Intersection and Stop Bar Position Extraction from Crowdsourced GPS Trajectories

Chao Wang, Ph.D. candidate
(Corresponding Author)
Department of Electrical and Computer Engineering, UC Riverside
1084 Columbia Avenue, Riverside, CA 92507
Tel: (951) 522-7381; Fax: (951) 781-5790
Email: cwang061@ucr.edu

Peng Hao, Ph.D.
Center for Environmental Technology and Research, UC Riverside
1084 Columbia Avenue, Riverside, CA 92507
Tel: (518) 364-3671; Fax: (951) 781-5790
Email: haop@cert.ucr.edu

Guoyuan Wu, Ph.D.
Center for Environmental Technology and Research, UC Riverside
1084 Columbia Avenue, Riverside, CA 92507
Tel: (951) 781-5630; Fax: (951) 781-5790
Email: gywu@cert.ucr.edu

Xuewei Qi, Ph.D., candidate
Center for Environmental Technology and Research, UC Riverside
1084 Columbia Avenue, Riverside, CA 92507
Tel: (706) 206-5972; Fax: (951) 781-5790
Email: xqi001@ucr.edu

Tingxu Lyu, M.S.
Department of Electrical and Computer Engineering, UC Davis
Tel: (530) 761-3439; Fax: (951) 781-5790
Email: txlyu@ucdavis.edu

Matthew Barth, Ph.D.
Department of Electrical and Computer Engineering, UC Riverside
1084 Columbia Avenue, Riverside, CA 92507
Tel: (951) 781-5782; Fax: (951) 781-5790
Email: barth@ece.ucr.edu

Total number of words: 4,250 (text) + 3,250 (2 tables, 11 figures) = 7,500

Submitted for presentation at the 96th Annual Meeting of Transportation Research Board and Publication in Transportation Research Record
ABSTRACT
Detailed road features (e.g., lane marks and stop bars) are crucial for many recent intelligent transportation system applications, especially for automated or autonomous driving systems. In this paper, a crowdsourcing based method is proposed to mark intersection areas and map stop bar positions without prior knowledge of road information. The proposed method includes an efficient approach for marking intersection areas by analyzing the entropy of moving direction, as well as a statistical model of stop positions for estimating the number and coordinates of stop bars. The proposed method is applied to the real-world dataset collected for the Safety Pilot Model Deployment Program (SPMDP). The numerical analysis results prove its applicability and robustness in processing GPS trajectories of an urban region (a 1.2 km by 2 km rectangular area). For the intersections covered well by trajectories, the accuracy of marking intersections is 95.7%. For stop bar positioning, the mean and standard deviation of the errors are 0.25 m and 0.32 m.

Keywords: Crowdsourced data, detailed road features, Global Positioning System (GPS), entropy, Gaussian Mixture Model (GMM), hierarchical clustering, maximum likelihood estimation (MLE), probability density function (PDF).
I. INTRODUCTION

The advances in connected vehicle (CV) technologies and autonomous vehicles are producing increasing demands on digital maps with detailed road features. High-resolution roadway and intersection information is required for many applications such as lane-level navigation and self-driving control. And such information usually requires continuous update to ensure its accuracy and reliability for safety purpose. Recently, LiDAR has been widely used for mapping with detailed road features (up to centimeter level) extracted from the laser point clouds (1-3), but the cost is high and the data processing is computationally expensive.

To achieve higher map updating efficiency with low cost, mining road features from crowdsourced data becomes promising and attracts more and more attentions from researchers and engineers. As the number of vehicles equipped with GPS sensors increases, more and more probe data are available for road feature extraction. It has been proven that the crowdsourcing approach is capable of constructing road-level maps from GPS data (4). However, it is challenging to extract other features such as stop bars and lane markers, because both GPS errors and driver behaviors introduce noises and uncertainties to the data. Therefore, advanced statistical analysis and modeling methods have been developed to improve the information accuracy and reliability.

For almost two decades, crowdsourcing methods have already been used to extract lane features. In 1998, Wilson et al. introduced the idea that uncoordinated probe vehicles with positioning capabilities and communications could be used to map the lane network with decimeter accuracy (5). In a continued work (6), a data-mining approach was proposed to build a lane model that gave predictions with high accuracy from a small number of trips passed over a particular road segment. The main idea is to cluster the traces for different lanes, and to use the mean of each cluster as the lane centerline (represented by the offset to road centerline). Similarly, the Kernel Density Estimation method (7) was applied to estimate the probability density function (PDF) of the vehicle’s position on a cross section. With the PDF, the lane centerlines could be well presented by the peaks of the function. This method is independent of the lane width and lane parallelism, and can handle lane splits and merges. In (8), a Gaussian Mixture Model (GMM) was used to cluster the GPS trajectories in the lateral direction, where the position of centerline was represented by the mean of each cluster.

While many research efforts have been conducted on mapping lateral traffic control factors (e.g., lane marks) on roadways, the longitudinal traffic control factors such as stop bars at intersections are lack of study. To our best knowledge, there has been very little research work aiming at estimating the stop bar position with crowdsourced data. In addition, most related studies are based on a priori knowledge of existing road-level maps, which limits their applicability.

In this paper, we propose a method that locates intersections and extracts stop bar locations only from crowdsourced GPS trajectories. A novel efficient model for marking intersection areas is developed by analyzing the diversity of vehicles’ moving directions in terms of entropy. For each intersection area, approaching directions are used to determine the number of stop bars. A statistical model is designed to describe the distribution of idling points in each approaching direction in order to estimate the positions of stop bar accurately. The remainder of the paper is organized as follows. Section II details the proposed methods for locating intersections and stop
bars. Section III shows a case study with data from the Safety Pilot Model Deployment Program (SPMDP). Section IV concludes this study and also provides the future directions.

II. METHODOLOGY

System Architecture
The overall process of intersection identification and stop bar localization is provided in Figure 1. And the detailed description for each step is given as follows:

1. Mark intersections
   a. Select the study region \( A \). Split \( A \) into square cells with grids.
   b. Calculate the entropy of the driving directions inside each cell, find the cells with local maximum entropies.
   c. Group together the cells with maximum entropy near to each other. Find the weighted average positions of the cells in each group, and mark as intersections. Assign every intersection a zone.

2. Determine the number of stop bars by direction grouping
   a. Inside each intersection zone, find all the idling points of the going-through trajectories (the trajectories not turning at the intersection, but passing through it), cluster them by approaching directions towards the intersection.

3. Estimate the stop bar position
a. For each approaching direction, project the idling points along the road. Model the probability distribution of the idling position as a Gaussian mixture model (GMM) with the stop bar position as a parameter.

b. Apply modified maximum likelihood estimator (MLE) to estimate the parameters of the GMM.

Please note that the only data source needed is the GPS trajectories and this method does not require any a priori knowledge on the map.

Intersection Marking

Entropy Analysis

Before we can identify the stop bars we need to mark the intersection areas first. Compared to the driving on non-intersection road segments, vehicles at intersection areas have more diverse moving directions. So finding intersection areas is equivalent to selecting areas with high diversity of moving directions. The moving direction can be measured by GPS receivers as GPS heading (0–360 degree). And a good index for diversity is the Shannon entropy (or Shannon index). Therefore, the 2-dimensional area of study is divided into square cells by grids (see Figure 2). For each cell, the GPS heading entropy for all the trajectory points inside that cell is calculated by the following equation (10):

\[ H = - \sum_{i=1}^{n} p_i \log p_i \]

where we split the round angle (360 degrees) into \( n \) intervals of equal size; \( p_i \) is the proportion of heading angle belonging to the \( i \)th angle interval. In the example shown in Figure 2, the total number of vehicle trajectory points is 1189 and the number of heading angle between 300° and 330° is 203, therefore the proportion of vehicle driving within that direction is 0.17.

![FIGURE 2 Entropy Calculation Example](image)

In this study, the cells within an intersection area would have higher entropy values compared to the cells on road segments (see Figure 3, Cell A is with higher entropy than Cell B). In order to find out all the cells within intersection areas, a strategy named “3-cell rule” is proposed: if the entropy value of one cell is higher than at least 3 of its neighbor cells (see Figure 3. Cell A is an example), then this cell is marked with intersection area.
**Cell Clustering at intersections**

In data mining and statistics, the hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters (I/I). It is commonly used when the number of clusters is unknown. Strategies for hierarchical clustering generally fall into two types:

1. Agglomerative ("bottom up" approach): each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

2. Divisive ("top down" approach): all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

During the clustering process, in order to decide which clusters should be combined (for agglomerative), or where a cluster should be split (for divisive), a measure of dissimilarity between sets of observations is required. This is achieved by the use of an appropriate metric (a measure of distance between pairs of observations), and a linkage criterion which specifies the dissimilarity of sets as a function of the pairwise distances of observations in the sets.

In this study, the “bottom up” approach is used to cluster the marked high-entropy cells by their locations. We choose the Euclidean distance as the metric. For two cells with centroid coordinates \(a(a_x, a_y)\) and \(b(b_x, b_y)\), the Euclidean distance is:

\[
d(a, b) = \|a - b\|_2 = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}
\]

Since the marked cells from different intersections are far from each other, the single-linkage criterion is adopted. The formula for single-linkage clustering is:

\[
\min\{d(a, b): a \in A, b \in B\},
\]

where \(A\) and \(B\) are the clusters we are merging. Then we gradually merge clusters with their linkage under a certain threshold (more detailed discussion on the threshold setting is provided in section III.). With this hierarchical clustering, the marked cells from the same intersection are merged together to form an intersection zone (see Figure 3 intersection zone).

**Detection of Driving Directions at Intersection**

**Idling Points identification**
Idling points are defined as the trajectory points with speed zero or very close to zero. The GPS heading information for these points are extracted for driving direction detection. To avoid extra noise caused by vehicles that perform turning movements at intersections, only idling points of straight trajectories within the determination area are used for further clustering analysis.

**Driving direction detection and grouping**

Within the marked intersection zone, the selected idling points are grouped into different approach directions by hierarchical clustering since the number of stop bars (4 for a 4-way intersection and 3 for a 3-way intersection) is unknown. Along each approach, the stop bar position is then estimated by fitting the idling points with our constrained GMM.

**Stop Bar Position Estimation for Each Approach Direction**

*Stop/idling position distribution with Gaussian Mixture Model*

Vehicles enter the intersection can stop either at the stop bar or stop behind a preceding idling vehicle. Assume the space headway (distance between the front bumpers of two consecutive vehicles in a queue) is a Gaussian random variable \(S \sim N(\mu_s, \sigma_s^2)\). For one dimensional, the noise caused by GPS errors can also be modeled as a Gaussian noise \(W_{gps} \sim N(0, \sigma_{gps}^2)\). If we set the stop bar position as the origin, then the idling position of the \(k\)-th vehicle in queue is

\[
X_k = B + (k - 1)S + W_{gps}
\]

where \(B\) is the bias that the first stop position shifted from the stop bar. It is not only caused by the distance from the GPS sensor to the vehicle’s front bumper, but also caused by the difference in driver behavior. Thus, \(B\) should be also a random variable and we model it as another Gaussian \(B \sim N(b, \sigma_b^2)\).

Since \((k - 1)S, W_{gps}\) and \(B\) are independent with each other, \(X_k\) is also in Gaussian distribution: \(X_k \sim N(\mu_k, \sigma_k^2)\), where \(\mu_k = b + (k - 1)\mu_s\) and \(\sigma_k^2 = \sigma_{gps}^2 + (k - 1)\sigma_s^2 + \sigma_b^2\). Therefore, along the vehicle driving direction, the position of an idling point is a random variable whose distribution is a Gaussian Mixture Model (GMM), a mixture of all \(X_k\)’s distributions:

\[
X = \sum_{k=1}^{m} \pi_k X_k
\]

where \(\pi_k\) is the mixture rate of the \(k\)-th Gaussian component and

\[
\sum_{k=1}^{m} \pi_k = 1.
\]

The stop bar position can be estimated by estimating the parameters of this GMM. Figure 4 provides an example of the probability distribution function of idling/stop positions and how it corresponds to the road geometry dimensions (bias, space headway).
Parameter estimation for Gaussian Mixture Model

Let \( \mathbf{x} = (x_1, x_2, \cdots, x_n) \) be a sample of \( n \) independent observations (\( n \) idling point positions) of a GMM with \( m \) components, and let \( \mathbf{z} = (z_1, z_2, \cdots, z_n) \) be the latent variables that determine the component from which the observation originates (12):

\[
X_i | (Z_i = k) \sim N(\mu_k, \sigma_k^2).
\]

where \( Z_i \in [1, m] \) and \( P(Z_i = k) = \pi_k, \sum_{k=1}^{m} \pi_k = 1 \). The goal is to estimate the unknown parameters:

\[
\theta = [\pi, \mu, \sigma^2]
\]

where \( \pi = [\pi_1, \pi_2, \cdots, \pi_m], \mu = [\mu_1, \mu_2, \cdots, \mu_m], \sigma^2 = [\sigma_1^2, \sigma_2^2, \cdots, \sigma_m^2] \).

The incomplete-data (not including \( \mathbf{z} \)) likelihood function is:

\[
L(\theta; \mathbf{x}) = p(\mathbf{x}|\theta) = \prod_{i=1}^{n} \sum_{k=1}^{m} \pi_k \cdot f(x_i, \mu_k, \sigma_k^2),
\]

where \( f \) is the Gaussian probability density function.

The unknown parameters can be determined by utilizing the maximum likelihood estimator (MLE) which is formulated as following:
\[
\hat{\theta} = \arg\max_{\theta} \{ l(\theta; x) \} = \arg\max_{\theta} \{ \log L(\theta; x) \} = \arg\max_{\theta} \{ \log p(x|\theta) \} = \arg\max_{\theta} \left\{ \sum_z \log p(x, z|\theta) \right\}.
\]

As above, we use natural logarithm of the likelihood function (log-likelihood) for calculation convenience. And since the logarithm is a monotonically increasing function, it will not affect the results. Usually, we need Expectation Maximization (EM) algorithm to get the iterative solution of the MLE. Since the number of values grows exponentially with the number of components, it makes the exact calculation of the sum extremely difficult. But our model is special. According to the definition above, the Gaussian components are correlated (Intuitively, the stop position of a car in queue depends on the one ahead). So the number of parameters is shrank to \( m + 5 \):

\[
\theta = [\pi, \mu_s, \sigma^2_s, \sigma^2_{gps}, b, \sigma^2_p].
\]

It is also noted that these parameters are usually constrained with real world limits, which allows us to get a sub-optimal solution of MLE by an exhaustive search within a space of all discretized parameters. As long as the discretized interval is small enough, the sub-optimal solution should be well acceptable.

### III. CASE STUDY AND RESULT ANALYSIS

The proposed methods are validated by applying to the data from the Safety Pilot Model Deployment (SPMD) program (13). The results are analyzed and presented in this section.

#### Dataset Introduction

The SPMD database contains two-month mobile data that were collected from over 2700 vehicles equipped with connected vehicle technologies, traveling in Ann Arbor, MI in October 2012 and April 2013. The data used in this research are derived from the onboard wireless safety unit (WSU), consisting of GPS-based data elements and those obtained from the vehicle’s Controller Area Network (CAN) Bus. In addition to GPS-based data, there is other information such as the real-time states and system performance of some critical components inside the vehicle. The parameters used in this study are listed in Table 1.

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Type</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>Integer</td>
<td>none</td>
<td>A unique numeric ID assigned to each DAS. This ID also doubles as a vehicle’s ID</td>
</tr>
<tr>
<td>Trip</td>
<td>Integer</td>
<td>none</td>
<td>Count of ignition cycles—each ignition cycle commences when the ignition is in the on position and ends when it is in the off position</td>
</tr>
</tbody>
</table>
Intersection Area Selection

The study region is located in the downtown area of Ann Arbor near the campus of University of Michigan. The boundary longitudes and latitudes are: west: -83.7532°; east: -83.7281°; north: 42.2845°; south: 42.2739°. The trajectory points’ density in the study region is shown in Figure 5. If we choose (longitude: -83.744, latitude: 42.281) as origin, and convert the (longitude, latitude) into (x, y) in meters, the boundaries of the study region are: west: -758.69 m; east: 131.21 m; north: 366.44; and south: -813.32.

We then discretize the study region by dividing into cells and determine which cell can represent the intersection. The size of cell has to be selected properly. If the cell is too large, the low resolution will result in large error when identifying the intersection area. If it is too small, there will be not enough samples in each cell for entropy calculation. Then the intersection cells will be less differentiable. In this study, we choose the cell size as 18 m × 18 m, which is similar to half of the intersection area size. In Figure 6, we show the GPS heading entropy heat map where the color transition is smoothed for a more comfortable look. As what we expect, the cell associated with an intersection always has higher GPS heading entropy.
Then we select the cells that are locally optimal (with entropy higher than 3 adjacent cells). Figure 7 shows the locations of those cells. As they are concentrated at the location of intersections, the hierarchical clustering can be directly applied to classify them with promising results.
The clustering process runs as the “bottom up” approach:

1. Initialize each cell as a cluster;
2. Repeat the following steps:
   a. Calculate all the minimum linkage (the minimum distance between cells from two different clusters) between each of two clusters;
   b. Merge clusters whose linkage in-between is less than a threshold (20 meters);
Until no merging can be executed.

After clustering, we identify each intersection using its center, which is calculated as the weighted average of the inside cells’ positions. The entropy then acts as a weighting factor:

\[ x_{ci} = \sum_{j=1}^{N_i} H_j x_j, \quad y_{ci} = \sum_{j=1}^{N_i} H_j y_j. \]

We then define the intersection region as the study area for intersection feature extraction. The size of the region is set to be 70 m × 70 m, about 8 times as large as a typical intersection size. We work on an extended intersection region in order to include more data samples for stop bar detection and to reduce the intersection positioning error. As shown in Figure 8(a), most intersections are accurately identified, while some of them are not detected mainly due to very little traffic passing through them (less than 2 trajectories). It is also noted that some regions with high driving direction entropy such as parking lot are mistakenly marked. For the intersections covered well by trajectories within the study region, the accuracy of intersection marking is as high as 95.7%. The average error on intersection center estimation is about 12m (see Figure 8(b)).
(a) Intersection center and region identification

(b) Error for intersection center estimation

**FIGURE 8 Error for Intersection Center Estimation**

**Stop Bar Number and Locations Determination**

To identify the number and location of stop bars, we choose the intersection of East Huron Street and Division Street as an example. Before we apply the bar position (associated with each approach) estimation algorithm to the intersection, we filter and process the raw data from the intersection region as follows:
(1) Movement filtering. We select the trajectories going straight through the intersection without making a turn. For each trajectory, we accumulate its turning angles (negative for left turn and positive for right turn) obtained by taking the difference of GPS headings. Trajectories with accumulated turning angle greater than 80 degrees or less than -80 degrees are excluded.

(2) Speed filtering. We then keep the idling points in the trajectories. Considering the errors and lag effects of GPS sensors, we select all data points with speed value less than a predefined threshold (0.1 m/s in this study).

(3) Approach clustering. We utilize the hierarchical clustering to group these idling points along each approaching direction based on GPS headings. The clustering process starts with each point as a cluster, and keeps merging clusters until the minimum linkage of any two clusters is smaller than the linkage threshold (e.g., 20°).

It is noted that there could be an issue if we directly apply clustering algorithm to the GPS headings. For example, GPS headings of 355° and 5° are both near northward, but they will not be grouped together because the linkage (distance) is |355° - 5°| = 350° > 20°. This problem can be solved by using vectors to express GPS headings: \( \vec{v} = (\cos h, \sin h) \), where \( h \) is the GPS heading. Then the 2-D vector \( \vec{v} \) is used as the feature for clustering. As shown in Figure 9, three approaches are clearly classified based on the heading vector, which indicates that the proposed method works satisfactorily without any mis-grouping.

**FIGURE 9 Idling Points Clustering**

For the filtered data of each approach, we apply the proposed method in Section II to estimate the location of corresponding stop bar. Here, we use approach 3 (cluster 3, S Division St) in the last step as an instance. All the idling points’ coordinates are projected to the longitudinal direction.
The histogram of the data samples with respect to the distance to the origin is shown in Figure 10. All the raw trajectory data are involved for analyzing, taking into account both light and heavy traffic conditions.

Before searching the optimal solution to the proposed Gaussian mixture model, we set constraints for the parameters to improve the computational efficiency with the integration of traffic knowledge. For example, the space headway is selected to be greater than the minimum possible car length but smaller than the maximum possible car length plus a large clearance. The number of Gaussian components is limited to 3, as the GPS position of the 4th vehicle (or vehicles behind) in a queue has little correlation with the stop bar position. We also discretize other parameters in the searching space.

Some parameters that are not site-specific can be calibrated using training data from other places, e.g. the distance between the stop bar location and the location of the GPS antenna. To identify the distribution of that distance, we use the same type of data collected in Riverside, CA as the training set. In that test, three drivers were asked to stop at a stop bar in the open area, repeating for 20 times. We recorded each bias distance from the stop bar to the GPS antenna and calculated the mean and standard deviation of the bias: $b' = 0.56\text{m}$, $\sigma_b' = 0.17\text{m}$. Table 2 shows the details of the parameter constraints. Figure 10 presents the fitting result. The longitudinal stop bar position is calculated by:

$$P_s = E[X_1] - E[B],$$

where $E[X_1]$ is the mean of the first Gaussian component, and $E[B]$ is the mean of the bias.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Description</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>Integer</td>
<td>Number of Gaussians</td>
<td>$1 \leq m \leq 3$</td>
</tr>
<tr>
<td>$b$</td>
<td>Real</td>
<td>Mean of the bias $B$</td>
<td>$b' - 4 &lt; b &lt; b' + 4$</td>
</tr>
<tr>
<td>$\sigma_b$</td>
<td>Real</td>
<td>Standard deviation of the bias $B$</td>
<td>$\sigma_b'$</td>
</tr>
<tr>
<td>$\mu_s$</td>
<td>Real</td>
<td>Mean of the space headway $S$</td>
<td>$5 \leq \mu_s \leq 10$</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>Real</td>
<td>Standard deviation of the space</td>
<td>$0 \leq \sigma_s \leq 1.5$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>headway $S$</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{gps}$</td>
<td>Real</td>
<td>Standard deviation of the 1</td>
<td>$0 \leq \sigma_{gps} \leq 2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dimensional GPS error $W_{gps}$</td>
<td></td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>Real</td>
<td>Mixture rate of first Gaussian</td>
<td>$0.4 \leq \pi_1 \leq 1$</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>Real</td>
<td>Mixture rate of second Gaussian</td>
<td>$0 \leq \pi_2 \leq \pi_1$</td>
</tr>
<tr>
<td>$\pi_3$</td>
<td>Real</td>
<td>Mixture rate of third Gaussian</td>
<td>$0 \leq \pi_3 \leq \pi_2$</td>
</tr>
</tbody>
</table>
Figure 11 shows final results for the estimated positions of 3 stop bars at the East Huron Street-Division Street intersection. The stop bars are orthogonal to the average moving direction of each approach. And the length of the stop bars are determined by the lateral range of straight through trajectories.

To evaluate the proposed method, the positions of 12 stop bars at 4 intersections on the corridor of East Huron Street are estimated. Using the stop bars on Google Earth as ground truth, we then get the mean (0.25 m) and standard deviation (0.32 m) of longitudinal position error, respectively. That
means the accuracy of the stop bar detection is in meter-level, which should satisfy many ITS applications such as pre-stop warning and eco-approaching.

IV. CONCLUSIONS AND FUTURE WORK
This study proposes a crowdsourcing based approach to detect stop bars at intersections. This method is capable of efficiently extracting stop bar positions of all the intersections in a given area by mining the probe vehicles’ GPS trajectories without any a priori map information. The dataset from Safety Pilot Model Deployment Program is used for testing. Numerical analysis shows that 95.7% of the intersections (covered by a certain amount of trajectories) within the test region is correctly marked with an average shift of 12 m in the center location (with respect to the ground truth). For the stop bar positioning, the mean and standard deviation of the error are 0.25 m and 0.32 m, respectively. Such accuracy can satisfy many ITS applications such as pre-stop warning and eco-driving at intersections. The future work will involve investigation of the impact of different road geometry designs on the stop bar positioning. And these differences can be potentially extracted from the same type of data. We may also take advantage of the additional information (e.g., space headway) learned by the GMM to better understand the traffic conditions at intersections in urban areas. Furthermore, the entropy analysis method can be used for freeway traffic information mining to find disordered and chaotic spots caused by accidents or heavy traffic.
REFERENCE


7. Tang, L., Yang, X., Dong, Z., & Li, Q. CLRIC: Collecting Lane-Based Road Information Via Crowdsourcing.


15. Cao, L., & Krumm, J. (2009, November). 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.