# Lane Marking Verification for High Definition Map Maintenance Using Crowdsourced Images

Binbin Li, Dezhen Song, Aaron Kingery, Dongfang Zheng, Yiliang Xu, and Huiwen Guo

***Abstract*— Autonomous vehicles often rely on high-definition (HD) maps to navigate around. However, lane markings (LMs) are not necessarily static objects due to wear & tear from usage and road reconstruction & maintenance. Therefore, the wrong matching between LMs in the HD map and sensor readings may lead to erroneous localization or even cause traffic accidents. It is imperative to keep LMs up-to-date. However, frequently recollecting data to update HD maps is cost-prohibitive. Here we propose to utilize crowdsourced images from multiple vehicles at different times to help verify LMs for HD map maintenance. We obtain the LM distribution in the image space by considering the camera pose uncertainty in perspective projection. Both LMs in HD map and LMs in the image are treated as observations of LM distributions which allow us to construct posterior conditional distribution (a.k.a Bayesian belief functions) of LMs from either sources. An LM is consistent if belief functions from the map and the image satisfy statistical hypothesis testing. We further extend the Bayesian belief model into a sequential belief update using crowdsourced images. LMs with a higher probability of existence are kept in the HD map whereas those with a lower probability of existence are removed from the HD map. We verify our approach using real data. Experimental results show that our method is capable of verifying and updating LMs in the HD map.**

## I. Introduction

The fast-evolving autonomous vehicle (AV) technology has the potential to drastically change modern transportation. Many AVs rely on a high-definition (HD) map to navigate around. HD maps include a highly accurate and realistic representation of the road, including many types of objects such as lane markings (LMs), traffic signs, street lamp posts, etc. In the absence of accurate GPS signals, the precision of LMs in HD maps is important for the vehicle to recognize lanes and plan for its motion. However, LMs are not necessarily constant because they wear out due to road usage and also vary due to road construction and maintenance. A set of outdated LMs may lead to erroneous localization results. Frequently recollecting LM data is cost-prohibitive.

We propose to utilize crowdsourced images to keep LMs up-to-date in the HD map. We view LMs in both the HD map and the crowdsourced images as observations of LM distribution. We model the posterior LM distribution in either source using Gaussian kernels. We take the uncertainty from camera poses into consideration for image-based observations. We examine their consistency within the same image coordinate using statistical hypothesis testing. We then establish a sequential Bayesian model for updating the posterior LM distributions using a sequence of crowdsourced images. We threshold the conditional probability to determine if each LM is consistent, inconsistent, or undetermined. We have implemented our map verification algorithm and tested it using real data. The experimental results show that the algorithm has achieved its design goal and outperformed commonly used intersection over union (IoU) metric in precision, recall and F1-measure.

B. Li, A. Kingery and D. Song are with CSE Department, Texas A&M University, College Station, TX 77843 US. (Emails: binbinli@tamu.edu; aaronmkingery@tamu.edu; dzsong@cs.tamu.edu).

D. Zheng, H. Guo and Y. Xu are with Tencent America, Palo Alto, CA 94306 US (Emails:{dzheng, huiwenguo, and yiliangxu}@tencent.com)

This work was supported in part by Tencent America, in part by GM/SAE Autodrive Challenge, in part by NSF under NRI-1925037.

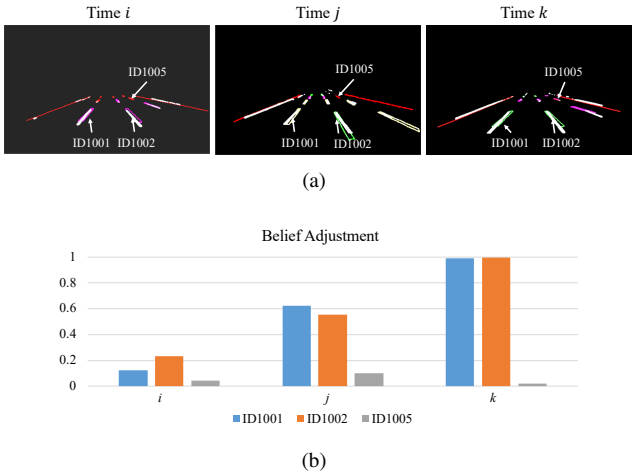



Fig. 1. (a) LMs are extracted from front-view camera images. We project LMs from the HD map, which have unique ID numbers in the map database, into the front-view camera images taken at different time with index $i < j < k$. Different colors stand for different stages for LM belief. LM in purple is "inconsistent", LM in Yellow is "undetermined", and LM in green is "consistent". (b) With accumulated observations such as camera images in (a) from left to right, the LMs with ID1001 and ID1002 are "consistent" and kept in the HD map, while the LM with ID1005 is labeled as "inconsistent" and removed from the HD map.

## II. Related Work

AVs require up-to-date high definition maps to ensure safe navigation and to cope with environmental changes [1]–[4]. To create updatable HD maps, it is necessary to 1) have the ability to detect LMs, 2) design a flexible data structure to represent maps, and 3) develop algorithms to validate and maintain HD maps.

LM detection and tracking play an important role in autonomous driving, which has been studied for years [5]–[7]. Andrade et al. [8] use Hough transform to track LMs through the shape-preserving spline interpolation. In [9], we fuse camera images and lidar point clouds to detect LMs and assess LM quality by proposing correctness, shape and visibility metrics. Our recent work [10] also generates

virtual LMs in sensor space while considering vehicle size and kinodynamic constraints. Huang et al. [11] detect and estimate multiple LMs by fusing calibrated video images and laser range data captured by a moving vehicle. Kang et al. [12] propose a probabilistic decision-making algorithm to track curbs that uses interacting multiple model method for autonomous mobile robot navigation. Li et al. [13] apply hierarchical neural networks to detect lane boundaries including those areas without any lane markings. Here we build on existing LM detection work to provide inputs for HD map maintenance.

In robotics, simultaneous localization and mapping (SLAM) has developed many map representations as a collection of landmarks which include occupancy grids [14], [15], sparse visual features [16]–[18], and point clouds [19]. However, most existing map representation are designed for stationary objects without consideration of frequent updates.

In recent developments, Ryde et al. [20] employ multi-resolution occupied voxel lists to represent 3D spatial maps, which detects changes by finding points that do not locate inside an occupied voxel after alignment. Aijazi et al. [21] extract temporarily static and mobile 3D point clouds by matching sensor's observations on at different times of the day, which yields the progressively modified 3D urban landscape. Wang et al. [22] detect and track dynamic objects in dynamic environments, and build a map that satisfies both navigation and safety requirements for autonomous driving in urban areas. Julie et al. [23] assign scores for features in the map, which depend on the geometric distribution and characteristics when the features are re-detected at a different time. Sun et al. [24] present a novel semantic mapping approach for the successfully mapping of a dynamic environment using more than two weeks of data. Nurminen et al. [25] propose methods to support spatial updating and rapid alignment of physical and virtual spaces in the 3D mobile maps. Unlike existing approaches, we employ Bayes' theorem to track the belief changes of LMs by fusing observations from crowdsourced data. Our method can remove or add LMs as needed over long periods of time.

## III. Problem Formulation

When a vehicle is driving on a street with HD maps, it takes images from its camera and verifies if LMs on the road are the same as those in its HD map. The HD map usually consists of a variety of objects such as LMs, traffic signs, street lamp posts, etc. LMs are the most common landmarks to help achieve high precision global localization. Here we focus on verifying LMs in the HD map.

### A. Assumptions and Coordinate Systems

The vehicle is equipped with a front facing camera to observe the LMs. We assume that the camera is pre-calibrated, and the nonlinear distortion of images has been removed [26].

All coordinate systems or frames are right-handed systems and defined as follows,

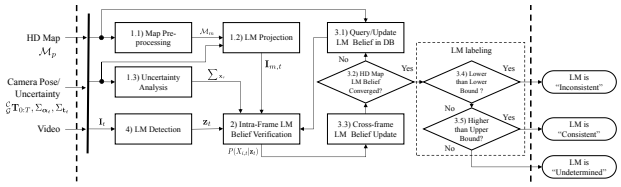



Fig. 2. System diagram.

- $\{\mathcal{C}\}$ defines the camera coordinate system with its origin at the camera center, $z$-axis pointing forward coinciding with the camera's principal axis, and its $x$-axis and $y$-axis parallel to the horizontal and vertical directions of camera imaging sensor, respectively.
- $\{\mathcal{I}\}$ defines the image coordinate system. Let ${}^{\mathcal{I}}\mathbf{x} = [u\ v]^{\mathsf{T}} \in \mathbf{I}_t$ be a pixel point in camera image $\mathbf{I}_t$ in $\{\mathcal{I}\}$ at time $t$ where $(u, v)$ is the image coordinate.
- $\{\mathcal{G}\}$ defines the vehicle global frame with the $x$-axis pointing to the east, $y$-axis pointing north, and $z$-axis pointing upward.

Note that we will attach frames to a variable as the left super and sub scripts to indicate which frame the variable is associated with.

### B. HD Map and Camera Inputs

We have inputs from both a HD map database and the on-board camera/sensors. From the HD map, we have,

- ${}^{\mathcal{G}}\mathbf{M}_i$ is the $i$-th LM consisting of a set of points ${}^{\mathcal{G}}\mathbf{M}_i := \{{}^{\mathcal{G}}\mathbf{M}_{i,j} \in \mathcal{R}^3 | j = 1, 2, ..., n_l\}$, where $\mathbf{M}_{i,j}$ is the $j$-th point in the LM and $n_l$ is the number of the points in the $i$-th LM. Correspondingly, we also have ${}^{\mathcal{I}}\mathbf{M}_i$ in camera frame.
- $\mathcal{M}_p$ is a HD map consisted of a set of LMs $\mathcal{M}_p := \{{}^{\mathcal{G}}\mathbf{M}_i \subset \mathcal{R}^3 | i = 1, 2, ..., n_m\}$ where ${}^{\mathcal{G}}\mathbf{M}_i$ is the $i$-th LM set and $n_m$ is the number of LMs.

Through an on-board map-based localization algorithm, the vehicle obtains the camera pose and its uncertainty range. Denote the camera's pose at time $t$ by ${}^{\mathcal{C}}_{\mathcal{G}}\mathbf{T}_t$. ${}^{\mathcal{C}}_{\mathcal{G}}\mathbf{T}_t$ is the rigid body transformation from frame $\mathcal{G}$ to frame $\mathcal{C}$,

$$ {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{T}_t = \begin{bmatrix} {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t & {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}, $$

where ${}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t$ is the translation vector from the origin of $\{\mathcal{G}\}$ to the origin of $\{\mathcal{C}\}$, and ${}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t \in \mathcal{SO}^3$ is the rotation matrix from $\{\mathcal{G}\}$ to $\{\mathcal{C}\}$ and represented in Euler angle $\boldsymbol{\alpha}_t = [\phi, \theta, \psi]^{\mathsf{T}}$ in Z-Y-X order,

$$ {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t = \begin{bmatrix} c\phi & s\phi & 0 \\ -s\phi & c\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\theta & 0 & s\theta \\ 0 & 1 & 0 \\ -s\theta & 0 & c\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\psi & -s\psi \\ 0 & s\psi & c\psi \end{bmatrix}, $$

where symbols $c$ and $s$ represent 'cos' and 'sin' operators, respectively.

For uncertainties of camera poses, we have multivariate Gaussian $\boldsymbol{\alpha}_t \sim \mathcal{N}(\bar{\boldsymbol{\alpha}}_t, \Sigma_{\boldsymbol{\alpha}_t})$ and ${}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t \sim \mathcal{N}({}^{\mathcal{C}}_{\mathcal{G}}\bar{\mathbf{t}}_t, \Sigma_{\mathbf{t}_t})$, where $\Sigma_{\boldsymbol{\alpha}_t}$ and $\Sigma_{\mathbf{t}_t}$ are the corresponding covariance matrices for rotation and translation, respectively. Here, the overhead symbol '–' represents the mean of the vector.

For camera image $\mathbf{I}_t$, we can extract LM points using lane detection algorithms [9]. It results in ${}^{\mathcal{I}}\mathbf{x}_s$ as LM points in the image (see Box 4 in Fig. 2). We do not need to group them into different LM sets. Assemble all LM points ${}^{\mathcal{I}}\mathbf{x}_s$, we obtain set $\mathbf{z}_t$ and its cardinality $n_h = |\mathbf{z}_t|$.

### C. Problem Definition

We want to use crowdsourced images to confirm or disconfirm each LM set ${}^{\mathcal{G}}\mathbf{M}_i$ in HD map. This will generate three labeled categories including "consistent", "inconsistent" or "undetermined." The "consistent" LM ${}^{\mathcal{G}}\mathbf{M}_i$ will be kept in the HD map while "inconsistent" LM points will be removed, and those "undetermined" LM points will require more observations in the future to ascertain its consistency.

Given a sequence of crowdsourced data ordered by time $t = 0, ..., T$ where $T$ the latest time index, our problem is defined as follows,

*Problem 1:* Given the HD map $\mathcal{M}_p$, camera images $\mathbf{I}_{0:T}$, and historical camera poses ${}^{\mathcal{C}}_{\mathcal{G}}\mathbf{T}_{0:T}$ with known covariance matrices, label consistence category for each LM set ${}^{\mathcal{G}}\mathbf{M}_i$ in the HD map $\mathcal{M}_p$.

## IV. Algorithm

Fig. 2 illustrates our system diagram. It mainly contains the following blocks: (1.1-1.3) we project LM points from HD map into the correct camera frame of the vehicle. We analyze and compute the uncertainty of the projected points; (2) We update LM point belief modeling given the current image observation; (3.1-3.5) LM point belief update by accumulating all the historical camera observations. We start with the first block.

### A. Lane Marking Projection and Uncertainty Analysis

Note that the vehicle samples data periodically. At discrete time $t$, we have the camera's pose $\{{}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t, {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t\}$. We extract a subset of LMs from the HD map based on vehicle speed $v_t$ and distance threshold $d_m$,

$$\mathcal{M}_m = \{{}^{\mathcal{G}}\mathbf{M}_{i,j} | \|{}^{\mathcal{G}}\mathbf{M}_{i,j} + {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t^{\mathsf{T}} {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t\| \leq d_m, {}^{\mathcal{G}}\mathbf{M}_{i,j} \in \mathcal{M}_p\}. \tag{1}$$

Here, $\|\cdot\|$ is the vector $l^2$-norm. Distance threshold $d_m$ is obtained as follows,

$$d_m = \begin{cases} \zeta \cdot v_t \tau & \text{if } v_t > 0 \\ \nu_v, & \text{otherwise} \end{cases}$$

where $\zeta$ controls the overlapping regions of the HD map between neighboring $\mathcal{M}_m$, $\tau$ is the sampling interval, and $\nu_v$ is a constant. Define $\mathbf{U} = \mathbf{K}\, {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t$, and $\mathbf{U}^{3\mathsf{T}}$ to be the third row of $\mathbf{U}$. Here, $\mathbf{K} \in \mathcal{R}^{3\times 3}$ is the intrinsic camera matrix under the pin hole model. We remove point ${}^{\mathcal{G}}\mathbf{M}_{i,j}$ in $\mathcal{M}_m$ that is in the back of the camera if the condition,

$$\mathbf{U}^{3\mathsf{T}}({}^{\mathcal{G}}\mathbf{M}_{i,j} + {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t^{\mathsf{T}} {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t) < 0,$$

is satisfied. Recall $\mathcal{M}_m$ is made of a set of LM points with known LM index, and we have grouped the points belonging to the $i$-th LM as ${}^{\mathcal{G}}\mathbf{M}_i$. We accumulate such LMs that can be projected into image $\mathbf{I}_t$ in set $\{{}^{\mathcal{G}}\mathbf{M}_i | i \in \mathcal{M}_t\}$, where $\mathcal{M}_t$ is the index set.

Given the camera pose $\{{}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t, {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t\}$, we can project LMs from $\{\mathcal{G}\}$ to $\{\mathcal{C}\}$ through perspective projection,

$$\tilde{\mathbf{x}}_r = c_p \mathbf{K}({}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t \mathbf{X}_r + {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t) \tag{2}$$

for each LM point $\mathbf{X}_r \in {}^{\mathcal{G}}\mathbf{M}_i$, where $c_p$ is a scalar, and a vector with symbol '$\sim$' on top is in its homogeneous representation. This generates a projected HD map pixel set $\mathbf{z}_{m,t} := \{{}^{\mathcal{I}}\mathbf{x}_r | \mathbf{X}_r \in {}^{\mathcal{G}}\mathbf{M}_i\}$ at time $t$.

The point positions of $\mathbf{X}_r$ are not noise free. We need to understand how it propagates to the image frame. The noise distribution of $\mathbf{X}_r$ is modeled as bounded Gaussian with a zero-mean and covariance matrix $\sigma_r^2 \mathbf{I}_3$, where $\mathbf{I}_3$ is a $3 \times 3$ identity matrix and $\sigma_r$ is determined by the accuracy of the HD map. As a function of $\boldsymbol{\nu} = [\boldsymbol{\alpha}_t^{\mathsf{T}}, \mathbf{X}_r^{\mathsf{T}}, {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t^{\mathsf{T}}]^{\mathsf{T}}$ in (2), we have

$$\operatorname{cov}\left(\begin{bmatrix} \boldsymbol{\alpha}_t \\ \mathbf{X}_r \\ {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{\boldsymbol{\alpha}_t} & \mathbf{0}_{3\times 3} & \operatorname{cov}(\boldsymbol{\alpha}_t, {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t) \\ \mathbf{0}_{3\times 3} & \sigma_r^2 \mathbf{I}_3 & \mathbf{0}_{3\times 3} \\ \operatorname{cov}(\boldsymbol{\alpha}_t, {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t) & \mathbf{0}_{3\times 3} & \Sigma_{\mathbf{t}_t} \end{bmatrix}, \tag{3}$$

by assuming that $\mathbf{X}_r$ is independent of the other two vectors. Then we have

$$\Sigma_{\mathbf{x}_r} = J_{\boldsymbol{\nu}} \operatorname{cov}\left(\begin{bmatrix} \boldsymbol{\alpha}_t \\ \mathbf{X}_r \\ {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{t}_t \end{bmatrix}\right) J_{\boldsymbol{\nu}}^{\mathsf{T}}, \tag{4}$$

under the first-order approximation (see Box 1.3 in Fig. 2) in error forward propagation, where $J_\nu$ is the Jacobian matrix of (2) by

$$J_{\boldsymbol{\nu}} = \frac{\partial \tilde{\mathbf{x}}_r}{\partial \boldsymbol{\nu}} = c_p \mathbf{K} \begin{bmatrix} \frac{\partial({}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t \mathbf{X}_r)}{\partial \boldsymbol{\alpha}_t} & {}^{\mathcal{C}}_{\mathcal{G}}\mathbf{R}_t & \mathbf{I}_{3\times 3} \end{bmatrix}. \tag{5}$$

The covariance matrix $\Sigma_{\mathbf{x}_r}$ characterizes the uncertainty of the projected LM points from HD map to the current camera frame. It allows us to establish a belief model for LM points.

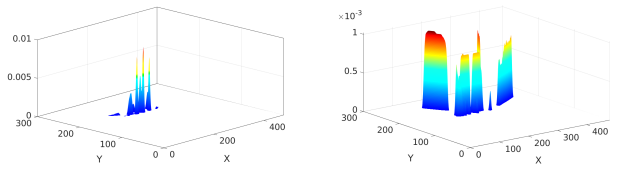
### B. Intra-Frame Lane Marking Verification

It is worth noting that verifying LMs between the HD map and those in the current camera frame is not a point-to-point verification. In fact, this is a set-to-set association and requires a new belief model to facilitate this. We first establish a pixel-wise intra-frame belief function to verify LMs using a single frame.

Due to the existence of noises in HD maps, the projected HD map $\mathbf{z}_{m,t}$ can be understood as an observation of actual LMs in the current camera coordinate system. We model the conditional probability distribution of a pixel being a true LM pixel as a weighted sum of Gaussian functions established given the pixel and its neighbors (see Fig. 3(a)) in the observation $\mathbf{z}_{m,t}$,

$$f_m({}^{\mathcal{I}}\mathbf{x} | \mathbf{z}_{m,t}) = \sum_{{}^{\mathcal{I}}\mathbf{x} \in N_e({}^{\mathcal{I}}\mathbf{x}_r)} w_a \frac{\exp(-\frac{1}{2} d({}^{\mathcal{I}}\mathbf{x}, {}^{\mathcal{I}}\mathbf{x}_r))}{2\pi\sqrt{|\Sigma_{\mathbf{x}_r}|}} \tag{6}$$

where $d({}^{\mathcal{I}}\mathbf{x}, {}^{\mathcal{I}}\mathbf{x}_r) = ({}^{\mathcal{I}}\mathbf{x} - {}^{\mathcal{I}}\mathbf{x}_r)^{\mathsf{T}} \Sigma_{\mathbf{x}_r}^{-1} ({}^{\mathcal{I}}\mathbf{x} - {}^{\mathcal{I}}\mathbf{x}_r)$, $N_e({}^{\mathcal{I}}\mathbf{x}_r)$ is the neighboring set with $d({}^{\mathcal{I}}\mathbf{x}, {}^{\mathcal{I}}\mathbf{x}_r) \leq \kappa_m^2$, $\kappa_m$ is a threshold, $|\Sigma_{\mathbf{x}_r}|$ is the determinant of $\Sigma_{\mathbf{x}_r}$, $w_a$ is a normalization factor, $a = 1, 2, ..., n_g$, and $n_g$ is the set cardinality of $N_e({}^{\mathcal{I}}\mathbf{x}_r)$. Here we set $\kappa_m^2 = F^{-1}(\alpha, 2)$, where $F^{-1}(\alpha, 2)$ is



(a) (b)

Fig. 3. Pixel-wise LM probability distribution $f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})$ from the HD map in (a) and $f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t)$ from the image in (b).

the inverse cumulative $\chi^2$ distribution function with a desired confidence level of $\alpha$ and 2 degrees of freedom.

Similarly, the current camera image also provides an observation $\mathbf{z}_t$, we can model the conditional probability distribution $f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t)$ where $f_s(\cdot)$ shares the same format of $f_m(\cdot)$ (6) except that the noise covariance matrix is $\sigma_s^2\mathbf{I}_2$ instead of $\Sigma_{\mathbf{x}_r}$, and $\sigma_s$ is determined by the accuracy of our LM segmentation model. An example of $f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})$ and $f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t)$ is shown in Fig. 3(b).

For a pixel ${}^{\mathcal{I}}\mathbf{x} \in \mathbf{z}_t$, we can obtain both $f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})$ and $f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t)$. This allows us to test if ${}^{\mathcal{I}}\mathbf{x}$ is a consistent LM pixel across the HD map data and the camera image by verifying if $f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})$ and $f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t)$ are the same distribution through goodness of fit test,

$\mathbf{H_0}$: ${}^{\mathcal{I}}\mathbf{x}$ is a consistent LM pixel.
$\mathbf{H_1}$: Otherwise.

Through chi-square goodness of fit test [27], we have

$$\chi^2 = \frac{\left(f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t) - f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})\right)^2}{f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})}.$$

We reject $\mathbf{H_0}$ if

$$\chi^2 > \chi^2_{1-\beta,1},$$

where $\beta$ is the significance level.

Thus for each LM $\mathbf{z}_{m,t}$ at time $t$, we obtain the consistent pixel set as,

$$\mathcal{X}_{i,t} := \Big\{{}^{\mathcal{I}}\mathbf{x}|f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t) \leq f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t}) + \sqrt{f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})\chi^2_{1-\beta,1}}, {}^{\mathcal{I}}\mathbf{x} \in \mathbf{z}_{m,t}\Big\}.$$

Define $X_{i,t}|\mathbf{z}_t$ as the conditional spatial distribution of the $i$-th LM ${}^{\mathcal{G}}\mathbf{M}_i$ in camera frame given the observation $\mathbf{z}_t$. Then we have

$$P(X_{i,t}|\mathbf{z}_t) = \begin{cases} \frac{1}{\xi}\sum_{{}^{\mathcal{I}}\mathbf{x}\in\mathcal{X}_{i,t}} f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t), & \text{if } \mathcal{X}_{i,t} \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where $\xi$ is a normalization factor.

Note that as $P(X_{i,t}|\mathbf{z}_t)$ is easily influenced by the current observation $\mathbf{z}_t$, and current camera pose with respect to $\{\mathcal{G}\}$. We need to fuse observations from multiple vehicles at different times to ensure we can identify correct consistency category so that the $i$-th LM should be kept or removed from the HD map $\mathcal{M}_p$.

### C. Cross-frame Lane Marking Belief Update

LMs in the HD map can be classified into three categories: 1) LMs with no matchings in the image; 2) LMs which have appeared in the image; and 3) false-positive LMs caused by noises, which will be filtered out with more observations. Here we combine all the historic observations for the LM ${}^{\mathcal{G}}\mathbf{M}_i$ from the crowdsourced images to establish a robust belief function (see Box 3.3 in Fig. 2) and verify the existence of LMs in the map.

Accumulating all images up-to-date into current observation $\mathbf{z}_t$ at time $t$, we have $\mathbf{Z}_{0:t} = \bigcup_t \{\mathbf{z}_t\}$. Note that $\mathbf{Z}_{0:t-1} = \mathbf{Z}_{0:t} \setminus \mathbf{z}_t$. Define $P(X_i|\mathbf{Z}_{0:t-1})$ as the conditional spatial probability of the LM ${}^{\mathcal{G}}\mathbf{M}_i$ given the observation set $\mathbf{Z}_{0:t-1}$. Similarly, we define the conditional probability $P(X_i|\mathbf{Z}_{0:t})$ here. To verify the existence of the LM ${}^{\mathcal{G}}\mathbf{M}_i$, our problem becomes how to compute $P(X_i|\mathbf{Z}_{0:t})$ given the current observation $\mathbf{z}_t$, previous observation set $\mathbf{Z}_{0:t-1}$ and the conditional probability $P(X_i|\mathbf{Z}_{0:t-1})$.

We decompose the conditional probability $P(X_i|\mathbf{Z}_{0:t})$ and have,

$$\begin{aligned} P(X_i|\mathbf{Z}_{0:t}) &= P(X_i|\mathbf{Z}_{0:t-1}, \mathbf{z}_t) = \frac{P(\mathbf{z}_t, \mathbf{Z}_{0:t-1}, X_i)}{P(\mathbf{z}_t, \mathbf{Z}_{0:t-1})} \\ &= \frac{P(\mathbf{z}_t, \mathbf{Z}_{0:t-1}|X_i)P(X_i)}{P(\mathbf{z}_t, \mathbf{Z}_{0:t-1})}. \end{aligned} \quad (8)$$

Since observations in $\mathbf{Z}_{0:t-1}$ are independent of each other, we have $P(\mathbf{z}_t, \mathbf{Z}_{0:t-1}|X_i) = \prod_{t=0}^{t} P(\mathbf{z}_t|X_{i,t})$ and $P(\mathbf{z}_t, \mathbf{Z}_{0:t-1}) = \prod_{t=0}^{t} P(\mathbf{z}_t)$. Plug them into (8), we obtain

$$P(X_i|\mathbf{Z}_{0:t}) = \frac{P(X_i)\prod_{t=0}^{t} P(\mathbf{z}_t|X_{i,t})}{\prod_{t=0}^{t} P(\mathbf{z}_t)}. \quad (9)$$

Similarly, we obtain $P(X_i|\mathbf{Z}_{0:t-1})$,

$$P(X_i|\mathbf{Z}_{0:t-1}) = \frac{P(X_i)\prod_{t=0}^{t-1} P(\mathbf{z}_t|X_{i,t})}{\prod_{t=0}^{t-1} P(\mathbf{z}_t)}. \quad (10)$$

Combine (9) and (10), we have

$$\begin{aligned} \frac{P(X_i|\mathbf{Z}_{0:t})}{P(X_i|\mathbf{Z}_{0:t-1})} &= \frac{P(X_i)\prod_{t=0}^{t} P(\mathbf{z}_t|X_{i,t})}{P(X_i)\prod_{t=0}^{t-1} P(\mathbf{z}_t|X_{i,t})} \frac{\prod_{t=0}^{t-1} P(\mathbf{z}_t)}{\prod_{t=0}^{t} P(\mathbf{z}_t)} \\ &= \frac{P(\mathbf{z}_t|X_{i,t})}{P(\mathbf{z}_t)}. \end{aligned} \quad (11)$$

Plug $P(\mathbf{z}_t|X_{i,t}) = P(X_{i,t}|\mathbf{z}_t)P(\mathbf{z}_t)/P(X_{i,t})$ into (11) and we have

$$P(X_i|\mathbf{Z}_{0:t}) = \zeta P(X_{i,t}|\mathbf{z}_t)P(X_i|\mathbf{Z}_{0:t-1}), \quad (12)$$

where $\zeta$ is a normalization factor.

With more observations, we update the conditional probability $P(X_i|\mathbf{Z}_{0:t})$ for the $i$-th LM until it converges. For initialization, we set $P(X_i|\mathbf{Z}_{0:t-1})$ to be 1, and utilize (7) to update $P(X_i|\mathbf{Z}_{0:t})$ in (12).

We threshold $P(X_i|\mathbf{Z}_{0:t})$ to determine if the $i$-th LM is consistent or not. Define $\epsilon_u$ and $\epsilon_v$, $1 > \epsilon_u > \epsilon_v > 0$, as thresholds to determine if an LM is consistent or not. If $P(X_i|\mathbf{Z}_{0:t}) \geq \epsilon_u$, then the $i$-th LM is consistent; if $P(X_i|\mathbf{Z}_{0:t}) \leq \epsilon_v$, then the $i$-th LM is inconsistent; otherwise, the $i$-th LM ${}^{\mathcal{G}}\mathbf{M}_i$ is undetermined and we expect more observations to confirm its consistency.

**Algorithm 1:** Lane Marking Verification

**Input:** $\mathcal{M}_p, \mathbf{T}_{0:t}, \sum_{\boldsymbol{\alpha}_{0:t}}, \sum_{\mathbf{t}_{0:t}}, \mathbf{Z}_{0:t}$
**Output:** The $i$-th LM ${}^{\mathcal{G}}\mathbf{M}_i$ is consistent or not

| | |
|---|---|
| **for** $t \in \{0, 1, ..., t\}$ **do** | $O(t)$ |
| Obtain set $\mathcal{M}_m$ using (1); | $O(n_m \log n_m)$ |
| Compute ${}^{\mathcal{I}}\mathbf{x}_r$ through (2); | $O(1)$ |
| Generate $f_m({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_{m,t})$ by (6) ; | $O(n_g)$ |
| Get $f_s({}^{\mathcal{I}}\mathbf{x}|\mathbf{z}_t)$ through $\mathbf{z}_\tau$; | $O(n_h)$ |
| Attain $P(X_i|\mathbf{z}_t)$ in (7); | $O(1)$ |
| Obtain $P(X_i|\mathbf{Z}_{0:t-1})$ and $P(X_i|\mathbf{Z}_{0:t})$; | $O(1)$ |
| **if** $P(X_i|\mathbf{Z}_{0:t}) \geq \epsilon_u$ **then** | |
| Report ${}^{\mathcal{G}}\mathbf{M}_i$ as "consistent"; | $O(1)$ |
| Stop updating $P(X_i|\mathbf{Z}_{0:t})$; | $O(1)$ |
| **else if** $P(X_i|\mathbf{Z}_{0:t}) \leq \epsilon_v$ **then** | |
| Mark ${}^{\mathcal{G}}\mathbf{M}_i$ as "inconsistent"; | $O(1)$ |
| **else** | |
| Mark ${}^{\mathcal{G}}\mathbf{M}_i$ as "undetermined"; | $O(1)$ |

TABLE I
DATASETS FOR COMPARISON

| Dataset | Date | Length | #Images | Weather |
|---|---|---|---|---|
| A | 2019_07_01 | 634s | 4953 | Partially sunny |
| B | 2019_07_29 | 352s | 3088 | Light rain |

### D. Algorithm

We summarize our our LM verification algorithm in Algorithm 1. It is noted that we stop updating $P(X_i|\mathbf{Z}_{0:t})$ for the LM ${}^{\mathcal{G}}\mathbf{M}_i$ if $P(X_i|\mathbf{Z}_{0:t}) \geq \epsilon_u$, thus the computation complexity can be greatly decreased with the increasing number of consistent LMs. Besides, we also utilize a local database to store the LM belief every time with new observations to decrease the memory usage (see Box 3.1 in Fig. 2). If an LM's belief by using crowdsourced images is still below the pre-selected threshold and required to be removed, we query the HD map and remove the corresponding LM.

The computational complexity of our algorithm is,

*Lemma 1:* Our lane marking verification algorithm runs in $O(tn_m \log(n_m))$.

## V. EXPERIMENTS

We have implemented our algorithm in C++ under Ubuntu 16.04. It is tested on a Laptop PC with an Intel® Core™ i5-8265U CPU@1.60GHz and 8 GB RAM. We collect images using forward-looking cameras mounted on data collection vehicles. The data have been collected on the north segment of the 4th ring road in the Beijing. The vehicles runs on the same part of road back and forth at different days and different times. We collected two datasets with different weather conditions (see Tab. I). The image resolution is 300 $\times$ 480. We plan to release our data and algorithm output, a total of 14815 frames to the pubilc[1].

We set $\epsilon_u = 0.99$ and $\epsilon_v = 0.01$ in experiments. To evaluate the performance of our approach quantitatively,

[1]http://telerobot.cs.tamu.edu/lane/

TABLE II
EVALUATION USING REAL DATA

| Dataset | Methods | Precision | Recall | F1-measure |
|---|---|---|---|---|
| A | Ours | **91.13%** | **92.47%** | **91.80%** |
| | r-IoU | 88.32% | 90.13% | 89.22% |
| | p-IoU | 76.24% | 79.38% | 77.78% |
| B | Ours | **92.36%** | **93.75%** | **93.05%** |
| | r-IoU | 87.22% | 88.17% | 87.69% |
| | p-IoU | 72.13% | 75.26% | 73.66% |

three metrics, including precision, recall, and F1-measure [28], are employed. The F1-measure is the harmonic mean of precision and recall. Recall that set $\mathbf{z}_{m,t}$ contains all the project pixels for the $i$-th LM and $\mathbf{z}_t$ has the LM pixels extracted from the images. For comparison, we compare our algorithms to the following approaches.

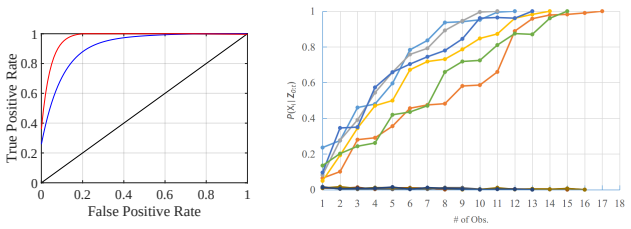
- r-IoU: we replace (7) using a common similarity metric: intersection over union (IoU),

$$P(X_{i,t}|\mathbf{z}_t) = \frac{|\mathbf{z}_{m,t} \bigcap \mathbf{z}_t|}{|\mathbf{z}_{m,t} \bigcup \mathbf{z}_t|}.$$

- p-IoU: we project the all the historical LM pixels from the camera images for the ${}^{\mathcal{G}}\mathbf{M}_i$ to the latest image frame $\mathbf{I}_T$ and use IoU metric above to verify its existence.

The experimental results in Tab. II show that our approach outperforms the r-IoU or p-IoU based approach. Set similarity based approach generates a relatively lower $P(X_{i,t}|\mathbf{z}_t)$ by disregarding potential pixels that belong to the LM and require more observations than our algorithm even for consistent LMs. Besides, the p-IoU based method ignores the camera pose uncertainty, map accuracy and lane marking pixel noise, and yields poor verification performance.

In Fig. 4(a), we plot the receiver operating characteristic (ROC) curves for each method by varying their respective thresholds. In the ROC plane, the upper left corner represents the ideal result. It is clear that our method outperforms the IoU by a large margin. In fact, this is not surprising because IoU produces too many false negatives.




(a) (b)

Fig. 4. (a) ROC curve for lane marking verification in comparison. The curve in red is by our method, and the curve in blue is from IoU-based approach (best viewed in color). (b) Examples of LM belief adjustment with more and more observations.

We present ten example LMs from the HD map to illustrate the belief update process: six consistent LMs have been identified and kept in the map by our algorithm and four LMs have been identified as inconsistent and hence removed. We plot their belief changes as the number of the observations increase in Fig. 4(b). Initially, most LMs

are "undetermined" due to the lack of enough observations. With more and more incoming observations, LM statuses converge to either "consistent" or "inconsistent." For an LM that is kept in the map, it is clear that $P(X_i|\mathbf{Z}_{0:t})$ grows monotonically toward $\epsilon_u$. For LM that is inconsistent with the map, its belief is mostly at a lower level all the times and below the threshold $\epsilon_v$, as expected.

## VI. Conclusion and Future Work

Here we presented an algorithm for updating LMs in HD maps using crowdsourced images. Realizing LMs in both the HD map and camera images contain noises, we modeled them respectively as observations of two LM spatial distributions in the camera frame and check if they agreed with each other via a goodness of fit test. We modeled the Gaussian belief functions by considering noises from camera motion. We derived a sequential Bayes' model to allow the belief functions to be updated using crowdsourced images. We implemented and tested our algorithms using data collected from testing vehicles and the results showed that our approach is successful and outperformed the counterpart. In the future, we will extend approach to other objects in HD maps to ensure HD maps can be kept up-to-date at low cost.

## Acknowledgment

Thanks to H. Cheng, S. Yeh, A. Angert, D. Wang, S. Xie and Z. Wang for their inputs and contributions to the NetBot Laboratory, Texas A&M University.

## References

[1] M. Schreiber, C. Knöppel, and U. Franke, "Laneloc: Lane marking based localization using highly accurate maps," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 449–454.

[2] A. Zang, Z. Li, D. Doria, and G. Trajcevski, "Accurate vehicle self-localization in high definition map dataset," in *Proceedings of the 1st ACM SIGSPATIAL Workshop on High-Precision Maps and Intelligent Applications for Autonomous Vehicles*, 2017, pp. 1–8.

[3] M. Aldibaja, N. Suganuma, and K. Yoneda, "Lidar-data accumulation strategy to generate high definition maps for autonomous vehicles," in *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2017, pp. 422–428.

[4] H. Chu, L. Guo, B. Gao, H. Chen, N. Bian, and J. Zhou, "Predictive cruise control using high-definition map and real vehicle implementation," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 377–11 389, 2018.

[5] K. Abdulrahim and R. A. Salam, "Traffic surveillance: A review of vision based vehicle detection, recognition and tracking," *International journal of applied engineering research*, vol. 11, no. 1, pp. 713–726, 2016.

[6] A. Petrovai, R. Danescu, and S. Nedevschi, "A stereovision based approach for detecting and tracking lane and forward obstacles on mobile devices," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*. IEEE, 2015, pp. 634–641.

[7] A. Joshi and M. R. James, "Generation of accurate lane-level maps from coarse prior maps and lidar," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 19–29, 2015.

[8] D. C. Andrade, F. Bueno, F. R. Franco, R. A. Silva, J. H. Z. Neme, E. Margraf, W. T. Omoto, F. A. Farinelli, A. M. Tusset, S. Okida *et al.*, "A novel strategy for road lane detection and tracking based on a vehicle's forward monocular camera," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–11, 2018.

[9] B. Li, D. Song, H. Li, A. Pike, and P. Carlson, "Lane marking quality assessment for autonomous driving," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, October, 1-5, 2018.

[10] B. Li, D. Song, A. Ramchandani, H.-M. Cheng, D. Wang, Y. Xu, and B. Chen, "Virtual lane boundary generation for human-compatible autonomous driving: A tight coupling between perception and planning," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019.

[11] A. S. Huang, D. Moore, M. Antone, E. Olson, and S. Teller, "Finding multiple lanes in urban road networks with vision and lidar," *Autonomous Robots*, vol. 26, no. 2-3, pp. 103–122, 2009.

[12] Y. Kang, C. Roh, S.-B. Suh, and B. Song, "A lidar-based decision-making method for road boundary detection using multiple kalman filters," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4360–4368, 2012.

[13] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 3, pp. 690–703, 2016.

[14] A. Elfes, "Occupancy grids: A probabilistic framework for robot perception and navigation," Ph.D. dissertation, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1989.

[15] H. Moravec, "Sensor fusion in certainty grids for mobile robots," *AI Magazine*, no. 9, pp. 61–74, 1988.

[16] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 2007, pp. 225–234.

[17] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[18] Y. Lu and D. Song, "Visual navigation using heterogeneous landmarks and unsupervised geometric constraints," in *IEEE Transactions on Robotics (T-RO)*, vol. 31, no. 3, June 2015, pp. 736 —- 749.

[19] D. Hahnel, W. Burgard, D. Fox, and S. Thrun, "An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements," in *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 1. IEEE, 2003, pp. 206–211.

[20] J. Ryde and N. Hillier, "Alignment and 3d scene change detection for segmentation in autonomous earth moving," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 1484–1490.

[21] A. Aijazi, P. Checchin, and L. Trassoudaine, "Automatic removal of imperfections and change detection for accurate 3d urban cartography by classification and incremental updating," *Remote Sensing*, vol. 5, no. 8, pp. 3701–3728, 2013.

[22] C.-C. Wang, C. Thorpe, S. Thrun, M. Hebert, and H. Durrant-Whyte, "Simultaneous localization, mapping and moving object tracking," *The International Journal of Robotics Research*, vol. 26, no. 9, pp. 889–916, 2007.

[23] J. S. Berrio, J. Ward, S. Worrall, and E. Nebot, "Identifying robust landmarks in feature-based maps," *arXiv preprint arXiv:1809.09774*, 2018.

[24] L. Sun, Z. Yan, A. Zaganidis, C. Zhao, and T. Duckett, "Recurrent-octomap: Learning state-based map refinement for long-term semantic mapping with 3-d-lidar data," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3749–3756, 2018.

[25] A. Nurminen and A. Oulasvirta, "Designing interactions for navigation in 3d mobile maps," in *Map-based mobile services*. Springer, 2008, pp. 198–227.

[26] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[27] W. G. Cochran, "The $\chi 2$ test of goodness of fit," *The Annals of Mathematical Statistics*, pp. 315–345, 1952.

[28] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.