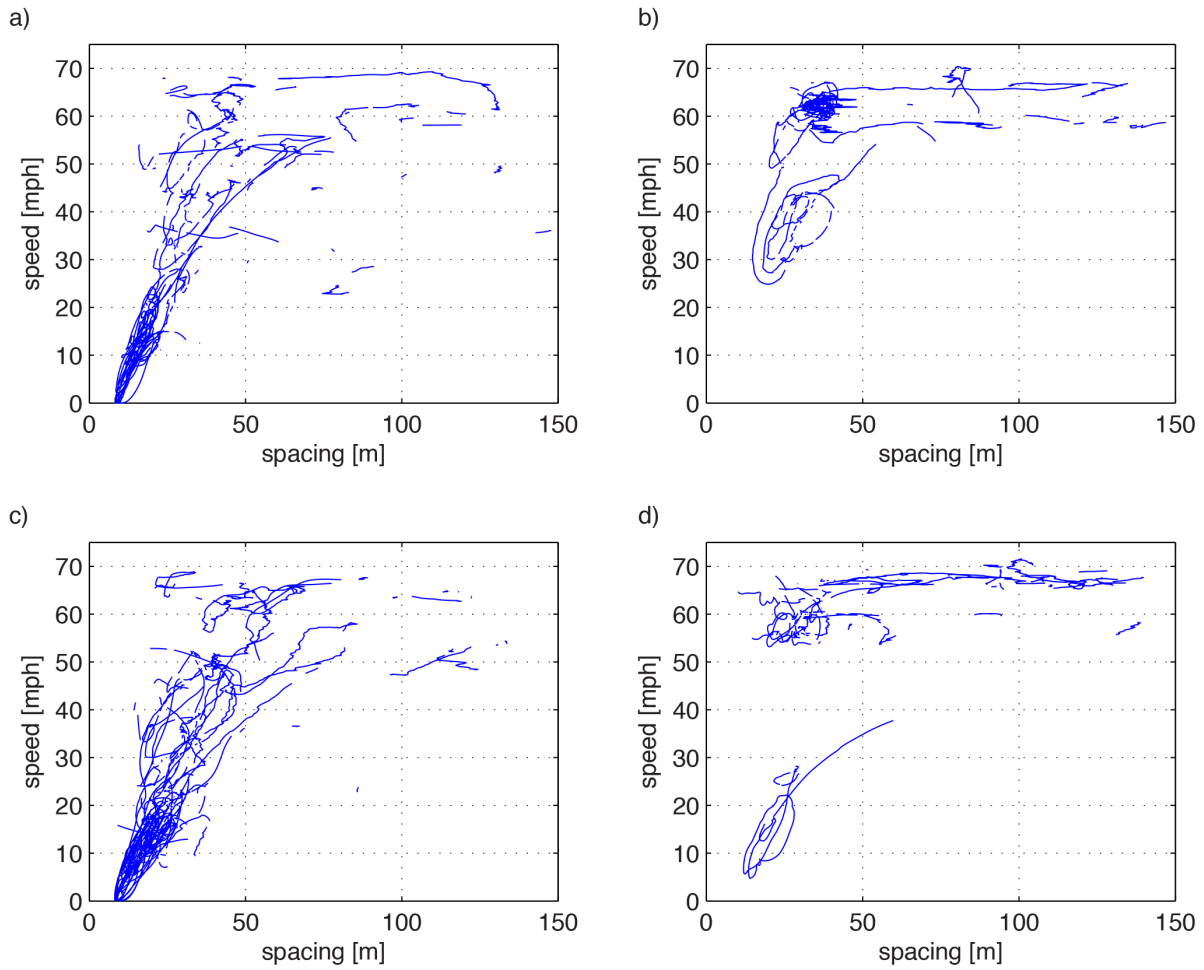
26

Figure 8, The resulting speed-spacing relationships exhibited by the instrumented probe vehicle as measured from the combined LIDAR and radar data over the four passes, (a) northbound 1, (b) southbound 1, (c) northbound 2, (d) southbound 2. The various plots show many individual curves rather than points, where each curve is one contiguous period with a single tracked leader. In the free flow regime note that the exhibited free speeds reflect the two different speed limits along the tour.
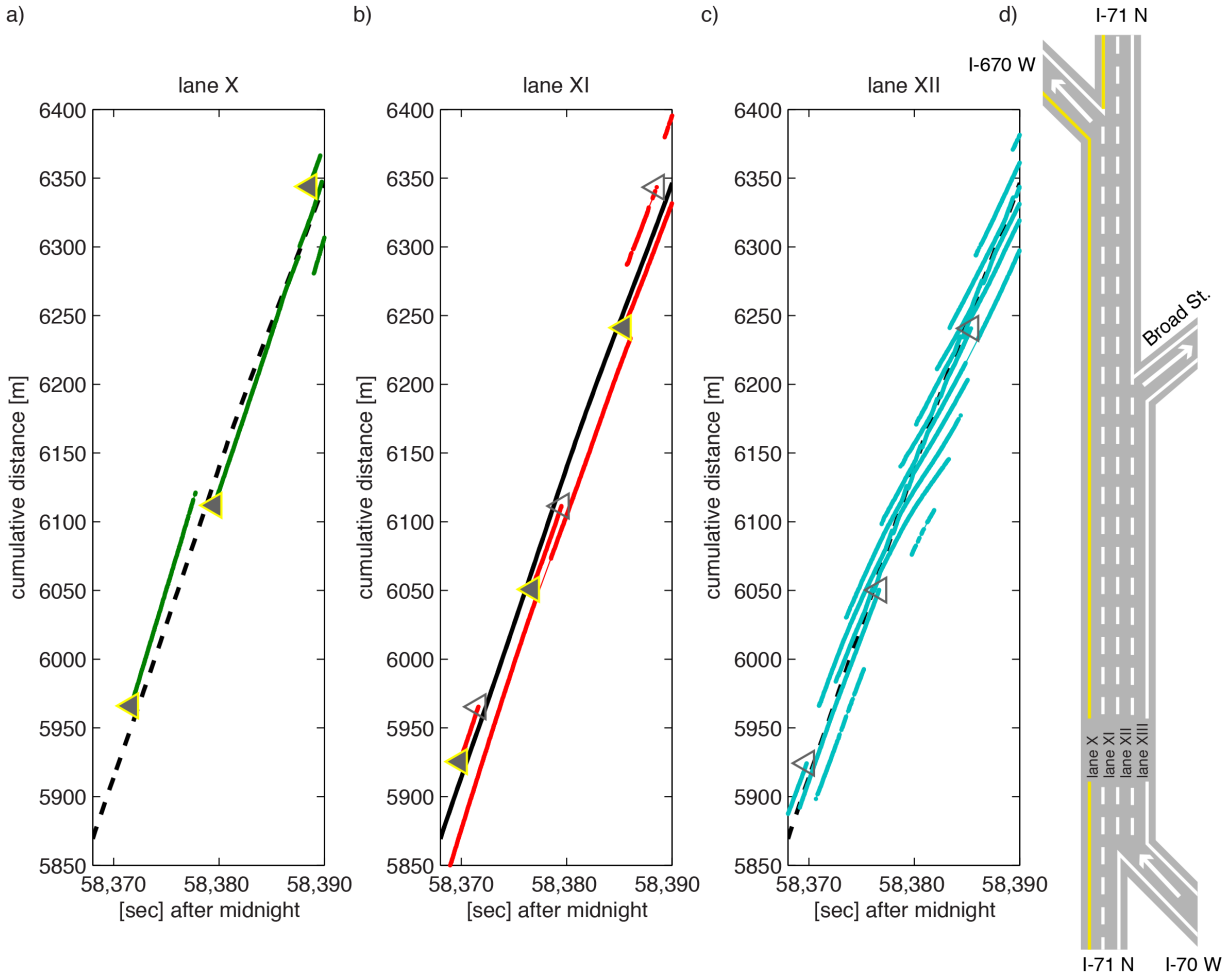
Figure 9, Ambient vehicle trajectories around the probe vehicle on I-71 from the LIDAR data in (a) lane X, (b) lane XI and (c) lane XII. Traffic moves from bottom to top with the probe vehicle strictly in lane XI, as indicated with a bold curve (this trajectory is repeated for reference in the other lanes with a dashed curve). The schematic in (d) is roughly to scale with the distance axis in the plots and shows that lanes XII and XIII join from I-70 westbound around 5,850 m, lane XIII exits at Broad St. around 6,200 m, lane X leaves and lane XII splits to I-670 westbound around 6,400 m. Triangles indicate the direction of lane change maneuvers with empty triangles denoting an exit and solid triangles denoting an entrance.
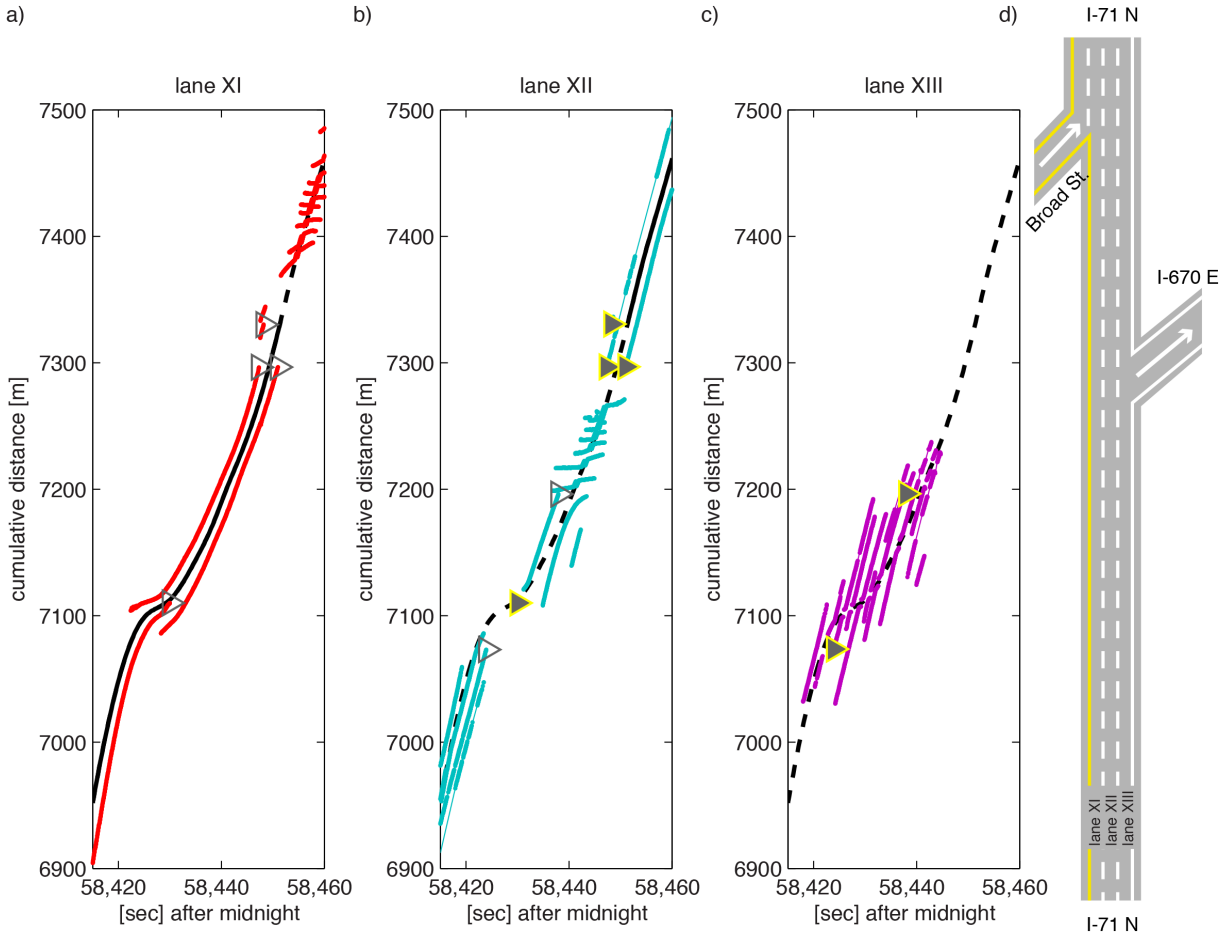
28

Figure 10,   Ambient vehicle trajectories around the probe vehicle on I-71 from the LIDAR data in (a) lane XI, (b) lane XII and (c) lane XIII. Traffic moves from bottom to top and the notation follows from Fig. 9. The probe vehicle's trajectory is shown with a bold curve starting in lane XI and moving to lane XII around 7.300 m, just past a major connector ramp to I-670 eastbound, as shown in (d) the schematic, which is roughly to scale with the distance axis in the plots. Most of the vehicles seen in lane XIII were destined for this ramp. A box truck in lane XII had stopped at around 7,280 m in lane XII and was waiting to move to cross lane XIII to go to the I-670 ramp. At the end of this segment a ramp joins on the left with a merge lane, the resulting queue is evident in lane XI.
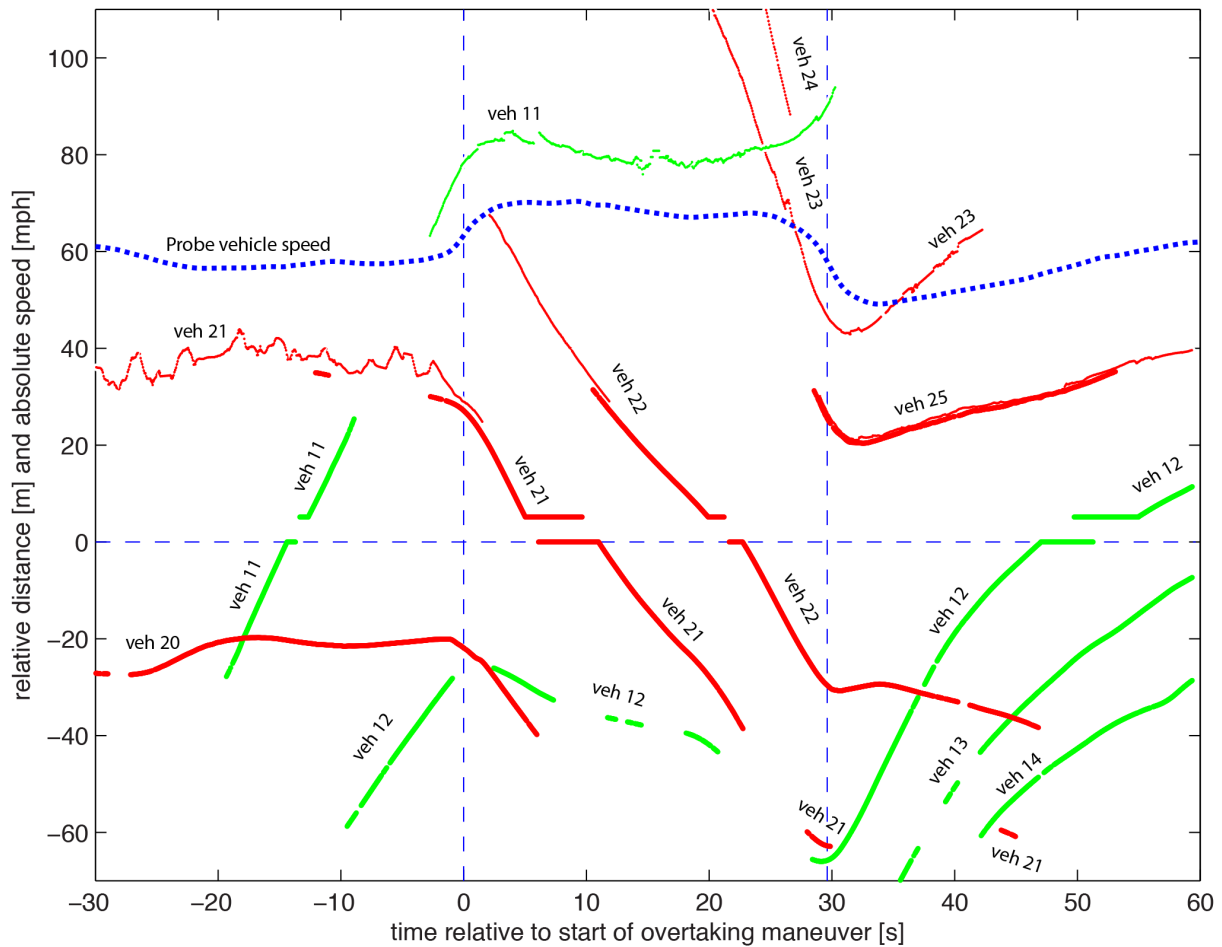
Figure 11, An example the relative inter-vehicle relationships as the instrumented probe vehicle completes an overtaking maneuver. Vertical distance is shown relative to the rear of the probe at 0 with distance increasing in the forward direction of travel. The initial lane change maneuver from the center lane to the left lane occurs at 0 sec in this plot and the subsequent return to the center lane at 30 sec, as indicated by vertical dashed lines in the figure. Vehicles in the exited center lane are labeled 2• and the left lane are labeled 1•. The LIDAR vehicle positions are shown with bold curves and the radar positions with lighter weight curves. The probe vehicle's speed is shown with a bold dashed curve, relative to the same vertical scale. Time zero in this plot corresponds to 60409 sec in the data set.

Contacts

For more information:

Benjamin Coifman, PhD
The Ohio State University
Department of Civil, Environmental, and Geodetic Engineering
Hitchcock Hall 470
2070 Neil Ave, Columbus, OH 43210

(614) 292-4282
Coifman.1@OSU.edu

https://ceg.osu.edu/~coifman

NEXTRANS Center
Purdue University - Discovery Park
3000 Kent Ave.
West Lafayette, IN 47906

nextrans@purdue.edu
(765) 496-9724

www.purdue.edu/dp/nextrans