Automatic Intersection Map Generation

V2I Safety Applications Project
Task 10 Report

www.its.dot.gov/index/htm
Final Report – February 29, 2016
FHWA-JPO-16-416

U.S. Department of Transportation
Federal Highway Administration

Produced by Crash Avoidance Metrics Partners LLC in response to Cooperative Agreement Number DTFH6114H00002

U.S. Department of Transportation
Federal Highway Administration

Notice

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-JPO-16-416	(Remove; Insert Information Here or leave blank)	(Remove; Insert Information Here or leave blank)

4. Title and Subtitle	5. Report Date
Automatic Intersection Map Generation Task 10 Report	February 29, 2016
	6. Performing Organization Code (Remove; Insert Information Here or leave blank)
7. Author(s) Jianfeng Zheng and Henry X. Liu	**8. Performing Organization Report No.** (Remove; Insert Information Here or leave blank)
9. Performing Organization Name And Address Crash Avoidance Metrics Partners LLC on behalf of the Vehicle-to-Infrastructure Consortium 27220 Haggerty Road, Suite D-1 Farmington Hills, MI 48331	**10. Work Unit No. (TRAIS)** (Remove; Insert Information Here or leave blank) **11. Contract or Grant No.** DTFH6114H00002
12. Sponsoring Agency Name and Address **FHWA Headquarters** 1200 New Jersey Avenue, SE West Building Washington, DC 20590	**13. Type of Report and Period Covered** Final Report June 2015 – February 2016 **14. Sponsoring Agency Code** (Remove; Insert Information Here or leave blank)

15. Supplementary Notes

(Remove; Insert Information Here or leave blank)

16. Abstract

This report describes the work conducted in Task 10 of the V2I Safety Applications Development Project. The work was performed by the University of Michigan Transportation Research Institute (UMTRI) under contract to the Crash Avoidance Metrics Partners LLC (CAMP) Vehicle to Infrastructure (V2I) Consortium. Participating companies in the V2I Consortium were FCA US LLC, Ford, General Motors, Hyundai-Kia, Honda, Mazda, Nissan, Subaru, Volvo Truck, and VW/Audi. This project investigated the feasibility of automatically generating intersection maps in SAE J2735 MAP format using Basic Safety Messages (BSMs) received by Roadside Units (RSUs). A procedure for classifying vehicle trips through the intersection and subsequently estimating lane centerlines was developed. In addition, a method for associating vehicle movements to the green signal phase was also developed using Signal Phase and Timing (SPaT) messages. BSM data from five intersections used in the Safety Pilot Model Deployment Project were analyzed. The estimated maps were compared to reference maps produced from LIDAR surveys of the intersections. With 48 traffic lanes in total, 43 lanes and associated vehicle movements were correctly identified using the estimation approach. Five lanes were not successfully identified due mainly to a lack of BSM data. For the identified lanes, two measurements of accuracy were calculated: the mean distance of the estimated geometry node points to closest points in the surveyed lane centerline geometry and the maximum distance of estimated points and surveyed lane geometry. On average, the mean distance was found to be 0.5 meter and maximum distance was found to be 1.2 meters between the estimated maps and reference maps. Overall, this project demonstrated the significant potential of using BSM data for estimation of an intersection map as well as association of the vehicle lanes (vehicle movements) to the traffic signal phases.

17. Key Words	18. Distribution Statement
Connected Vehicle, Intersection Map, Map Generation, V2I System, Roadside Equipment, Basic Safety Message	(Remove; Insert Information Here or leave blank)

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
(Remove; Insert Information Here or leave blank)	(Remove; Insert Information Here or leave blank)	**37**	(Remove; Insert

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Table of Contents

List of Figures

List of Tables

Executive Summary

Connected Vehicle (CV) technology has gained significant attention in the U.S. due to its potential to improve safety and mobility performance of the transportation system. One crucial element of many CV applications is an accurate representation of road segments, i.e., location maps. At traffic signal intersections, such location maps or Geometric Intersection Description (GID) maps, are conveyed in MAP messages for broadcasting by Roadside Units (RSUs), as defined by the SAE J2735 standard. However, if GID maps are manually surveyed, creation and maintenance of MAP messages could be costly and challenging for public agencies, especially for a large scale deployment.

This report describes the work completed during Task 10 of the Vehicle-to-Infrastructure Safety Applications (V2I-SA) Project. The overall goal of the project is to develop and evaluate safety applications in prototype vehicles using Dedicated Short Range Communication (DSRC)-based Vehicle-to-Infrastructure (V2I) communication. The V2I-SA Project is being conducted by the Crash Avoidance Metrics Partners LLC (CAMP) V2I Consortium. The participating companies in the V2I Consortium are FCA US LLC, Ford, General Motors, Hyundai-Kia, Honda, Mazda, Nissan, Subaru, Volvo Technology of America, and VW/Audi. The project is sponsored by the Federal Highway Administration (FHWA) through Cooperative Agreement DTFH611H0002, Work Order 0003. The project started September 15, 2014 and is scheduled to conclude in June 2016. The work in Task 10 was performed by the University of Michigan Transportation Research Institute (UMTRI) under contract to the CAMP V2I Consortium.

To overcome the drawbacks of manual surveys, this project developed an approach for automatically generating intersection maps using Basic Safety Messages (BSMs) received by RSUs. In the developed approach, we propose three processing steps:

1. Pre-processing to clean and categorize data
2. Estimating approach geometry and lane centerline geometry using categorized data
3. Post-processing to reorganize the lane centerline geometry data to be compatible with the SAE J2735 MAP message

The proposed approach was implemented to estimate intersection maps for five selected intersections from the Safety Pilot Model Deployment (SPMD) Project. To evaluate performance of the proposed approach, a highly accurate LIDAR survey was conducted to obtain reference maps.

The evaluation showed promising results for the proposed map estimation approach. Using one month of data, the estimation approach was able to correctly identify the main intersection structure regarding the number of lanes associated with the different movements for all of the five intersections. With 48 traffic lanes in total, 43 lanes and associated vehicle movements were correctly identified. Five lanes were not successfully identified due mainly to a lack of CV data for these lanes. For the identified lanes, two measurements of accuracy were calculated. One measurement is the mean distance of node points in estimated lane centerline geometries to closest points in the surveyed lane geometry, indicating the average accuracy of the estimation. The other measurement is the maximum distance of points between the estimated and surveyed lane geometry, indicating the robustness of the estimation.

Overall, the mean distances between the estimated and surveyed lane centerline geometries were approximately 0.2 to 1.5 meters, with the majority (90%) less than 1 meter. The maximum distances between the estimated and surveyed lanes were from 0 to 3.5 meters, with the majority (90%) less than 2.2 meters. On average, the mean distance was found to be 0.5 meter and maximum distance was found to be 1.2 meters between the estimated maps and reference maps.

In addition to map generation, the feasibility of automatic Signal Phase and Timing (SPaT)-MAP association (i.e., connecting vehicle lanes with associated traffic signal phases) was also investigated using SPaT and Basic Safety Message (BSM) data. A simple procedure was proposed to estimate the SPaT-MAP association based on the logic of matching vehicle departures with a green signal (i.e., maximizing green departures from the SPaT and BSM data). A case study was then conducted using data from one of the selected intersections. For a selected intersection, the proposed procedure is able to correctly associate vehicle movements to all applicable signal phases. The applicability of automatic SPaT-MAP association was confirmed.

Overall, this project demonstrated the significant potential of using CV data for estimation of an intersection map, as well as the SPaT-MAP association. The developed procedures may present cost-effective alternatives to manual surveys to generate these important elements of a Vehicle-to-Infrastructure (V2I) system. Therefore, the developed procedures could be helpful for both initial deployment as well as long-term maintenance of the CV system.

1 Introduction

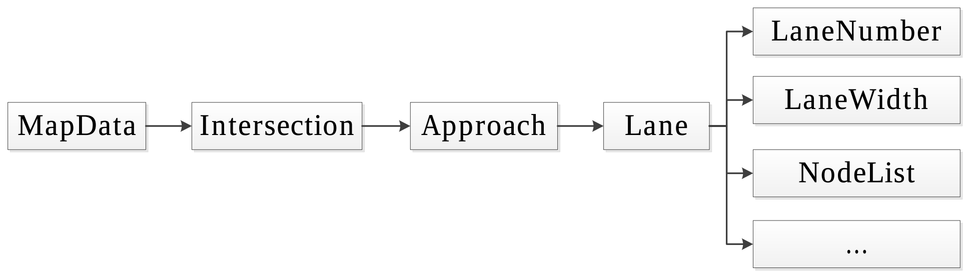
Connected Vehicle (CV) technology has gained significant attention in the U.S. due to its potential to improve safety and mobility performance of the transportation system. One crucial element of many CV applications is an accurate representation of the road segment (i.e., a location map). At traffic signal intersections the location maps, or Geometric Intersection Description (GID) maps, are conveyed in the SAE J2735 MAP Message (MAP) [1] for broadcasting by Roadside Units (RSUs). However, the creation and maintenance of maps could be very costly and challenging for agencies if the GID maps are manually surveyed.

This report describes the work completed during Task 10 of the Vehicle-to-Infrastructure Safety Applications (V2I-SA) Project. The overall goal of the project is to develop and evaluate safety applications in prototype vehicles using Dedicated Short Range Communication (DSRC)-based Vehicle-to-Infrastructure (V2I) communication. The V2I-SA Project is being conducted by the Crash Avoidance Metrics Partners LLC (CAMP) V2I Consortium. The participating companies in the V2I Consortium are FCA US LLC, Ford, General Motors, Hyundai-Kia, Honda, Mazda, Nissan, Subaru, Volvo Technology of America, and VW/Audi. The project is sponsored by the Federal Highway Administration (FHWA) through Cooperative Agreement DTFH611H0002, Work Order 0003. The project started September 15, 2014 and is scheduled to conclude in June 2016.

The work in Task 10 was performed by the University of Michigan Transportation Research Institute (UMTRI) under contract to the CAMP V2I Consortium. This effort developed an alternative to map surveying for the CV system, i.e., automatic map generation using Basic Safety Message (BSM) data collected by the RSU. In addition, as the MAP messages are being frequently used with SPaT messages at traffic signal intersections, a procedure for estimating SPaT-MAP association, i.e., lane-phase mapping, was also developed by using BSM and SPaT data jointly.

1.1 Background

The MAP message is one of the basic messages for communication between RSUs and Onboard Units (OBUs) for CV systems. As defined in SAE J2735, the MAP message (or MapData Message), contains mainly intersection information regarding the physical geometry and allowable maneuvers of each traffic lane. The core elements of the MapData are shown in Figure 1. There are three levels of information in the MapData, organized hierarchically: intersection level, approach level and lane level. Elements in the lane level are the most basic, including mainly information about vehicle lanes regarding lane width, allowable maneuvers, and a node list indicating the centerline positions within a lane. This information in the MAP message provides the basic location input for many Vehicle-to-Infrastructure (V2I) applications.



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 1: Essential Elements of MapData

To prepare map information, the conventional approach is to send personnel to the field to conduct manual surveys. Manual surveys are time consuming and costly, and could be challenging for large-scale deployment of a V2I system, particularly considering that there are approximately 300,000 signalized intersections in the U.S. High costs also prevent the map from being updated frequently and could easily result in outdated infrastructure information. As a consequence, availability and performance of CV applications would be compromised.

On the other hand, numerous research activities have been conducted for estimating a map using vast Global Positioning System (GPS) data as cost-effective alternatives to manual map surveys [2-7]. However, most existing studies have focused on estimating the topology of a large-scale traffic network rather than the geometry of individual intersections and are not suitable for deployment of V2I systems.

This project developed an approach to automatically generate an intersection map using BSM data from CVs. The benefits of this automatic map generation are twofold:

1. The need for manual surveying, and associated costs, could be largely eliminated for intersections which will be equipped with RSUs
2. Lane attribute changes, e.g., lane closure or modification, could be updated easily and promptly

The objective of the proposed approach is to assist in preparing and maintaining accurate, up-to-date intersection maps for V2I system deployment in the near future.

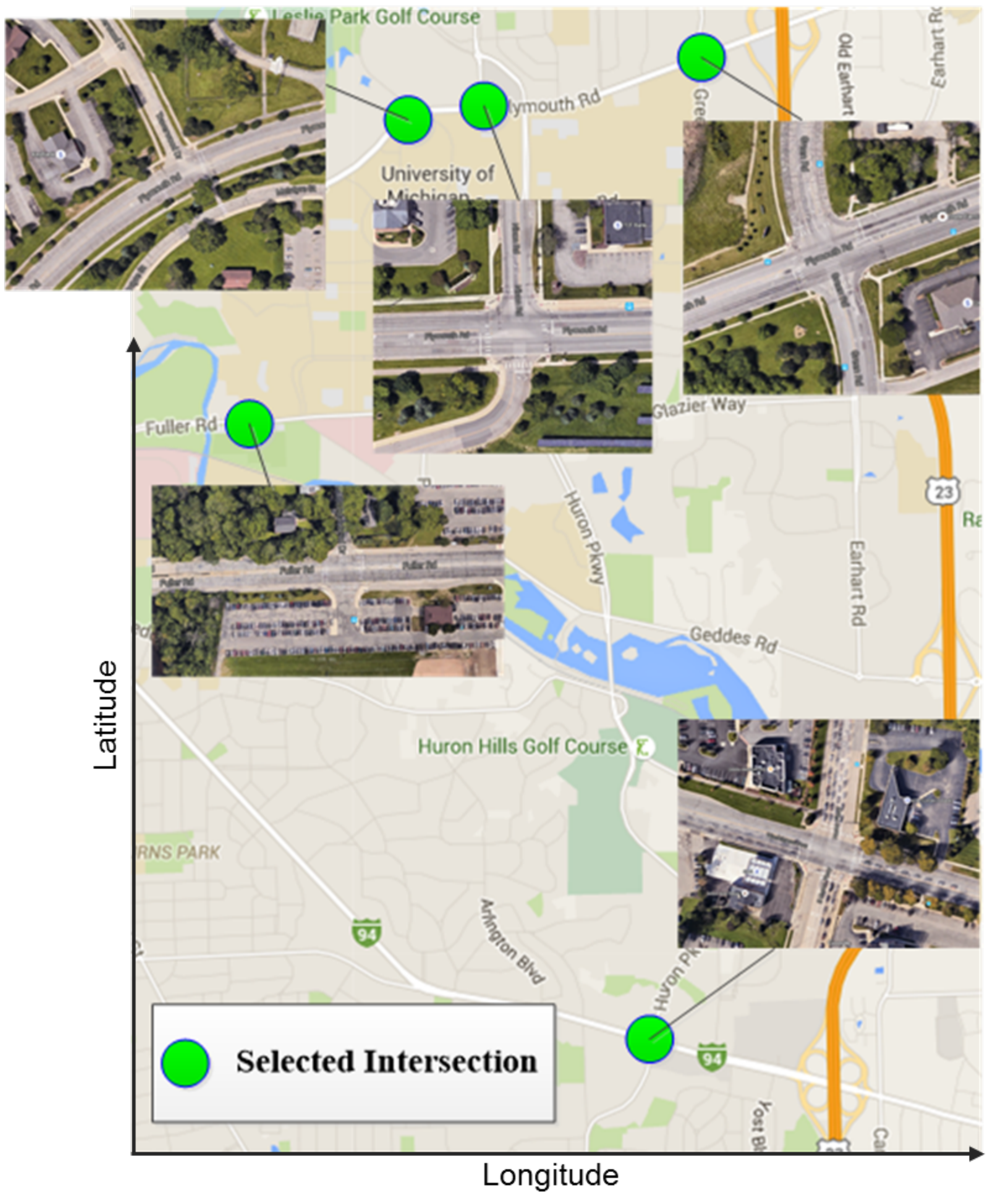
Since the SPaT messages are frequently used with MAP messages, the feasibility of automatic SPaT-MAP association (i.e., connecting vehicle lanes with traffic signal phases) using SPaT and BSM data was also investigated. A simple procedure was developed to estimate the association using the SPaT and BSM data, to facilitate preparation of the SPaT message for future deployment.

1.2 Organization of this Report

The rest of the report is organized as follows. Chapter 2 describes the details for the project's data preparation. Chapter 3 presents the main algorithm for intersection map generation. Chapter 4 presents the evaluation results by comparing estimated and surveyed maps. Chapter 5 presents the procedure and evaluation for automatic lane-phase mapping by jointly using SPaT data and BSM data. Lastly, conclusions are summarized in Chapter 6.

2 Data Preparation

In this project, five intersections equipped with RSUs during the Safety Pilot Model Deployment Project (SPMD) were selected for investigation. Aerial views of these five intersections are shown in Figure 2. All intersections are located in the Ann Arbor, Michigan area.



Source: © 2016 Google, Inc. Used with Permission.

Figure 2: Aerial Views of the Selected Intersections

The intersection list is shown in Table 1. The complexity levels of intersections were defined as follows:

1. Simple (S): intersections with at most two dedicated turning lanes
2. Moderately Complex (MC): intersections with 3-4 dedicated turning lanes
3. Complex (C): 4-leg intersections with four or more protected-turning lanes

Table 1: List of Candidate Intersections with RSUs from SPMD

Location	No. of LT	No. of RT	Other	Complexity
Fuller-Cedar Bend	2	0		S
Plymouth-Traverwood	3	1	T-intersection	MC
Plymouth-Nixon	4	0		MC
Plymouth-Green	4	2		C
Washtenaw-Huron	6	0		C

Source: University of Michigan Transportation Research Institute. Used with Permission.

With a range of different numbers of intersection legs, left-turn (LT) and right-turn (RT) lanes, these five selected intersections are good representatives of general traffic intersections. In this project, three types of data were prepared for analysis:

- BSMs collected by RSEs at intersections
- SPaT messages broadcasted by the RSUs
- High-precision reference maps generated by field surveys

2.1 BSM Data from RSUs

For each intersection, BSM data received by RSUs from the SPMD Project was prepared for later use. The data entries included the Identification Number (ID) for RSUs and CVs, GPS coordinates, speed and heading of CVs, as well as time stamps when the BSMs were generated, all extracted from received BSMs. One month of BSM data (January 2014) were retrieved from the SPMD database for this project. A sample of the BSM data from the SPMD database is shown in Table 2.

Table 2: Sample BSM Data Received by RSUs

RSUID	FileID	VehID	Gentime	TxRandom	MsgCount	Latitude	Longitude	Elevation
18001	1413	7568	271677215688596	1119	68	42.28638	-83.73286	203.2
18001	1413	7568	271677215788825	1119	69	42.28638	-83.73286	203

RSUID	FileID	VehID	Gentime	TxRandom	MsgCount	Latitude	Longitude	Elevation
18001	1413	7568	271677215888830	1119	70	42.28638	-83.73286	202.8
18001	1413	7568	271677215988588	1119	71	42.28639	-83.73286	202.6
18001	1413	7568	271677216088871	1119	72	42.28639	-83.73286	202.4

Source: University of Michigan Transportation Research Institute. Used with Permission.

2.2 SPaT Data

For selected intersections with capability of SPaT broadcasting, the SPaT data were also prepared for investigation of SPaT association. The data entries of SPaT data included time, signal phase ID, and signal status. A sample of the SPaT data is shown in Table 3.

Table 3: Sample SPaT Data from RSU

Time Stamp	Phase ID	Phase Status
2015-9-1,0:6:48.837	257	4
2015-9-1,0:6:48.837	1030	4
2015-9-1,0:6:48.837	259	1
2015-9-1,0:6:48.837	1035	1

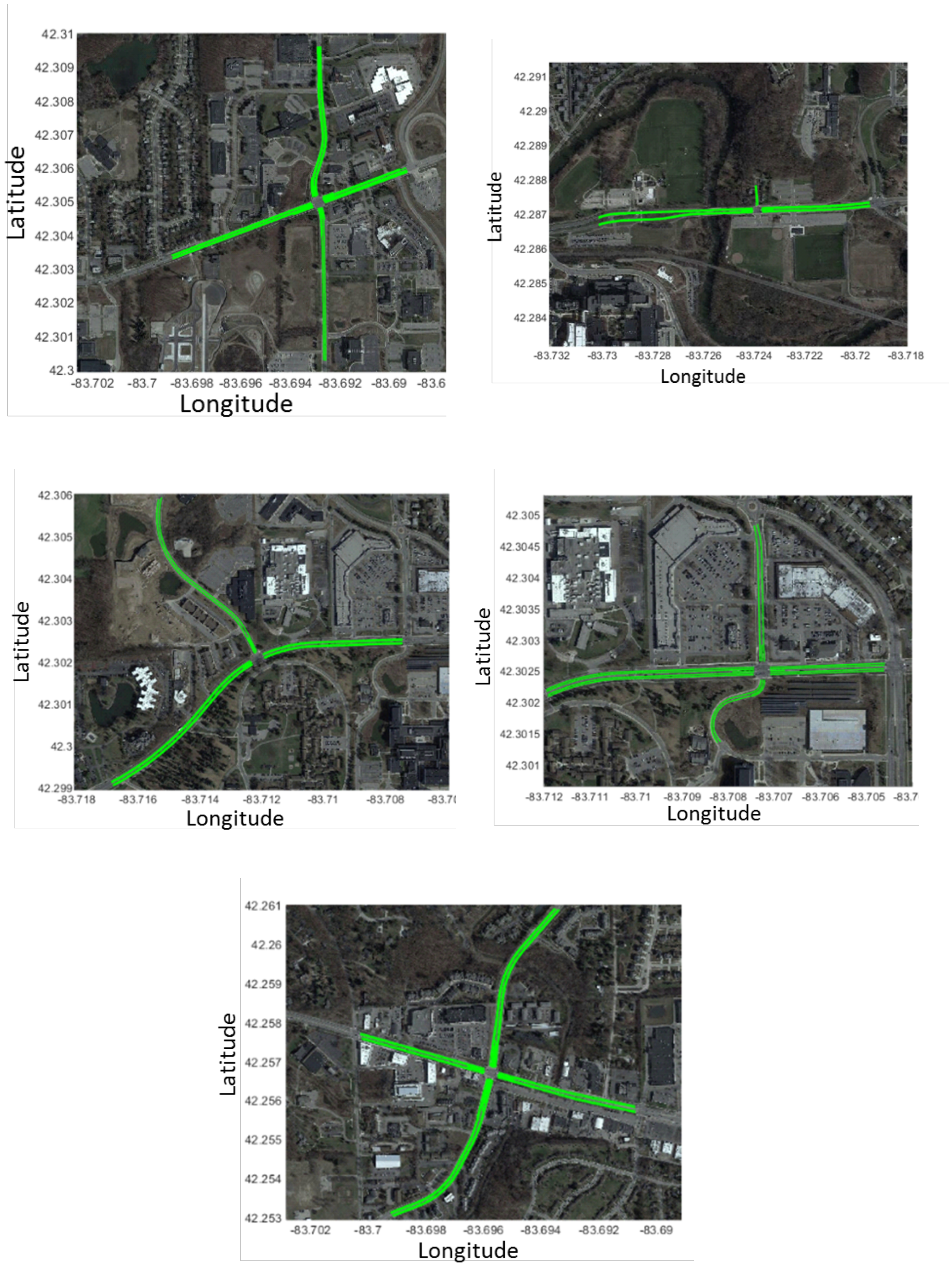
Source: University of Michigan Transportation Research Institute. Used with Permission.

2.3 Reference Intersection Map

To obtain the “ground truth” for the generated map, a local surveying company conducted a high-resolution Light Detection and Ranging (LIDAR) survey in the field to measure the geometry of the selected intersections. The LIDAR system uses laser light for detection and has the capability to measure both range and intensity of light return, combining advantages of both radar and vision detection systems. With these advantages, the LIDAR system can be used to detect road edges as well as lane markers within the road [8-10].

For the field survey, the LIDAR system could generate millions of data points as a “3D point cloud,” which can be used to precisely describe intersection geometry. In the survey system used, the LIDAR system was integrated with differential GPS to generate point lists of GPS coordinates for the lane geometry. In a prior project using this system, it was demonstrated that design-grade accuracies can be achieved by the survey system with an accuracy within 0.02 m [11]. More information about the LIDAR survey system can be found at [12].

As the final outcome from the field survey, Computer-Aided Design (CAD) drawing of traffic lanes and record files with digital coordinates for lane centerline geometries are generated as the reference intersection maps. Illustrations of the reference maps from the field survey are shown in Figure 3.

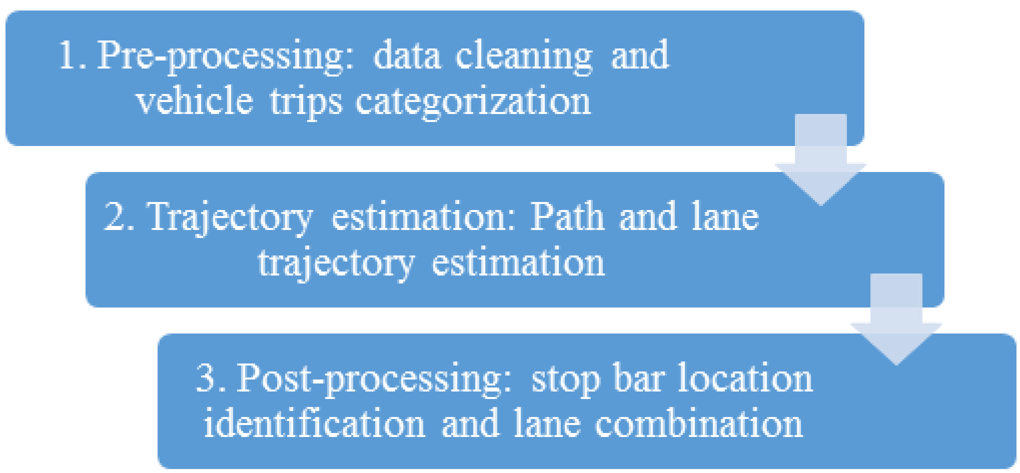


Source: © 2016 Google, Inc. Used with Permission.

Figure 3: Illustration of Reference Maps from Field Survey

3 Algorithm for Automatic Map Generation

The overall estimation process consists of three main steps, as shown in Figure 4. In Step 1, pre-processing cleans raw BSM data and also categorizes vehicle trips based on traveling direction and “movement type,” e.g., northbound (NB) through trip or westbound left-turn (WB-LT) trip. After categorizing vehicle trips, Step 2 aggregates the vehicle trips for each movement type. The geometry of a driving path is then estimated using aggregated vehicle position data. Lastly, lane positions are estimated along the generated driving path, and combined as individual lane centerline geometries. In Step 3, post-processing truncates the lane geometries, and organizes them to be compatible with the MAP message in the SAE J2735 standard. The details involved in each step are described in Sections 3.1 through 3.3.



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 4: Three Steps for Estimation Procedure

3.1 Pre-Processing Step: Data Cleaning and Trips Categorization Based On Movement Type

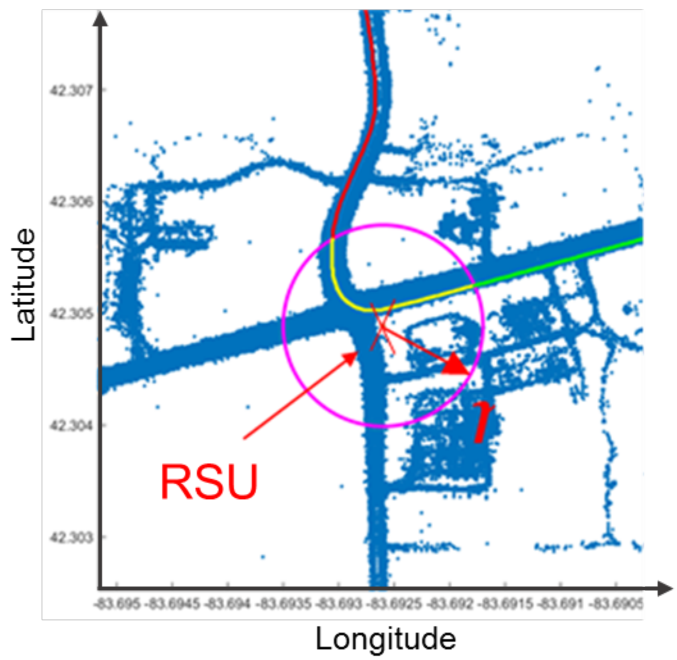
The first step of pre-processing was to clean raw data using a geofence to remove data outside of the geofence. The geofence boundary was defined as a square with the RSU at the center and extension of 0.01 unit for both latitude and longitude (roughly 780 meters).

The next step was to identify a topological representation of the intersection, i.e., the positions of approaches entering or exiting from the intersection. First, at each selected intersection, the root square distance (RS-Distance) between a CV with the RSU at the intersection based on their GPS coordinates, was calculated as:

$$\boldsymbol{d} = \sqrt{(\boldsymbol{Lat} - \boldsymbol{Lat_{RSU}})^2 + (\boldsymbol{Lng} - \boldsymbol{Lng_{RSU}})^2} \; (1)$$

Based on the RS-Distance, we next defined a circle with the RSU as the center. This circle was used as the “intersection center area,” where CVs enter and exit. The radius of the circle should be selected based on the consideration that the circle is large enough so that entering and exiting points of CVs on different approaches are sufficiently far from each other, but not so large that there are insufficient CV trips entering and exiting the circle.

The illustration of the circle is shown in Figure 5 using a data sample from the Green Road and Plymouth Road intersection. In the figure, the red mark indicates the position of RSU, and the blue points indicate raw GPS locations of CVs from BSMs received by the RSU. The purple circle indicates the intersection circle area, with RSU as the center. A sample CV trip is also shown, depicted by the red, yellow, and green line, representing the three portions of the trip: approaching, traversing, and leaving the circle, respectively.



Source: University of Michigan Transportation Research Institute. Used with Permission.

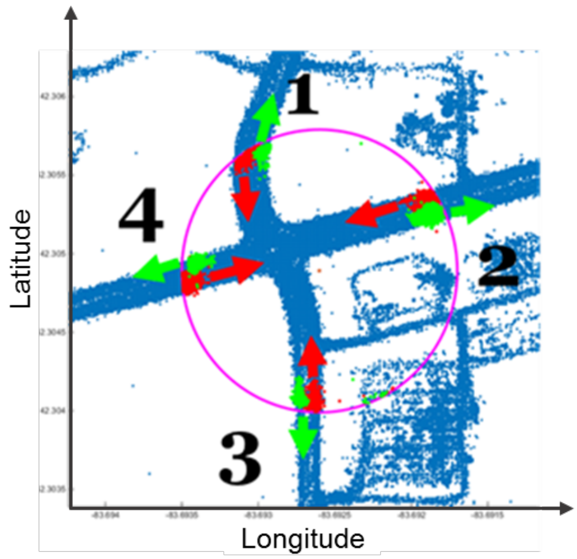
Figure 5: Illustration of the Circle for Intersection Center Area

Based on the intersection circle, two sets of data (one entering set and one exiting set) were prepared. The entering set included the first point of a CV entering the intersection circle for each CV. The exiting set included the last point of a CV within the intersection circle for each CV. The two sets are expressed as follows:

$$U_{enter} = \{[Lat_i, Lng_i] | [Lat_i, Lng_i] \text{ is the first point of CV trajectory in the circle}\}$$

$$U_{exit} = \{[Lat_j, Lng_j] | [Lat_j, Lng_j] \text{ is the last point of CV trajectory in the circle}\}$$

In both the entering and exiting sets, the points are further clustered into different groups, where a group represented a vehicle approach. Here we used the K-Means method for clustering, because of its simplicity and relatively low computational burden. For the Green Road and Plymouth Road intersection, the clustering results are shown in Figure 6. In the figure, the red points and green points are entering points and exiting points, respectively. For both entering and exiting points, four groups have been identified corresponding to the four approaches of the intersection, labeled as 1-4 clockwise.



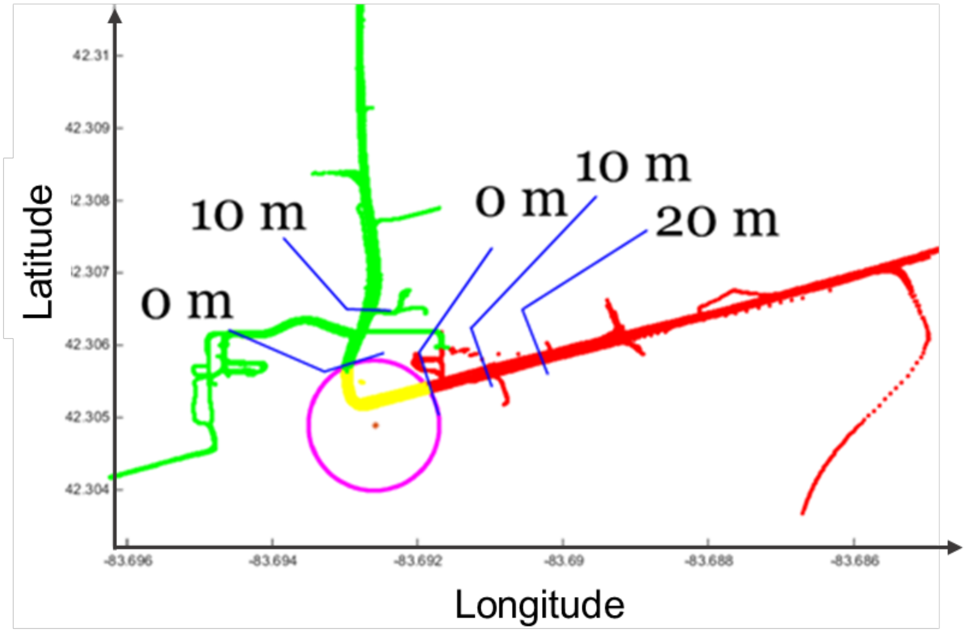
Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 6: Illustration of Clustering the Entering and Exiting Points

After clustering, the CV trips can be categorized according to the group of their entering points, and group of exiting points. For an intersection with N approaches, the trips can be categorized into N×(N-1) groups, excluding U-turn trips. For the example shown in Figure 6, there are 12 categories of trips. These categories were labeled as “N in, M out” category, where N is the group ID of an entry point of the trip, and M is the group ID of an exit point of the trip. For example, “2 in, 1 out” indicates CVs that entered through Group 2, and exited through Group 1, which represents westbound, right-turning (WB-RT) vehicles in Figure 6.

3.2 Geometry Estimation: Path Geometry and Lane Position Estimation

In the second step of the estimation process, the GPS data of trips from the same category were aggregated to estimate the geometry of the driving path. As mentioned in the previous section, based on the intersection circle, each of the trips can be divided into three segments: approaching, traversing, and leaving the circle. In this step, the path geometry was also estimated with these three segments. To do so, GPS points which were from the same part of trips belonging to a same category were grouped. Then, the GPS data were grouped into bins based on accumulated distance from either the entry point or the exit point. For the approach segment to the circle, the entry point was used as the reference point, because the first point of each of the CV trajectories could vary significantly with different V2I communication ranges for different CVs. Based on this reference point, the distance was accumulated working in the direction opposite the one the vehicle traveled. For the segment that a CV is traversing through the circle, either the entry or exit point can be used as the reference point. For the segment that a CV is leaving the circle, the exit point was used as the reference point for the calculation of accumulated distance. The illustration is shown in Figure 7.



Source: University of Michigan Transportation Research Institute. Used with Permission.

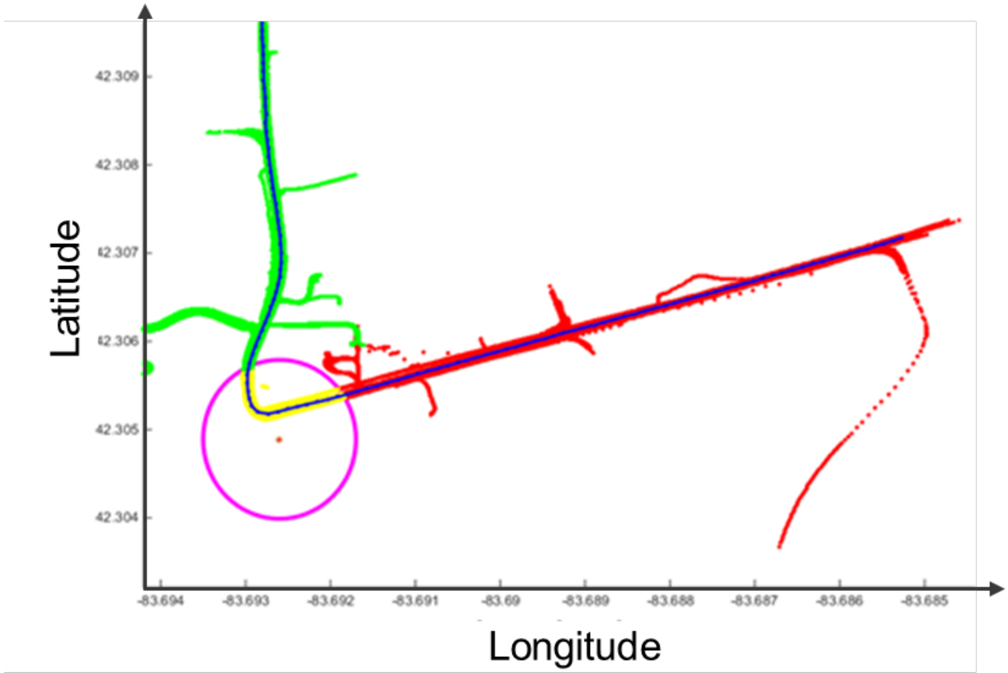
Figure 7: Illustration of Calculating Accumulated Distance and Grouping Data Based on Distance

For GPS data in a data bin based on accumulated distance, a weighted mean is calculated as a node point of the driving path geometry. The calculation is expressed in the following equation:

$$(\overline{Lat_i}, \overline{Lng_i}) = \left(\sum \frac{v_j}{\sum v_j} Lat_j, \sum \frac{v_j}{\sum v_j} Lng_j\right) \quad (2)$$

where: *i* is the index of the bin, *j* is the index of data point in the ith bin, v_j is the speed of j data point, (Lat_j, Lng_j) is the coordinate for the jth data point, and $(\overline{Lat_i}, \overline{Lng_i})$ is the coordinate for the node point.

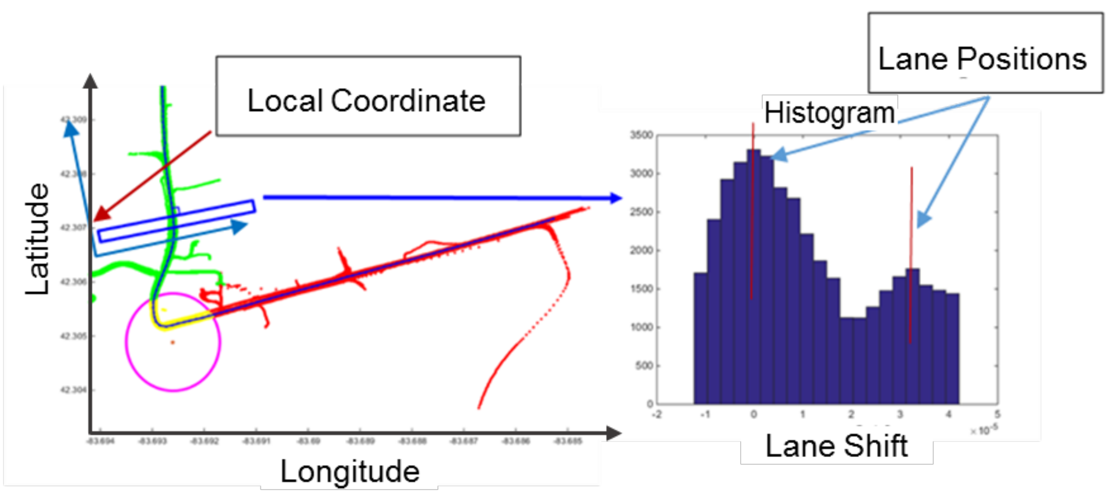
The calculation places greater weight on data points with higher speed, from CVs more likely traveling on main approaches. In this way the impact of data from CVs driving on a nearby driveway or parking lot, typically with low speed, was minimized. The calculation was repeated for each bin for all three parts of the driving path to obtain the path geometry for one trip category. The estimation of path geometry is shown in Figure 8.



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 8: Estimated Path Geometry Based on Weighted Mean

Next, the lane positions on the path were estimated. Along the path geometry, the relative shift of lane position to the estimated path geometry was estimated. The overall illustration of lane position estimation is shown in Figure 9. The histogram in the right portion of Figure 9 shows the distribution of vehicle positions in the local coordinate system. The histogram shows two peaks, which indicates the position of the lane center-line (shown by the two red lines in the figure). The objective of lane position estimation is to identify the number of peaks and the position of these peaks. The procedure is described below.



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 9: Illustration of Estimating Lane Position

A slice of data vertical to the path geometry was then transformed into individual local coordinates in which the vertical axis is the shift of CV position to the center of the approach.

The coordinate transformation was performed by finding the slope θ_i of the vehicle's travel direction of the *i*th slice from the approach geometry, from:

$$\theta_i = \arctan\left\{\frac{Lng_{a,i} - Lng_{a,i-1}}{Lat_{a,i} - Lat_{a,i-1}}\right\} + \frac{\pi}{2}$$

where: $Lat_{a,i}$, $Lng_{a,i}$ are the location of node points of an approach geometry.

The coordinate transformation matrix was calculated as:

$$M_c = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

The vehicle location in the local coordinate was next calculated as:

$$\begin{bmatrix} x_{s,i} \\ y_{s,i} \end{bmatrix} = M_c \times \begin{bmatrix} Lat_i - Lat_{a,i} \\ Lng_i - Lng_{a,i} \end{bmatrix}$$

where: $\begin{bmatrix} x_{s,i} \\ y_{s,i} \end{bmatrix}$ is the location of vehicle in the local coordinate system.

The number of lanes was determined based on the "spread" of vehicle locations, i.e., the difference between maximum and minimum value [13]. The calculation of spread $\boldsymbol{S}$ was expressed as:

$$\boldsymbol{S = max\left(x_{s,1}, x_{s,2}, \dots x_{s,i}\right) - min(x_{s,1}, x_{s,2}, \dots x_{s,i})} \text{ (3)}$$

where: $x_{s,i}$ is the horizontal distance of data point to the path node in an individual bin.

Then, the following criteria were used to determine the number of vehicle lanes, $\boldsymbol{N_{lanes}}$. The threshold was determined based on visual inspection of a small number of data samples with known lane numbers. Here we assumed that the maximum number of lanes is three for any one movement. This assumption can be relaxed with additional configuration for a larger number of vehicle lanes.

$$\boldsymbol{N_{lanes}} = \begin{cases} \boldsymbol{1, if\ S \leq 4.5 \times 10^{-5}} \\ \boldsymbol{2, if\ S \in (4.5, 7.5] \times 10^{-5}} \\ \boldsymbol{3, if\ S \geq\ 7.5 \times 10^{-5}} \end{cases} \text{ (4)}$$

Based on the number of lanes, we utilized the K-Means method to estimate the position of the individual lanes. The K-Means method is intended to minimize total distance to the so-called cluster centroids which are also the estimated lane position. This is as shown in the following:

$$\boldsymbol{Min \sum_i \sum_{x \in S_i} \left|\left|x - c_i\right|\right|^2} \text{ (5)}$$

where: x is the spread of a GPS point in a local coordinate; c_i is the centroid of the *i*th cluster, which is also the estimated lane position; S is a cluster; i is the cluster index.

The procedure of the K-Means algorithm is:

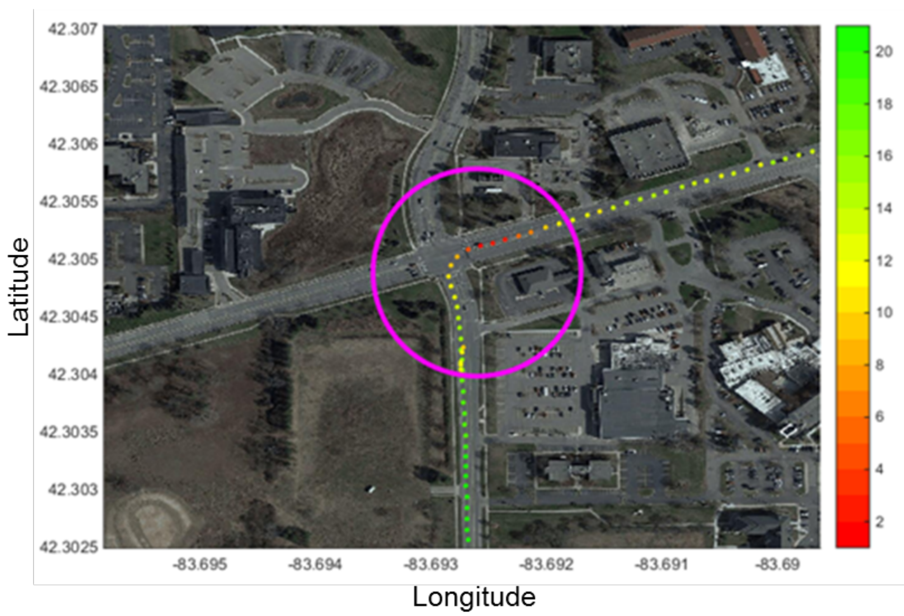
1. Randomly assign centroids for clusters
2. Assign points into clusters based on distances to each centroid
3. Recalculate new centroids
4. Calculate distance between new centroids with previous centroids. If distance is smaller than a threshold, end the algorithm. Otherwise, return to step 2.

In this project, the kmeans function of the Matlab (version 2015 b)[1] was used to perform K-Means clustering.

3.3 Post-process: Stop Bar Location Identification and Lane Combination

After the lane positions were estimated, the final phase was to organize the estimated data to be compatible with the MAP message. The first step was to identify the stop bar location so that the path geometry can be divided into ingress lanes and egress lanes. This project focused only on ingress lanes which are mandatory in the MAP message in the SAE J2735 standard [1], while leaving estimation of egress lanes for future investigation.

To identify the stop bar location, only points on the path geometry which are inside the intersection circle were considered as potential locations. Of these points, the one with minimum average speed was selected as the final stop bar location. Figure 10 illustrates this process.



Source: © 2016 Google, Inc. Used with Permission.

Figure 10: Illustration of Estimating Stop Bar Location

[1] http://www.mathworks.com/help/stats/kmeans.html

After identifying the stop bar location, the ingress lane portion was retrieved from the path geometry. To prepare an individual lane centerline geometry, lane position was connected within each bin along the driving path. Here, an assumption was made that the number of lanes upstream will not increase from the number of lanes at the stop bar location, i.e., the number of lanes at the stop bar location will always be maximum. Then starting with the stop bar, lane positions at adjacent bins were connected from downstream to upstream. Each connection was made based on the following equation to connect nearest points between the two bins:

$$\boldsymbol{m}^{*}(\boldsymbol{n},\boldsymbol{i}) = \underset{\boldsymbol{m}}{\boldsymbol{argmin}} \sqrt{\left(\boldsymbol{Lat}_{\boldsymbol{i},\boldsymbol{n}} - \boldsymbol{Lat}_{\boldsymbol{i}-\boldsymbol{1},\boldsymbol{m}}\right)^{\boldsymbol{2}} + \left(\boldsymbol{Lng}_{\boldsymbol{i},\boldsymbol{n}} - \boldsymbol{Lng}_{\boldsymbol{i}-\boldsymbol{1},\boldsymbol{m}}\right)^{\boldsymbol{2}}} \quad (6)$$

where: m is the index of lane position in $i-1$ bin, n is the index of lane position in i bin, and $m^{*}(n,i)$ is the index of lane position in bin $i-1$ that will be connected to n lane position of i bin.

There were cases when the number of lanes decreased along the path. In these cases, the number of lane positions was reduced across adjacent bins. Lane positions with maximum distance to the adjacent bin were not connected, and the associated lane centerline geometries were stopped. After connecting the nodes to form the lane centerline geometry, a local regression smoothing was then applied to further reduce outliers in the lane centerline geometry.

4 Evaluation and Validation of Map Estimation Procedure

This section presents results of the accuracy evaluation conducted by comparing the estimated map with the reference map. The comparison was conducted at two levels. The first level compared the structure of the intersection regarding the number of lanes for each approach and the vehicle movement. The second level compared the difference in individual lane centerline geometries between the estimated map and the reference map.

Based on distance between two GPS coordinates, two types of measurement were used to identify the average and worst accuracy of the estimated map. The first measurement was the Mean Distance of node points between estimated lane centerline geometries and surveyed lane centerline geometries. The second measurement was the maximum distance of node points between estimated lane centerline geometries and surveyed lane centerline geometries. The maximum distance is also called the Hausdorff Distance [7]. The calculations are as follows:

$$\boldsymbol{Mean\ Distance}: \boldsymbol{D_{mean}(X, Y)} = \underset{x \in X}{\textbf{mean}}\{\underset{y \in Y}{\textbf{min}}\{\boldsymbol{d(x, y)}\}\} \quad (7)$$

$$\boldsymbol{Max\ Distance}: \boldsymbol{D_{max}(X, Y)} = \underset{x \in X}{\textbf{max}}\{\underset{y \in Y}{\textbf{min}}\{\boldsymbol{d(x, y)}\}\} \quad (8)$$

where: X, Y are lane centerline geometries of the estimated map and reference map, respectively, and x, y are GPS coordinates of individual node points on geometry X, and geometry Y respectively.

The distance between two GPS points was calculated using the haversine formula as:

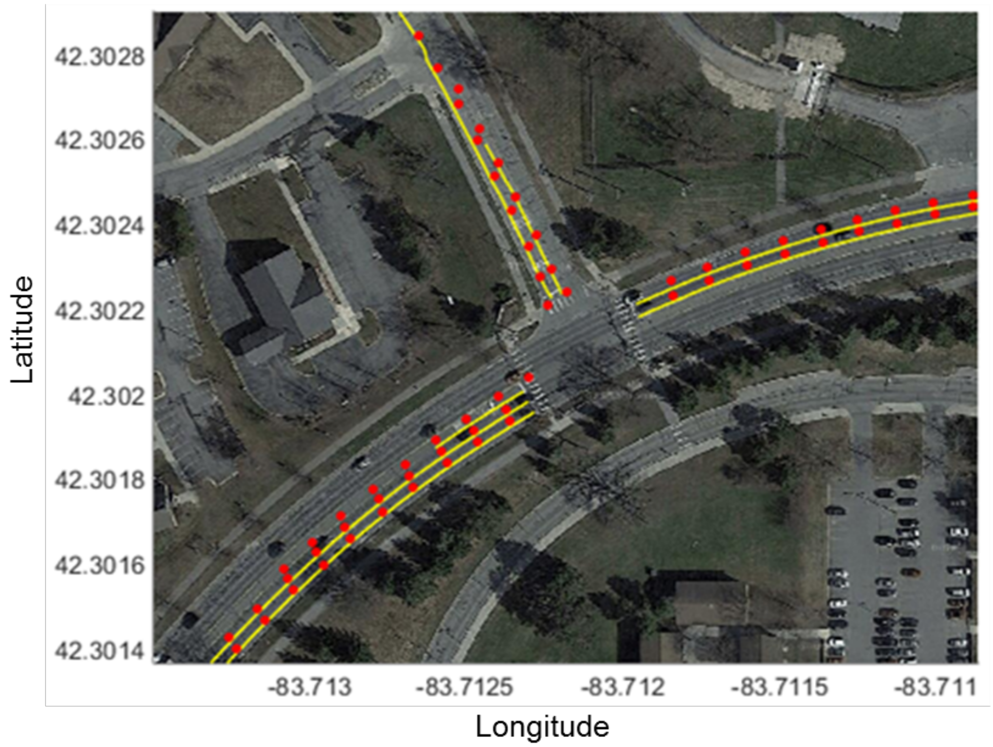
$$\begin{cases} \boldsymbol{d}\left([\boldsymbol{Lat_x}, \boldsymbol{Lng_x}], [\boldsymbol{Lat_y}, \boldsymbol{Lng_y}]\right) = \boldsymbol{2R} \times \textbf{arcsin}(\sqrt{\boldsymbol{h}}) \\ \boldsymbol{h} = \textbf{sin}^2\left(\frac{\boldsymbol{Lat_y} - \boldsymbol{Lat_x}}{2}\right) + \textbf{cos}(\boldsymbol{Lat_x})\,\textbf{cos}(\boldsymbol{Lat_y})\,\textbf{sin}^2\left(\frac{\boldsymbol{Lng_y} - \boldsymbol{Lng_x}}{2}\right) \end{cases} \quad (9)$$

4.1 Different Geodesic Datum and Coordinate Shift

Before comparing the estimated map with the reference map, an issue related to coordinate system shift needs to be addressed due to different geodesic datum used for the reference map and for the CV system.

Currently, there are two geodetic data being used most frequently for localization, World Geodetic System 1984 (WGS84) and North America Datum 1983 (NAD83). WGS84 is an Earth-centered Earth-fixed geodetic datum, which is not fixed to any tectonic plate. NAD83 Datum, however, is fixed to the North American Tectonic plate. Due to the tectonic forces of the Earth, in WGS84, the coordinates of the North American tectonic plate will be continuously moving, but will be static in NAD83. Therefore, the WGS84 and NAD83 will gradually shift from each other over time, or under severe events of tectonic motion (e.g., an earthquake).

In geo-referencing survey, the NAD83 are mostly used so that the surveyed coordinates will not shift over time with tectonic movement, while commercial GPS units use WGS84 Datum. In this project, the data collected from OBUs in CVs are all based on WGS84. However, the reference map obtained from geo-referencing survey is based on NAD83. As a result, a systematic shift exists between the estimated map and the reference map, which are in two different datum. The shift is illustrated in Figure 11, with red dots for the estimated lane centerline geometry and yellow solid lines for lane centerline geometry from the reference map.



Source: © 2016 Google, Inc. Used with Permission.

Figure 11: Shift between Estimated Map with Reference Map

The total shift was calculated for the Plymouth-Traverwood intersection to estimate the shift between the two datum in the investigated area. Shifts for latitude and longitude were enumerated to adjust the reference map. Then, the total distances between estimation lane centerline geometries and surveyed lane centerline geometries was calculated for each enumerated shift. The optimal shift adjustment will be the one with minimum distance between estimated geometries with surveyed geometries.

Mathematically, the shift is estimated as:

$$\Delta^{*} = \underset{\Delta}{\operatorname{argmin}} \sum_{x \in X} \min_{y \in Y} \{ d(x, \hat{y}(y, \Delta)) \}$$

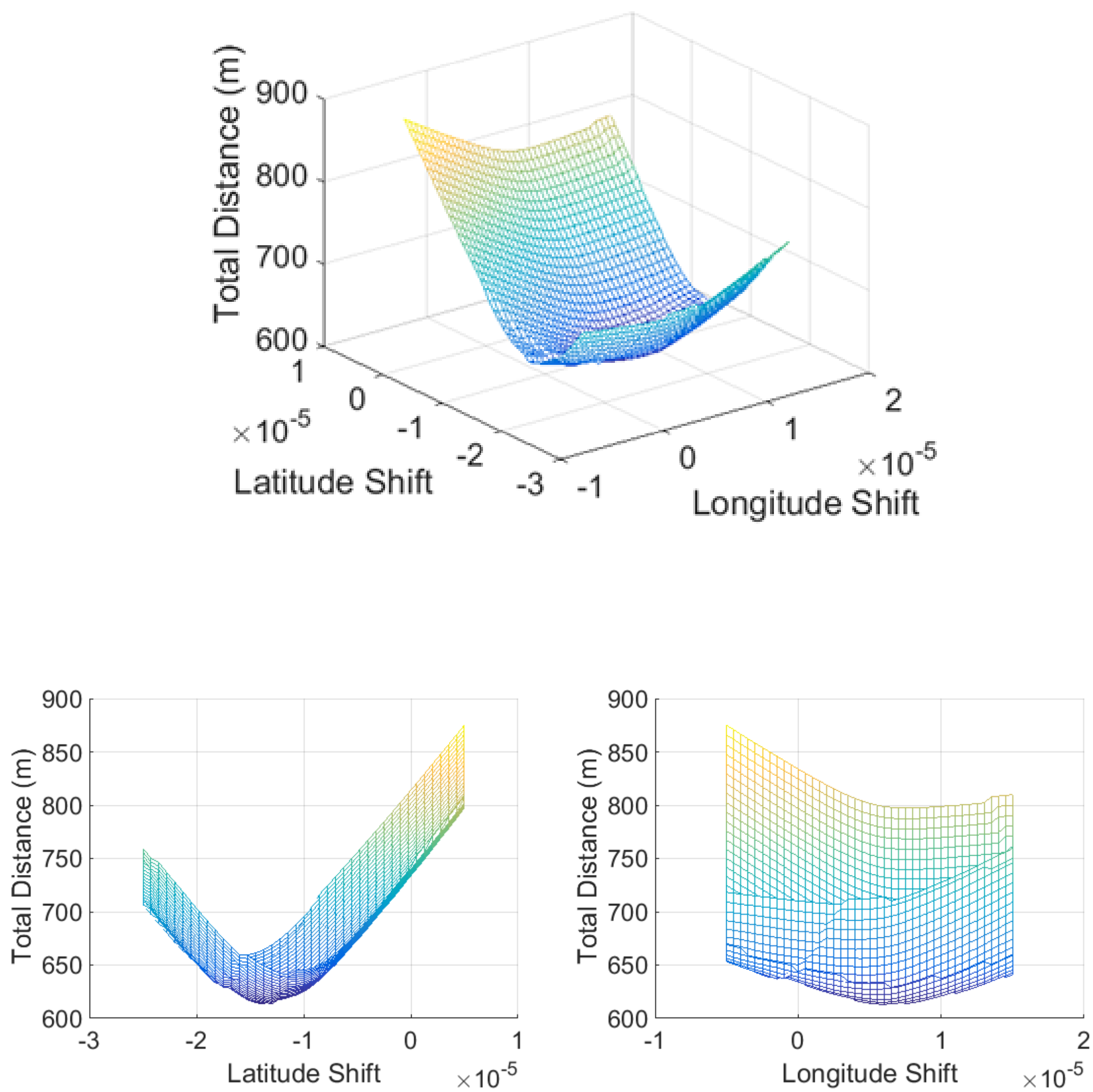
$$\Delta = (\Delta Lat, \Delta Lng)$$

$$\hat{y}(y, \Delta) = (Lat_y + \Delta Lat, Lng_y + \Delta Lng)$$

where: Δ^{*} is the optimal estimation of the shifts between the reference map and the estimated map and $\hat{y}(y, \Delta)$ is the GPS coordinates from the reference map with shift adjusted.

The total distances on different shifts are shown in Figure 12. The shift was estimated as:

$$\Delta^{*} = (\Delta Lat, \Delta Lng) = (-1.4 \times 10^{-5}, 0.6 \times 10^{-5})$$



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 12: Illustration of Identifying Optimal Estimation of the Shift

The two measurements were then calculated with the shift adjusted. The evaluation of the estimation is shown for each intersection, in the next four subsections of this chapter.

4.2 Evaluation of Accuracy

4.2.1 Fuller-Cedar Bend (Simple Intersection)

The estimated map and reference map for the Fuller-Cedar Bend intersection are shown together in Figure 13, with red dots for the estimated lane centerline geometry and yellow solid lines for lane centerline geometry from the reference map.

Source: © 2016 Google, Inc. Used with Permission.

Figure 13: Estimated Map and Reference Map for Fuller-Cedar Bend

The comparison results are summarized in Table 4. As mentioned earlier, two levels of evaluation were conducted. The first level was the intersection structure evaluation that validates a number of vehicle lanes estimated from the algorithm. The second level was to evaluate the accuracy of the estimated lanes with surveyed lanes. For accuracy of estimated lane, the mean distance and maximum distance between estimated lane centerline geometry with surveyed geometry are calculated based on Equation (7) and (8). The yellow cells indicate that this is a false identification of the number of lanes for an approach.

Table 4: Comparison of Intersection Structure and Accuracy for Fuller-Cedar Bend

*Estimated number of lanes (True number of lanes); ** Mean Distance (Maximum Distance) between estimated and survey lane centerline geometries;

Performance Measure		WB	EB
Intersection Structure	LT lane	0(1)*	0(1)
	Through lane	2	2
	RT lane	0(1)	0(1)
	LT, through shared	0	0
	RT, through shared	0(1)	0(1)
	LT, RT shared	0	0
Accuracy of Estimated Geometry of Lane Centerline	Distance (unit: m)	1.0 (2.4)**	0.4 (1.0)
		0.3 (0.8)	0.9 (2.6)

False Identification

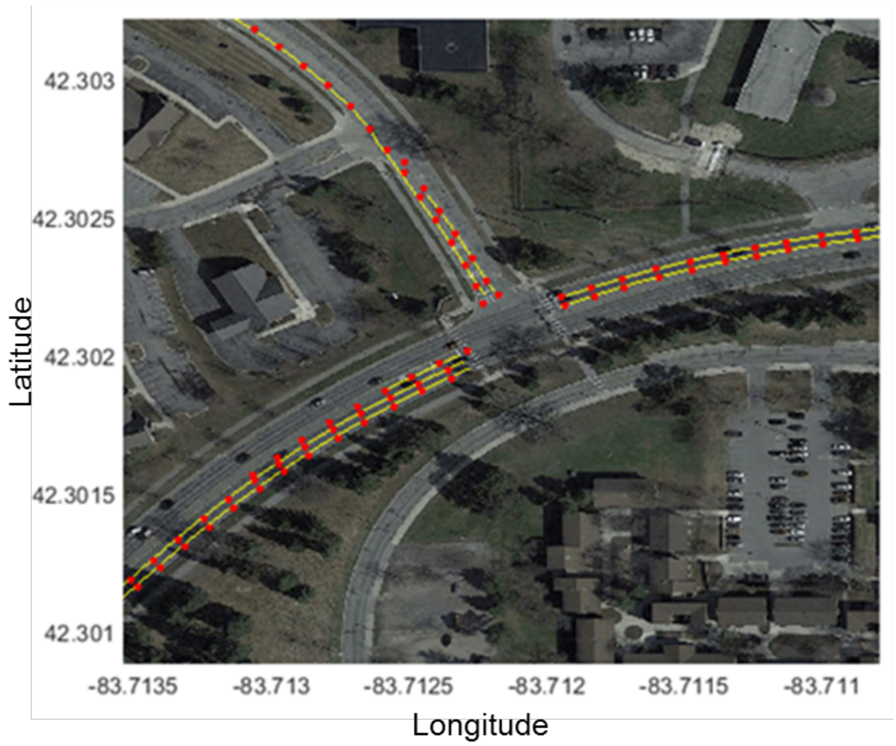
Source: University of Michigan Transportation Research Institute. Used with Permission.

As shown in Table 4, the developed procedure was able to correctly identify the lanes for WB and EB through movement, but not able to identify the rest of the lanes. This is because there are insufficient trips entering/exiting the SB approach, which is a small driveway. The NB approach contains the access to a parking lot. The vehicle trajectories collected from this approach scattered widely within the parking lot and could not be used to identify traffic lanes. Hence, only WB and EB approaches could be estimated from the data.

The accuracy is calculated for each identified lane, and is listed in the lower half of the table. The mean distances between estimated lanes centerline geometry and reference lane centerline geometry are around 0.3m-1m, while the maximum distances are around 0.8m-2.6m.

4.2.2 Plymouth-Traverwood (Moderately Complex Intersection)

The estimated map and reference map for Plymouth-Traverwood are shown in Figure 14.



Source: © 2016 Google, Inc. Used with Permission.

Figure 14: Estimated Map and Reference Map for Plymouth-Traverwood

The evaluation result is summarized in Table 5. For the Plymouth-Traverwood intersection, the lane structure has been correctly identified. As shown in the table, accuracy similar to the prior intersection was found at this intersection, with mean distance approximately 0.2 to 0.9 meters and maximum distance approximately 0.5 to 1.6 meters.

Table 5: Comparison of Intersection Structure and Accuracy for Plymouth-Traverwood

Measurement Description		WB	EB	SB
Intersection Structure	LT lane	0	1	1
	Through lane	2	2	0
	RT lane	1	0	1
	LT, through shared	0	0	0
	RT, through shared	1	0	0
	LT, RT shared	0	0	0
Accuracy of Estimated Geometry of Lane Centerline	Distance (unit: m)	0.1 (1.1)	0.2 (0.5)	0.9 (1.6)
		0.2 (0.9)	0.3 (0.8)	0.2 (1.0)
			0.5 (0.8)	

Source: University of Michigan Transportation Research Institute. Used with Permission.

4.2.3 Plymouth-Nixon (Moderately Complex Intersection)

The estimated map and reference map for the Plymouth-Nixon intersection are shown in Figure 15.

Source: © 2016 Google, Inc. Used with Permission.

Figure 15: Estimated Map and Reference Map for Plymouth-Nixon

The evaluation result is summarized in Table 6. For this intersection, the lane structure was correctly identified, except for the lanes in the NB approach. The NB approach, consisting of two lanes, was mistakenly identified as a 1-lane approach. The main reason is likely due to the fact that the lane marker was blurred in this approach, making it difficult for drivers to identify the two lanes correctly. For this intersection, the mean distances were found to be about 0.2 to 0.6 meters and maximum distances were about 0.4 to 2.9 meters.

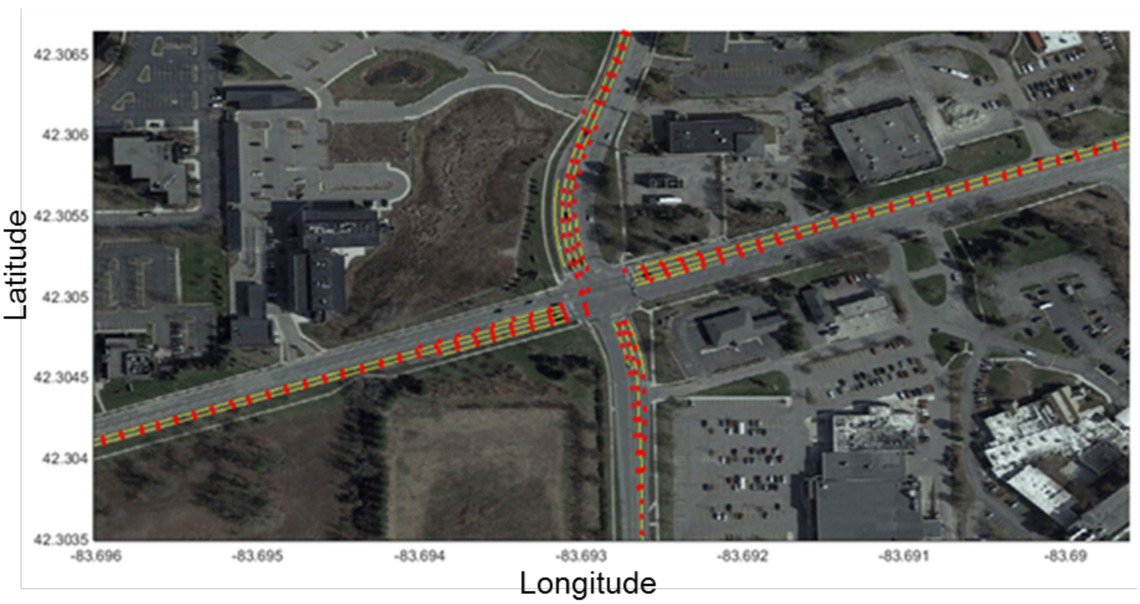
Table 6: Comparison of Intersection Structure and Accuracy for Plymouth-Nixon

Measurement Description		WB	NB	EB	SB
Intersection Structure	LT lane	1	1	1	1
	Through lane	2	1	2	1
	RT lane	1	1	1	1
	LT, through shared	1(0)	1(0)	0	0
	RT, through shared	1	1	1	1
	LT, RT shared	0	0	0	0
Accuracy of Estimated Geometry of Lane Centerline	Distance (unit: m)	0.6 (2.9)		0.3 (0.8)	0.4 (0.7)
		0.2 (0.4)	0.8(1.9)	0.4 (0.9)	0.6 (1.2)
				0.5 (1.0)	

Source: University of Michigan Transportation Research Institute. Used with Permission.

4.2.4 Plymouth-Green (Complex Intersection)

The estimated map and reference map for Plymouth-Nixon are shown in Figure 16.



Source: © 2016 Google, Inc. Used with Permission.

Figure 16: Estimated Map and Reference Map for Plymouth-Nixon

The evaluation result is summarized in Table 7. At this intersection, there was one lane not successfully identified, which serves dedicated RT movement at the SB approach. The identification was unsuccessful because there were only seven CV trips observed in the one month of data, an insufficient number for lane centerline geometry estimation.

For the accuracy measurements, the mean distances are found to be approximately 0.1 to 1.6 meters and maximum distances are approximately 0.3 to 2.2 meters.

Table 7: Comparison of Intersection Structure and Accuracy for Plymouth-Green

Measurement Description		WB	NB	EB	SB
Intersection Structure	LT lane	1	1	1	2
	Through lane	2	1	3	2
	RT lane	1	1	1	0(1)
	LT, through shared	0	0	0	1
	RT, through shared	0	0	1	0(1)
	LT, RT shared	0	0	0	0
Accuracy of Estimated Geometry of Lane Centerline	Distance (unit: m)	0.1 (0.3)	0.2 (0.2)	0.2 (0.6)	0.3 (0.9)
		0.7 (1.9)	0.1 (0.6)	0.4 (0.7)	0.6 (1.6)
		0.3 (1.1)	1.4 (1.7)	0.3 (1.0)	0.7 (1.9)
		1.6 (2.2)		0.8 (1.4)	

Source: University of Michigan Transportation Research Institute. Used with Permission.

4.2.5 Washtenaw-Huron (Complex Intersection)

The estimated map and reference map for Washtenaw-Huron are shown in Figure 17.

Source: © 2016 Google, Inc. Used with Permission.

Figure 17: Estimated Map and Reference Map for Washtenaw-Huron

The evaluation result is summarized in Table 8. At this intersection, one lane was not successfully identified, which is the shared RT and through lane of the SB approach. It was found that the data was for through traffic only, and was distributed on the left through lane instead of both of the lanes. This could be due to possible construction and lane blockage during the data collection period for this particular intersection. For the accuracy measurements, the mean distances between estimated lane centerline geometries to the surveyed lane centerline geometries were found to be approximately 0.2

to 1.3 meters and maximum distances between estimated lane centerline geometries to the surveyed geometries were approximately 0.3 to 2.2 meters.

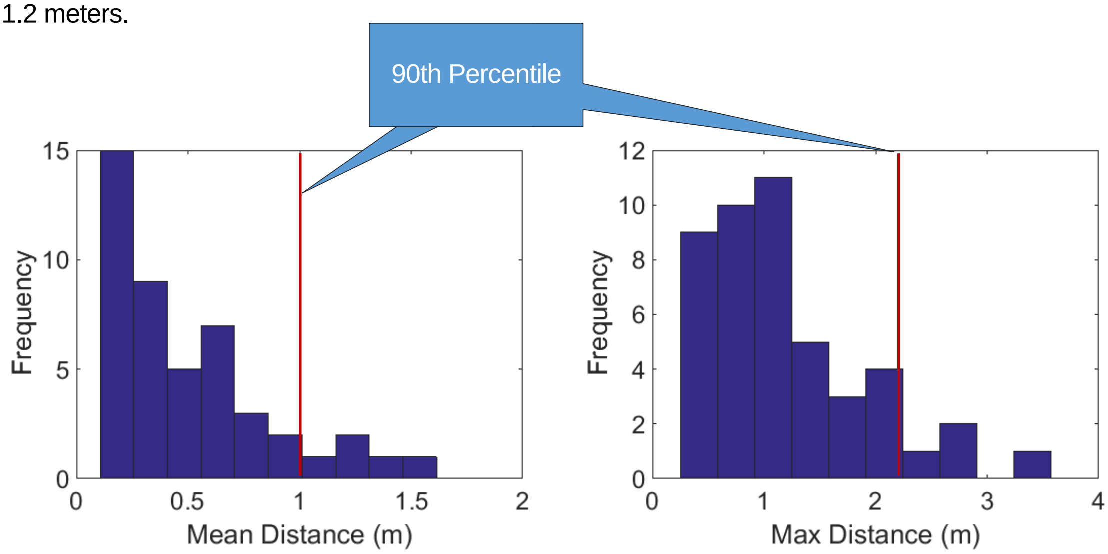
Table 8: Comparison of Intersection Structure and Accuracy for Washtenaw-Huron

Measurement Description		WB	NB	EB	SB
Intersection Structure	LT lane	1	2	1	2
	Through lane	2	2	2	1(2)
	RT lane	1	1	1	0(1)
	LT, through shared	0	0	0	0
	RT, through shared	1	1	1	0(1)
	LT, RT shared	0	0	0	0
Accuracy of Estimated Geometry of Lane Centerline	Distance (unit: m)	0.3 (0.6)	1.2 (2.2)	0.3 (0.6)	0.1 (0.3)
		0.2 (1.5)	0.8 (1.3)	0.5 (1.0)	0.3 (1.6)
		0.5 (0.9)	0.2 (0.5)	0.1 (0.6)	1.3 (3.5)
			0.8 (1.2)		

Source: University of Michigan Transportation Research Institute. Used with Permission.

4.2.6 Summary of Measurements of Accuracy

The accuracy results for all five intersections are summarized in this section. The histograms of both mean distances and maximum distances are shown in Figure 21. Overall, the mean distances ranged from 0 to 1.5 meters, with the majority (90%) under 1 meter. The maximum distances ranged from 0.5 to 3.5 meters, with the majority (90%) under 2.2 meters. The average value of mean distance for all estimated maps was found to be 0.5 meters, while the average maximum distance was found to be 1.2 meters.



	Mean Distance (m)	Max Distance (m)
Average Value	0.5	1.2
90% Percentile Value	1.0	2.2

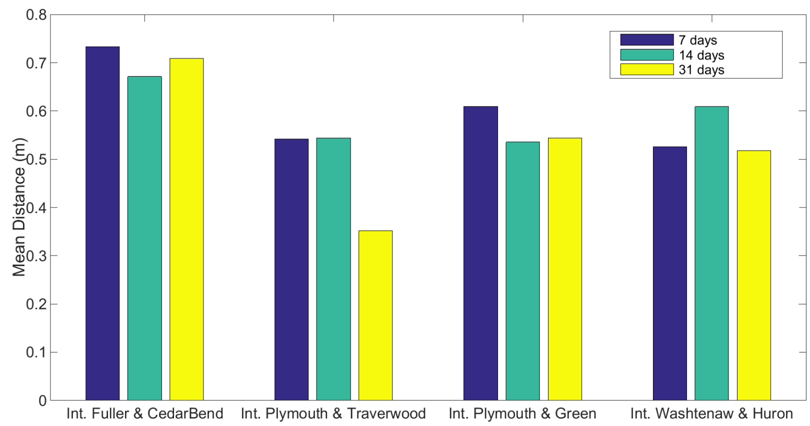
Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 18: Summaries of Mean Distances and Max Distances

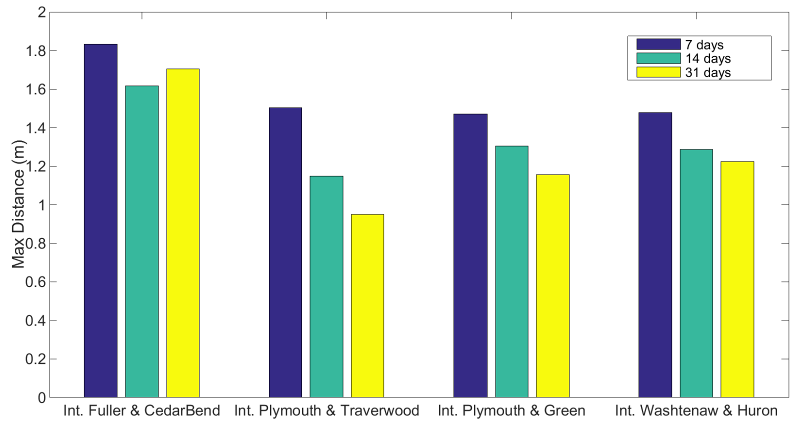
4.2.7 Impact of Data Size on Estimation Accuracy

To evaluate the impact of data size on the estimation accuracy, data with three different periods were selected from the 1-month data for map estimation. The three periods were: 7 days, 14 days, and 31 days. For each intersection, the mean and maximum distances between estimated lane centerline geometries to the surveyed geometries in three different periods are shown in Figure 19. However, only the results for four of the selected intersections are actually shown in the figure. The estimation for Plymouth-Nixon is not available for 7-day and 14-day time periods due to a lack of trips on the NB approach of the intersection. Hence, the analysis of the Plymouth-Nixon intersection was excluded from Figure 19.

As can be seen from Figure 19, overall the mean distances ranged from 0.35 to 0.75 meters across the selected intersections. Although slight improvements can be seen as the number of days increased, the improvement is minimal and inconsistent. This would suggest that seven days of data may be sufficient for estimation for these selected intersections. However, improvements of the maximum distance are observed with increasing number of days for three of the four intersections, indicating the potential that increasing data size would be helpful to improve the robustness of the estimation.



(A)



(B)

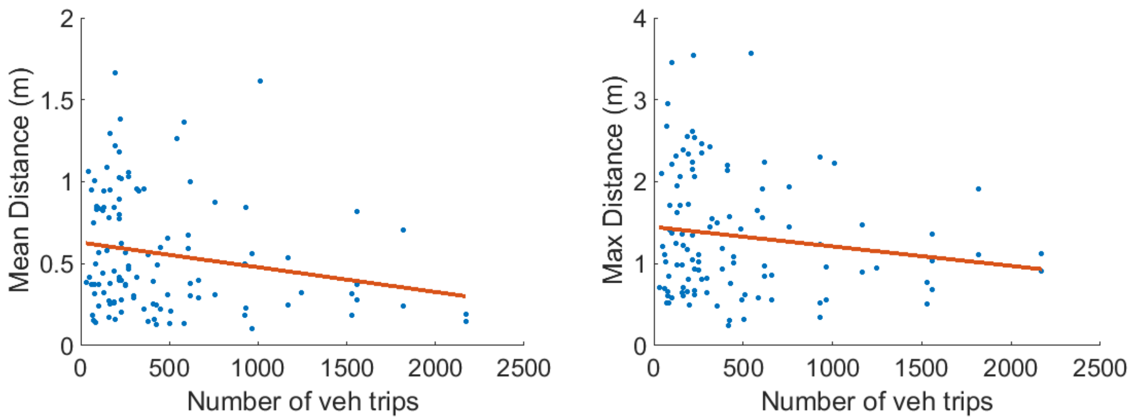
	Number of Vehicle Trips		
Intersection	**7 Days**	**14 Days**	**31 Days**
Fuller & Cedar Bend	273	498	1,073
Plymouth & Traverwood	1,022	2,478	6,101
Plymouth & Green	1,641	3,931	9,701
Washtenaw & Huron	899	2,225	5,566

(C)

Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 19: Mean Distances (A), Maximum Distances (B) and Number of Vehicle Trips (C) Using Data with Three Different Periods

To further shed light on the impact of data size on estimation accuracy, the relationships regarding the number of vehicle trips used for estimation with the mean distance and maximum distance were examined. The plots are shown in Figure 20.

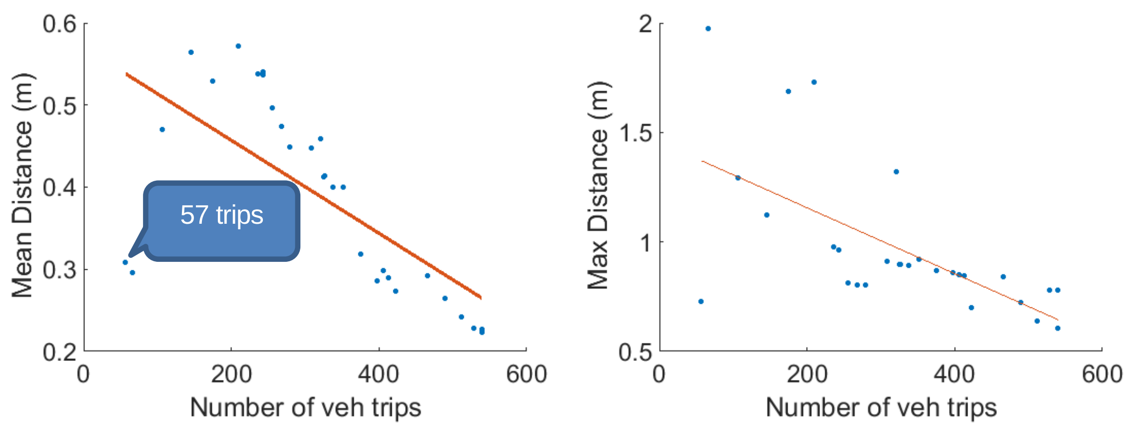


Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 20: Relation of Mean Distances (Left) and Max Distances (Right) With Number of Vehicle Trips Used for Estimation

As shown in Figure 20, a slight descending trend can be found on both the mean distance and maximum distance when the number of vehicle trips used for estimation increases. Regarding the relation between accuracy with the data size shown in Figure 20, a slight improvement can be observed for both mean distance and maximum distance when the number of vehicle trips used for estimation increases. Here, the minimum number of trips needed for successful estimation is found to be 32. However, with a same number of vehicle trips used, the range of the distances still varied from 0.5 to 1.5 meters, indicating there are other factors affecting estimation accuracy. These factors may include curvature of the path, nearby driving environments (e.g., parking lots and private driveways) and may still need to be evaluated case by case.

To take a closer look at the relationship between data size and accuracy, a lane in the Fuller-Cedar Bend intersection was selected. This intersection has a simple geometry, and the leftmost through lane in the EB approach was chosen for the analysis. Lane centerline geometry was estimated using data collected during different time periods, ranging from two days to the full month of data obtained for the project. The result is shown in Figure 21. For this intersection, the minimum number of days required for successful estimation was two days with approximately 60 trips. As shown in Figure 21, improvement can be observed for this case when number of trips increases, with reduction of both mean distance and maximum distance, indicating the validity of improving estimation accuracy by increasing amount of data used for estimation.



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 21: Relation of Mean Distances (Left) and Maximum Distances (Right) with Number of Vehicle Trips for Fuller-Cedar Bend

5 Automatic SPaT-MAP Association Using SPaT and BSM Data

This section presents details of the procedure for SPaT-MAP association. As defined by the SAE J2735 standard, the SPaT message conveys traffic signal status for the intersection and is broadcast together with MAP messages by RSUs. For CV applications, the MAP message indicates which lane the CV is on, based on positioning information, and what maneuvers are allowed in the lane, while the SPaT message helps to indicate the signal status for the associated lane and vehicle maneuvers. Only with correct interpretations of both messages and lane-level positioning can the on-board CV applications (e.g. red light violation warning) be aware of the correct traffic signal status and operate properly. Therefore, it is critical to correctly associate vehicle lanes in the MAP message with signal phases in the SPaT message.

5.1 Technical Approach

The problem of the SPaT-MAP association is to identify the signal phase ID for a selected vehicle lane, using data of signal status sequence and BSMs. For simplicity in the project, it was assumed that only protected phases will be considered.

Mathematically, the list of start times for the Green signal, Yellow signal and Red signal for different phase i, was formulated as:

$$\{G_{i,1}, G_{i,2} \dots, G_{i,n} \dots\}, \{Y_{i,1}, Y_{i,2} \dots, Y_{i,n} \dots\}, \{R_{i,1}, R_{i,2} \dots, R_{i,n} \dots\}$$

where: G, Y, R are the start of the green, yellow and red signal, respectively, and i is index of signal phase, and n is index of signal cycle. Here, the green start of a particular signal phase was selected as the start of signal cycle.

From BSM data, we can calculate the time when the vehicle traverses the intersection, as:

$$\{A_{j,1}, A_{j,2} \dots, A_{j,m}, \dots\}$$

where: A is the time when a vehicle passed the stop bar, j is the index of vehicle movement, and m is the index of vehicle arrivals.

The problem here is to associate signal phase for each movement, i.e., to find the following mapping:

$$U(j) = i, if\ movement\ j\ is\ allowed\ during\ green\ of\ phase\ i$$

To determine $U(i)$, a green arrival indicator is first defined from:

$$\boldsymbol{I(A_{j,m}, i) = \begin{cases} 1, if\ A_{j,m} \in [G_{i,k}, Y_{i,k}] \\ 0, O.W. \end{cases}, where\ k\ is\ the\ index\ of\ cycle\ when\ A_{j,m}\ occurs} \quad (10)$$

Then, total green arrivals for a selected phase I and movement j, $N_g(i, j)$, can be calculated as:

$$N_g(i,j) = \sum_m I(A_{j,m}, i) \quad (11)$$

Next, the associated signal phase for movement j can be determined as the one yielding maximum green arrivals, as:

$$U(j) = \operatorname*{argmax}_i N_g(i,j) \quad (12)$$

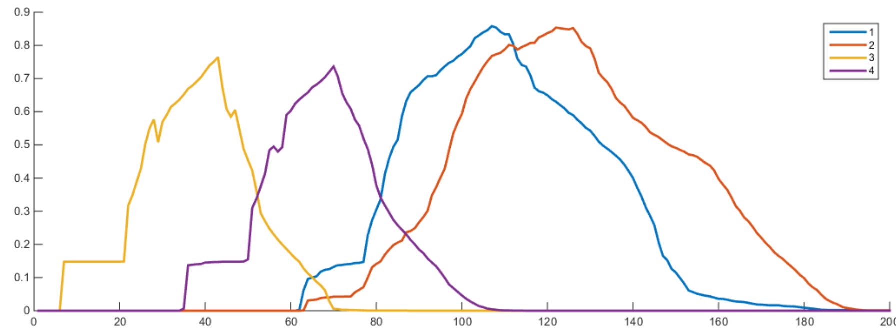
5.2 Investigation Example

The Green Road and Plymouth Road intersection was selected to investigate the feasibility of SPaT-MAP association. The objective is to infer the associated signal phase ID for vehicle lanes. In the investigated case, only signal phases 1-4 are available in the SPaT data. Therefore, the analysis will only focus on these four signal phases. The illustration of the problem is shown in Figure 22, as well as the layout and traffic signal information of the intersection.

Source: Map images © 2016 Google, Inc. Used with Permission.
Traffic signal information from University of Michigan Transportation Research Institute. Used with Permission.

Figure 22: Layout and Signal Phase Diagram of Investigated Intersection

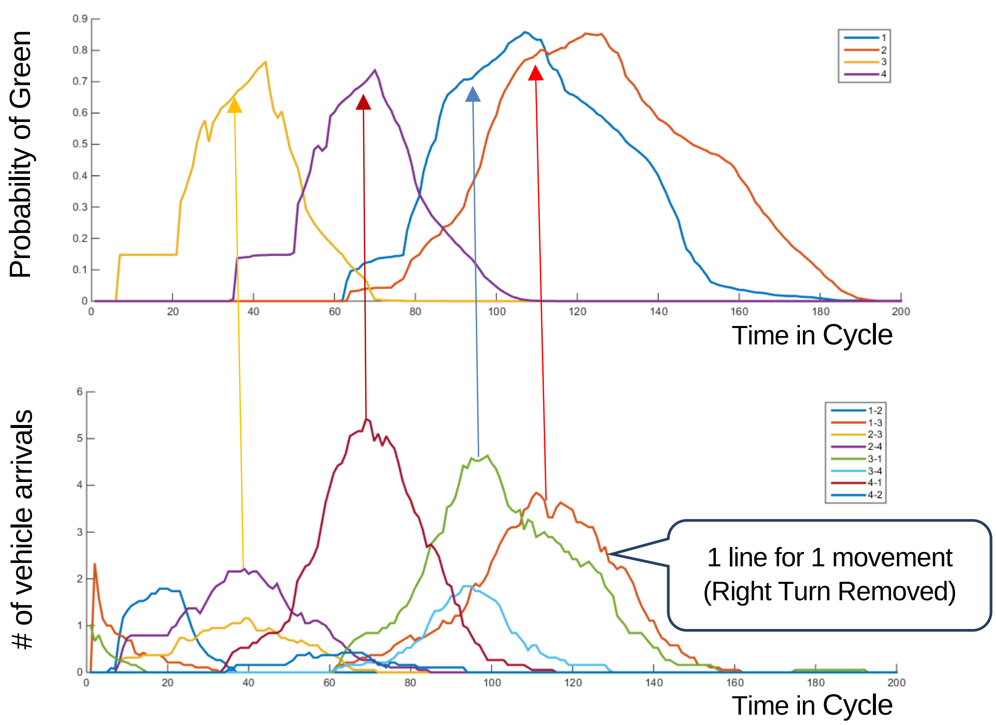
SPaT and BSM data collected during the afternoon peak hours, 3:00 PM to 7 PM, for two weeks were used for this SPaT association analysis. The signal data are visualized in Figure 23.



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 23: Probability of Green Signal on Time in Cycle for the Four Phases

For each vehicle movement, we will identify the signal phase that yields maximum green arrivals. The association between signal phases with vehicle arrivals is illustrated in Figure 24.



Source: University of Michigan Transportation Research Institute. Used with Permission.

Figure 24: Illustration of Associating Signal Phase with Traffic Arrivals

In Figure 24, the upper figure shows the probability of green on-time within a cycle for all four applicable phases, and the lower figure shows arrival volume during 5-second intervals for all movements at the intersection. It can be seen that arrival volumes also show similar patterns with the green probability profile. For example, a clear match of pattern between signal phase 4 and volume for movement 4-1 (group 4 in, group 1 out, i.e., southbound left turn, SB-LT), indicates an association between the signal phase and movement. The estimated association between vehicle movements with signal phases are summarized in Table 9.

Table 9: Movement Phase Association

Movement	WB-LT	WB-TH	SB-LT	SB-TH	EB-LT	EB-TH	NB-LT	NB-TH
Estimated Phase ID	N/A	1	4	4	N/A	2	3	3
Real Phase ID	N/A	1	4	4	N/A	2	3	3

Source: University of Michigan Transportation Research Institute. Used with Permission.

As shown in the table, the correct signal phase ID has been identified for all of the movements. The validity of the proposed SPaT-MAP association procedure is therefore confirmed.

6 Conclusion

In this project, an approach for automatically generating intersection maps using BSMs received by the RSUs was developed. The proposed approach consists of three steps to estimate an intersection map for automated creation of a MAP message: pre-processing to clean and categorize BSM data; estimating approach and lane centerline geometry using categorized data; and post-processing to reorganize lane geometries. The proposed approach was applied to five intersections selected from the SPMD Project database to estimate intersection maps. The estimated maps were then compared with reference maps obtained from LIDAR surveys to evaluate the effectiveness and accuracy of the proposed approach.

Overall, the estimation approach was able to correctly identify the main structure of the intersections regarding the number of lanes for different movements. For all of the five intersections, with 48 traffic lanes in total, five lanes were not correctly identified. The primary reason for these incorrect identifications was a lack of BSM data in these lanes. For correctly identified lanes, the accuracy is within 0.5 meters on average, with the majority (90%) within 1.0 meter.

In addition to map generation, the feasibility of automatic SPaT-MAP association, i.e., connecting vehicle lanes with traffic signal phases, was also investigated using SPaT and BSM data. A simple procedure was developed to estimate the association based on the logic of matching vehicle departures with green signal, i.e., maximizing green departures from the SPaT and BSM data. A case study was then conducted using data from one of the five intersections identified for analysis. For the selected intersection, the proposed procedure was able to correctly associate vehicle movements to all applicable signal phases. The applicability of automatic SPaT-MAP association was confirmed.

Overall, this project demonstrated the feasibility and significant potential of using BSM and SPaT data to generate intersection maps and phase-lane mapping automatically, thereby automating the process of generating the MAP message. The proposed procedure could be useful for V2I system deployment at the traffic intersections in the near future.

This project was a first step in mining CV data for infrastructure systems. A number of research efforts could be pursued in the future. While the current project focused mainly on static map generation using varying amounts of data (1 week to 1 month), in reality, CV applications could be dramatically affected by temporary lane closure or geometry modification due to road construction or traffic incidents. It would be beneficial to investigate the feasibility of using CV data to identify these dynamic changes to a map. Another potentially interesting application would be to integrate map estimation with estimation of real-time traffic conditions. Traffic conditions are frequently critical constraints for smooth and safe road driving. To generate and organize both map and traffic information for developing effective and robust CV applications could also be a promising research direction.

7 References

1. Std, S. A. E. J2735 Revised APR2015. (2015-04). SAE International. Dedicated short-range communications (DSRC) message set dictionary.
2. Ahmed, M., & Wenk, C. (2012). Probabilistic street-intersection reconstruction from gps trajectories: approaches and challenges. In *Proceedings of the Third ACM SIGSPATIAL International Workshop on Querying and Mining Uncertain Spatio-Temporal Data* (pp. 34-37). ACM.
3. Cao, L., & Krumm, J. (2009). From GPS traces to a routable road map. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 3-12). ACM.
4. Guo, T., Iwamura, K., & Koga, M. (2007). Towards high accuracy road maps generation from massive GPS traces data. In *2007 IEEE International Geoscience and Remote Sensing Symposium*.
5. Liu, X., Zhu, Y., Wang, Y., Forman, G., Ni, L. M., Fang, Y., & Li, M. (2012). Road recognition using coarse-grained vehicular traces. *HP Labs, HP Labs2012*.
6. Schroedl, S., Wagstaff, K., Rogers, S., Langley, P., & Wilson, C. (2004). Mining GPS traces for map refinement. *Data mining and knowledge Discovery*, *9*(1), 59-87.
7. Ahmed, M. et al., 2014. A comparison and evaluation of map construction algorithms using vehicle tracking data. *GeoInformatica*.
8. Chen, X., Kohlmeyer, B., Stroila, M., Alwar, N., Wang, R., & Bach, J. (2009). Next generation map making: geo-referenced ground-level LIDAR point clouds for automatic retro-reflective road feature extraction. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 488-491). ACM.
9. Kammel, S., & Pitzer, B. (2008). Lidar-based lane marker detection and mapping. In Intelligent Vehicles Symposium, 2008 IEEE (pp. 1137-1142). IEEE.
10. Huang, A. S., Moore, D., Antone, M., Olson, E., & Teller, S. (2009). Finding multiple lanes in urban road networks with vision and lidar. Autonomous Robots, 26(2-3), 103-122.
11. A Benchmark for Data Fusion (SSI's survey project with MDOT). http://www.pobonline.com/articles/96492-a-benchmark-for-data-fusion (Accessed on Mar, 2016)
12. SSI Mobile LiDAR (MoLi) Service. http://www.ssi-mi.com/services/lidar-services/mobile-lidar-moli (Accessed on Mar, 2016)
13. Chen, Y. & Krumm, J., 2010. Probabilistic modeling of traffic lanes from GPS traces. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 81-88). ACM.

APPENDIX A. List of Acronyms

Acronym	Definition
BSM	Basic Safety Message
CAD	Computer-aided Design
CV	Connected Vehicle
DOT	Department of Transportation
DSRC	Dedicated Short-Range Communications
EB	Eastbound
EB-LT	Eastbound – Left Turn
EB-TH	Eastbound – Through
FHWA	Federal Highway Administration
GID	Geometric Intersection Description
GPS	Global Positioning System
ID	Identification Number
IEEE	IEEE (Formerly the Institute of Electrical and Electronics Engineers)
ITS	Intelligent Transportation Systems
K-Means	A data partitioning method
LIDAR	Light Detection and Ranging
LT	Left Turn
MAP	SAE J2735 Map Message (a MapData Message)
MC	Moderately Complex
N/A	Not Applicable
NAD83	North American Datum 1983
NB	Northbound
NB-LT	Northbound – Left Turn
NB-TH	Northbound – Through
OBU	Onboard Unit

Acronym	Definition
RSU	Roadside Unit
RT	Right Turn
SAE	SAE International
SB	Southbound
SB-LT	Southbound – Left Turn
SB-TH	Southbound – Through
SPaT	Signal Phase and Timing
SPMD	Safety Pilot Model Deployment (Project)
V2I	Vehicle-to-Infrastructure
WB	Westbound
WB-LT	Westbound – Left Turn
WB-TH	Westbound – Through
WGS84	World Geodetic System 1984

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free "Help Line" 866-367-7487
www.its.dot.gov

FHWA-JPO-16-416

U.S. Department of Transportation