Bayesian Simultaneous Localization and Multi-Lane Tracking Using Onboard Sensors and a SD Map

Yuxuan Xia[†*], Erik Stenborg[†], Junsheng Fu[†], and Gustaf Hendeby[*]

[†]Zenseact, Gothenburg, Sweden
[*]Department of Electrical Engineering, Linköping University, Linköping, Sweden
Email: [†]firstname.lastname@zenseact.com, [*]firstname.lastname@liu.se

Abstract—High-definition map with accurate lane-level information is crucial for autonomous driving, but the creation of these maps is a resource-intensive process. To this end, we present a cost-effective solution to create lane-level roadmaps using only the global navigation satellite system (GNSS) and a camera on customer vehicles. Our proposed solution utilizes a prior standard-definition (SD) map, GNSS measurements, visual odometry, and lane marking edge detection points, to simultaneously estimate the vehicle's 6D pose, its position within a SD map, and also the 3D geometry of traffic lines. This is achieved using a Bayesian simultaneous localization and multi-object tracking filter, where the estimation of traffic lines is formulated as a multiple extended object tracking problem, solved using a trajectory Poisson multi-Bernoulli mixture (TPMBM) filter. In TPMBM filtering, traffic lines are modeled using B-spline trajectories, and each trajectory is parameterized by a sequence of control points. The proposed solution has been evaluated using experimental data collected by a test vehicle driving on highway. Preliminary results show that the traffic line estimates, overlaid on the satellite image, generally align with the lane markings up to some lateral offsets.

Index Terms—Autonomous driving, lane-level mapping, multi-lane tracking, simultaneous localization and mapping, B-spline.

I. INTRODUCTION

High-definition (HD) maps with detailed lane-level information are an enabling factor for autonomous driving, providing the necessary precision, safety, and efficiency in localization, navigation, and decision-making processes [1]. However, the creation of HD maps is an expensive and resource-intensive process, which typically involves using vehicles equipped with high-precision equipment, such as lidar, to continuously collect and process the data. An appealing, cost-effective alternative is to instead use onboard sensors available in massive customer vehicles, such as the global navigation satellite system (GNSS) and camera, to jointly create large-scale, lane-level geometric maps. This approach is known as crowdsourced mapping [2], whereby multiple sources of low-fidelity data are aggregated to create high-fidelity maps.

In this paper, we focus on the problem of creating lane-level geometric maps, which are built upon standard-definition (SD) maps with low geometric fidelity, using single-antenna GNSS measurements and mono camera data collected from a single vehicle traversal. This can serve as a solid foundation for the follow-up map change detection, update, and merging tasks. In particular, we aim to estimate the 3D geometry of traffic lines using lane marking edge detection points obtained from a computer vision algorithm. Since the accurate estimation of traffic

line geometries relies on the precise vehicle pose, which could hardly be obtained using only GNSS, the vehicle's pose and the traffic line geometries need to be simultaneously estimated. To this end, we propose an effective solution that leverages a prior SD map, GNSS measurements, visual odometry, and lane marking edge detection points to simultaneously estimate the vehicle's 6D pose, its position within a SD map, and the 3D geometry of traffic lines.

Our proposed solution is achieved using a Bayesian simultaneous localization and multi-object tracking (SLAMOT) filter, where the estimation of traffic lines is formulated as a multiple extended object tracking (EOT) problem. Note that, although traffic lines remain in fact static, in most cases they can only be partially observed by the camera due to its limited field-of-view. Thus, a dynamic model for the traffic lines is needed to make predictions of their unseen geometries before each measurement update. The multiple EOT problem is addressed using a trajectory Poisson multi-Bernoulli mixture (TPMBM) filter, which is a state-of-the-art method for tracking objects that may generate multiple detections per sensor scan [3], [4].

In Bayesian filtering, the state of each traffic line is modeled as a uniform B-spline trajectory, parameterized by a sequence of control points. Compared with models based on polylines or clothoids, B-splines can offer greater flexibility and smoothness, making them well-suited for modeling traffic lines with varying curvature [5]. The B-spline trajectory representations also integrate smoothly with the sets of trajectories framework [6], and a Bernoulli object spawning model can be easily incorporated to handle lane splitting [7]. Furthermore, by leveraging the road orientation information provided by the SD map in the lane marking measurement model, we can achieve recursive B-spline estimation [8] under Gaussian approximations using a linear Kalman filter.

A. Related Work

In many existing works, e.g., [9]–[12], the problem of lane-level map generation using only onboard sensors is typically addressed using graph-based optimization, whose performance can be highly dependent on the quality of data associations of lane marking detection points. However, the data association problem in these works is only addressed using simple heuristics, such as the nearest neighbor approach, which may lead to inaccurate estimates in some challenging cases, e.g., when the vehicle is moving on roads with multiple parallel dashed lane markings. In contrast, our proposed solution uses a Bayesian SLAMOT filter, which can handle the data association problem

in a principled way by leveraging data association techniques developed for multiple EOT.

Multi-lane estimation has also been formulated as a multiple EOT problem in [13] with polyline lane shape modeling and in [14] with Gaussian process lane shape modeling. However, these works only consider multi-lane estimation on a frame-by-frame basis, and they do not provide continuous, multi-lane estimates over time, and thus cannot generate lane-level maps. The modeling of each traffic line as a spline trajectory is also considered in [15], [16] using Catmull-Rom splines. Compared to B-splines, Catmull-Rom splines do not guarantee curvature continuity, which means that the transition between segments may not be smooth.

B. Contributions

Our proposed method stands out from the existing works by providing a fully Bayesian solution to generate lane-level maps using onboard sensors. Importantly, the multi-trajectory posterior density captures the uncertainties in both the number of traffic lines and their geometries, which are important for fusing the maps generated from multiple vehicle traversals.

The main contributions of this paper are as follows:

1) We present a TPMBM-based SLAMOT filter and apply it to the problem of simultaneous localization and multi-lane tracking using onboard sensors and a SD map.
2) We integrate recursive B-spline estimation into the extended object TPMBM filter to estimate the 3D geometry of traffic lines using lane marking detection points.
3) The efficacy of proposed solution has been verified using real-world data.

The rest of the paper is organized as follows. In Section II, we describe the system model. We present the problem formulation in Section III and the TPMBM-based SLAMOT filter in Section IV. In Section V, we present the experimental results. Finally, we conclude the paper in Section VI.

II. SYSTEM MODEL

In this section, we introduce the system modeling, which includes the SD map, the state representations of vehicle and traffic lines, as well as their dynamic and measurement models. We assume that the vision data has already been transformed from the camera coordinate to the vehicle coordinate. States and measurements in the vehicle coordinate are denoted with a superscript l, and, if not otherwise specified, they are given in an East-North-Up coordinate relative to a local origin.

We use $k \in \mathbb{N}_0$ to denote discrete time steps, and the time instant at time step k is t_k, with uniform time interval $\tau = t_{k+1} - t_k$. The cardinality of a set $\mathbf{x}$ is denoted $|\mathbf{x}|$. We use $\uplus$ to denote the union of sets that are mutually disjoint and $\otimes$ to denote the Kronecker product. In addition, we use $\delta_x[\cdot]$ and $\delta_x(\cdot)$ to represent the Kronecker delta and the Dirac delta functions, respectively, centered at x.

A. SD Map

The SD map can be represented as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges.

Each node $v_i \in \mathcal{V}$ is associated with a 2D position $[x_i, y_i]^T \in \mathbb{R}^2$, and each edge $e_{i,j} \in \mathcal{E}$, with start and end nodes v_i and v_j, represents a street segment with direction from v_i to v_j. Each edge $e_{i,j}$ is described by its length $l_{e_{i,j}}$ and its orientation $\phi_{e_{i,j}}$ via $l_{e_{i,j}} = \|(x_j - x_i, y_j - y_i)\|_2$ and $\phi_{e_{i,j}} = \arctan\left(\frac{y_j - y_i}{x_j - x_i}\right)$. As in [17], all the street segments are assumed to be one-way, and two-way streets are converted to two one-way streets.

B. Vehicle State

The vehicle's 6D pose $p_k = [x_k, y_k, z_k, \alpha_k, \beta_k, \gamma_k]^T$ at time step k consists of its 3D position $[x_k, y_k, z_k]^T \in \mathbb{R}^3$, and its yaw α_k, pitch β_k, and roll γ_k. The position of the vehicle $q_k = [e_k, d_k]^T$ within a SD map is defined by the street segment $e_k \in \mathcal{E}$ that the vehicle is moving on and the distance d_k from the start node of e_k. The inclusion of d_k to the vehicle's state representation may seem redundant, but as we shall see later, it is helpful for modeling the transition of e_k. Also note that, as the geometrical information provided by SD map may be inaccurate, the distance d_k may differ from the actual distance between the vehicle's position (projected onto e_k) and the start node of e_k. The complete vehicle state $o_k = [p_k^T, q_k^T]^T$ at time step k is then 8D, and it consists of the vehicle's global 6D pose p_k and its location q_k within a SD map $\mathcal{G}$.

1) Vehicle Measurement Model: In this work, we assume that the GNSS only measures[1] the vehicle's 3D position, and the measurement at time step k is denoted as ζ_k. Assuming that the measurement noise is zero-mean Gaussian, the measurement model of GNSS is given by

$$\ell_k(\zeta_k|p_k) = \mathcal{N}\left(\zeta_k; [x_k, y_k, z_k]^T, \Omega_k\right), \quad (1)$$

where Ω_k is the measurement noise covariance.

In addition to GNSS, the SD map $\mathcal{G}$ can also be regarded as a measurement source, in the sense that the vehicle's yaw α_k should not diverge too much from the orientation of the street segment e_k that the vehicle is moving on. This measurement likelihood is specified by a von Mises distribution [18]

$$\ell_k(\mathcal{G}|o_k) = \mathcal{VM}(\phi_{e_k}|\alpha_k, \kappa), \quad (2)$$

where $\mathcal{VM}(\cdot|\mu, \kappa)$ denotes a von Mises distribution with mean μ and concentration parameter κ.

2) Vehicle Motion Model: In this work, we assume that the vehicle motion estimates obtained from the visual odometry can be used as control input[2], and that the SD map does not impose any constraints on the vehicle's motion. The vehicle state transition density can then be factorized as

$$\begin{aligned} g_{k+1}&\left(p_{k+1}, q_{k+1}|p_k, q_k, u_{k+1}^l, \mathcal{G}\right) \\ &= g_{k+1}^p\left(p_{k+1}|p_k, u_{k+1}^l\right) g_{k+1}^q(q_{k+1}|p_{k+1}, q_k, \mathcal{G}), \quad (3) \end{aligned}$$

where u_{k+1}^l represents the control input in the vehicle coordinate at time step $k+1$, and it can be transformed to the global

---

[1]The GNSS operates at a lower frequency than the camera. We assume that the GNSS data has been interpolated to be synchronized with the vision data. It is also possible to directly work with low-frequency GNSS data by considering continuous-discrete particle filtering.

[2]Vehicle odometry from wheel encoders is used if visual odometry is not available, e.g., when the light condition is poor.

coordinate u_{k+1} using the vehicle's pose p_k. Assuming that the process noise is zero-mean Gaussian, the transition density of the vehicle's pose at time step $k+1$ is given by

$$g_{k+1}^p \left(p_{k+1}|p_k, u_{k+1}^l\right) = \mathcal{N}(p_{k+1}; p_k + u_{k+1}, Q_{k+1}), \quad (4)$$

where Q_{k+1} is the motion noise covariance.

The transition density $g_{k+1}^q(\cdot|p_{k+1}, q_k, \mathcal{G})$ of the vehicle's position within a SD map needs to consider the following two cases depending on the distance the vehicle has moved on the street segment: 1) the vehicle stays on the same street segment, and 2) the vehicle moves to a new street segment. In the first case, we have $e_{k+1} = e_k$, and the vehicle's predicted distance to the start node of e_k is modeled as $d_{k+1} = d_k + \epsilon_{k+1} + \sigma_{k+1}$, where ϵ_{k+1} is the vehicle's moved distance along the street segment e_k, given by

$$\epsilon_{k+1} = \cos(\phi_{e_k})(x_{k+1} - x_k) + \sin(\phi_{e_k})(y_{k+1} - y_k), \quad (5)$$

and σ_{k+1} is a zero-mean Gaussian noise. Then for $d_k + \epsilon_{k+1} \leq l_{e_k}$, we have

$$g_{k+1}^q(q_{k+1}|p_{k+1}, q_k, \mathcal{G}) = \delta_{e_k}[e_{k+1}]\mathcal{N}(d_{k+1}; d_k + \epsilon_{k+1}, \sigma_{k+1}). \quad (6)$$

In the second case, we have $e_{k+1} \in N(e_k)$, where $N(e_k)$ is the set of street segments to which e_k connects. Following [18], we assume that the difference between the orientation of the new street segment e_{k+1} and the vehicle's predicted yaw α_{k+1} follows a zero-mean von Mises distribution. That is, for each $e \in N(e_k)$,

$$P(e_{k+1} = e) \propto \mathcal{VM}(\phi_e|\alpha_{k+1}, \kappa). \quad (7)$$

After moving to e_{k+1}, the vehicle's distance d_{k+1} to the start node of e_{k+1} needs to have l_{e_k} subtracted, and it is modeled as $d_{k+1} = d_k + \epsilon_{k+1} - l_{e_k} + \sigma_{k+1}$ [17]. Then for $d_k + \epsilon_{k+1} > l_{e_k}$, $e \in N(e_k)$, we have

$$g_{k+1}^q(q_{k+1}|p_{k+1}, q_k, \mathcal{G})$$
$$= P(e_{k+1} = e)\mathcal{N}(d_{k+1}; d_k + \epsilon_{k+1} - l_{e_k}, \sigma_{k+1}). \quad (8)$$

C. Traffic Line State

A traffic line is modeled as a B-spline with uniform knots. A B-spline of degree d has minimum $d+1$ control points and continuous time derivatives up to the $(d-1)$-th order. In this work, we consider 3D quadratic B-splines with $d = 2$, which we find sufficient for modeling lane geometry.

A B-spline trajectory can be parameterized by $C = (\varepsilon, c^{1:\nu})$, where ε is the initial time step of the trajectory C, $\nu \geq 3$ is its length, and $c^{1:\nu} = (c^1, \ldots, c^\nu)$, with $c^i \in \mathbb{R}^3$, denotes a finite sequence of control points. In the sequence $c^{1:\nu}$, the first three control points $c^{1:3}$ correspond to time instant $t_\varepsilon^c = t_\varepsilon$, and the rests $c^{4:\nu}$ are at time instants $t_{\varepsilon+3}^c, \ldots, t_{\varepsilon+\nu-1}^c$, with uniform interval $\tau^c \geq \tau$. Note that different from the trajectory representation in [6], here each trajectory has minimum length 3, and the time interval τ^c for placing the control points can be larger than the measurement sampling interval τ [8].

The trajectory in continuous time can be obtained by interpolating the control points using the B-spline basis function.

Specifically, given $C = (\varepsilon, c^{1:\nu})$, the position of a point on the trajectory at an arbitrary time instant $\varsigma \in [t_{\varepsilon+i-1}^c, t_{\varepsilon+i}^c), i \in \{1, \ldots, \nu - 2\}$, is determined by the subsequence of control points $c^{i:i+2}$, according to

$$c^\varsigma = \left[c^i, c^{i+1}, c^{i+2}\right] \cdot \Sigma \cdot \iota(\varsigma, t_{\varepsilon+i-1}^c), \quad (9a)$$

$$\iota(\varsigma, t_{\varepsilon+i-1}^c) = \left[1, \frac{\varsigma - t_{\varepsilon+i-1}^c}{\tau^c}, \left(\frac{\varsigma - t_{\varepsilon+i-1}^c}{\tau^c}\right)^2\right]^T, \quad (9b)$$

$$\Sigma = \frac{1}{2}\begin{bmatrix} 1 & -2 & 1 \\ 1 & 2 & -2 \\ 0 & 0 & 1 \end{bmatrix}. \quad (9c)$$

where Σ is the basis function matrix for quadratic B-splines.

We consider B-spline trajectories up to the current time step k. Trajectory $C = (\varepsilon, c^{1:\nu})$ is considered alive at time step k if and only if the latest measurement time instant t_k satisfies that $t_k \in [t_{\varepsilon+\nu-3}^c, t_{\varepsilon+\nu-2}^c)$, i.e., $\nu - \lfloor \frac{\tau(k-\varepsilon)}{\tau^c} \rfloor = 3$. In the rest of the paper, we use the notation $\lfloor \frac{\tau(k-\varepsilon)}{\tau^c} \rfloor \triangleq \Xi_{k,\varepsilon}^{\tau,\tau^c}$ for brevity, where $\Xi_{k,\varepsilon}^{\tau,\tau^c} + 3$ represents the maximum length of trajectory $C = (\varepsilon, c^{1:\nu})$ at time step k. Then, the space of a single trajectory up to time step k can be defined as $T_{(k)} = \uplus_{(\varepsilon,\nu) \in I_{(k)}} \{\varepsilon\} \times \mathbb{R}^{3\nu}$, where $I_{(k)} = \{(\varepsilon, \nu) : 0 \leq \varepsilon \leq k \text{ and } 3 \leq \nu \leq 3 + \Xi_{k,\varepsilon}^{\tau,\tau^c}\}$. The set of all trajectories that have existed up to time step k is denoted as $\mathbf{C}_k$. Definitions of integrals and densities of single trajectories and sets of trajectories can be found in [6].

1) Multi-Lane Measurement Model: From (9), we can see that, for an alive trajectory $C = (\varepsilon, c^{1:\nu})$ at time step k, its interpolation at time step k is determined by its latest three control points $c^{\nu-2:\nu}$. This means that the detections of C at time step k only depend on $c^{\nu-2:\nu}$ (and the vehicle state).

A single B-spline trajectory $C = (\varepsilon, c^{1:\nu})$ at time step k is detected with probability

$$P_k^D(C, p_k) = \begin{cases} P_k^D\left(c^{\nu-2:\nu}, p_k\right), & \nu - \Xi_{k,\varepsilon}^{\tau,\tau^c} = 3, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

and generates an independent set $\mathbf{w}_k^{L,l} \in \mathbb{R}^3$ of lane marking edge detection points in the vehicle coordinate, or is misdetected with probability $1 - P_k^D(C, p_k)$. For a detection point w_k^l, it can be transformed to the global coordinate $w_k \in \mathbf{w}_k^L$ using the vehicle's pose p_k, and the dependence of w_k on p_k is made implicit in the following expressions.

We assume that each individual lane marking edge detection point w_k originates from a measurement source ϖ_k, corrupted by a zero-mean Gaussian noise with covariance $\Omega_k^{w_k}$. Each measurement source ϖ_k is considered uniformly distributed along the two lane marking edges. This modeling assumption allows us to rewrite the single measurement likelihood as

$$\ell_k\left(w_k|\varepsilon, c^{\nu-2:\nu}\right) = \int \mathcal{N}\left(w_k; \varpi_k, \Omega_k^{w_k}\right)\mathcal{U}\left(\varpi_k|\varepsilon, c^{\nu-2:\nu}\right) d\varpi_k. \quad (11)$$

To enable the tractable computation of (11), we further approximate the uniform distribution $\mathcal{U}(\varpi_k|\varepsilon, c^{\nu-2:\nu})$ as a Gaussian $\mathcal{N}(\varpi_k; h(\varepsilon, c^{\nu-2:\nu}), \Omega_k^\varpi)$, which yields

$$\ell_k\left(w_k|\varepsilon, c^{\nu-2:\nu}\right) = \mathcal{N}\left(w_k; h\left(\varepsilon, c^{\nu-2:\nu}\right), \Omega_k^{w_k} + \Omega_k^\varpi\right). \quad (12)$$

In Gaussian distribution $\mathcal{N}(\varpi_k; h(\varepsilon, c^{\nu-2:\nu}), \Omega_k^{\varpi})$, the mean $h(\varepsilon, c^{\nu-2:\nu})$ is the interpolated point on trajectory C at time step k, and it can be computed using (9), which gives

$$h(\varepsilon, c^{\nu-2:\nu}) = \left(\Sigma \cdot \iota(t_k, t^c_{\varepsilon+\nu-3})\right)^T \otimes I_3 c^{\nu-2:\nu}, \qquad (13)$$

where I_3 is an identity matrix. The covariance matrix Ω_k^{ϖ} can be factorized as [19]

$$\Omega_k^{\varpi} = R_k^{\varpi}\mathrm{diag}\left([l_x, l_y, 0]/2\right)\cdot\Lambda\cdot\mathrm{diag}\left([l_x, l_y, 0]/2\right)R_k^{\varpi T}, \qquad (14)$$

where R_k^{ϖ} is a 3D rotation matrix that aligns the measurement source ϖ_k with the orientation of the lane marking, and Λ is a diagonal matrix, which can be set to $\Lambda = I_3/4$ to match the Gaussian distribution to a uniform distribution [19]. The factorization (14) is motivated by the fact that each observed lane marking within a certain distance of the vehicle can be approximated as a very thin rectangle with length l_x and width l_y. The length l_x setting should consider the trade-off between detection range and approximation error due to road curvature, whereas the width l_y can be determined by lane marking type.

The rotation matrix R_k^{ϖ} in (14) depends on the orientation of the (observed) lane marking at time step k, which can be approximately obtained by computing the tangent vector of the velocity of trajectory C (the first derivative of (9) with respect to time) at time step k. However, in this case the measurement noise covariance Ω_k^{ϖ} would become state-dependent, resulting in a (highly) nonlinear measurement model [20].

To compute R_k^{ϖ} in a simple yet effective way, we assume that the yaw of the B-spline trajectory evaluated at the interpolated point $h(\varepsilon, c^{\nu-2:\nu})$ is given by the orientation $\phi_{e'}$ of the street segment e' that $h(\varepsilon, c^{\nu-2:\nu})$ is on, and that the pitch and roll of the B-spline trajectory evaluated at $h(\varepsilon, c^{\nu-2:\nu})$ align with the vehicle's. Under these assumptions, we have that

$$R_k^{\varpi} = R(\phi_{e'}, \beta_k, \gamma_k), \qquad (15)$$

where $R(\cdot)$ is a 3D rotation matrix determined by yaw, pitch, and roll. This assumption is reasonable for typical cases where the lane markings are in parallel with the street segment and the street segments are flat within a short range of distance. To find e', we first compute the distance ϵ' between the vehicle's position and the interpolated point $h(\varepsilon, c^{\nu-2:\nu})$ projected on the street segment e_k that the vehicle is currently moving on, similar to (5). If $d_k + \epsilon' \le l_{e_k}$, we have $e' = e_k$. Otherwise, we find the street segment $e \in N(e_k)$ that maximizes (7) and check if $d_k + \epsilon' - l_{e_k} \le l_e$. This process is repeated until we find e'. In addition to the detection points generated by lane markings, we assume that the lane marking detector can also report clutter detection points. The set $\mathbf{w}_k^{C,l}$ of clutter detection points is modeled as a Poisson point process (PPP) with Poisson clutter rate λ_k^C. The set of all detection points at time step k is denoted as $\mathbf{w}_k^l$.

2) Multi-Lane Dynamic Model: Given the set of all trajectories $\mathbf{C}_k$ at time step k, each $C = (\varepsilon, c^{1:\nu}) \in \mathbf{C}_k$ survives to time step $k+1$, i.e., remains in the set $\mathbf{C}_{k+1}$, with probability one, and its transition density is given by

$$g_{k+1}\left(\varepsilon_+, c_+^{1:\nu_+}|C\right) = \delta_\varepsilon[\varepsilon_+]$$

$$\times \begin{cases} \delta_\nu[\nu_+]\delta_{c^{1:\nu}}\left(c_+^{1:\nu}\right), & \begin{aligned}&\nu < 3 + \Xi_{k,\varepsilon}^{\tau,\tau^c} \text{ or,}\\ &\nu = 3 + \Xi_{k,\varepsilon}^{\tau,\tau^c} \text{ and}\\ &t_{k+1} < t^c_{\varepsilon+\nu-2},\end{aligned}\\[2em] \begin{aligned}&\delta_\nu[\nu_+]\delta_{c^{1:\nu}}\left(c_+^{1:\nu}\right)\\ &\times\left(1 - P_1^S(c^\nu)\right),\end{aligned} & \begin{aligned}&\nu = 3 + \Xi_{k,\varepsilon}^{\tau,\tau^c} \text{ and}\\ &t_{k+1} \ge t^c_{\varepsilon+\nu-2} \text{ and}\\ &t_{k+1} \ge t^c_{\varepsilon+\nu_+-2},\end{aligned}\\[2em] \begin{aligned}&\delta_{\nu+1}[\nu_+]\delta_{c^{1:\nu}}\left(c_+^{1:\nu_+-1}\right)\\ &\times P_1^S(c^\nu)g\left(c_+^{\nu_+}|c^{\nu-1:\nu}\right),\end{aligned} & \begin{aligned}&\nu = 3 + \Xi_{k,\varepsilon}^{\tau,\tau^c} \text{ and}\\ &t_{k+1} \ge t^c_{\varepsilon+\nu-2} \text{ and}\\ &t_{k+1} < t^c_{\varepsilon+\nu_+-2}.\end{aligned} \end{cases} \qquad (16)$$

From (16), we can see that if trajectory C is not alive at time step k, or if it is alive at time step k but the time instant t_{k+1} is smaller than $t^c_{\varepsilon+\nu-2}$, then trajectory C remains unaltered with probability one. If trajectory C is alive at time step k and the time instant t_{k+1} is no smaller than $t^c_{\varepsilon+\nu-2}$, then with probability $1 - P_1^S(c^\nu)$, trajectory C remains unaltered, and with probability $P_1^S(c^\nu)$, it is extended by appending one control point with a transition density $g(c_+^{\nu+}|c^{\nu-1:\nu})$.

The state of the new control point $c^{\nu+}$ in $g(c_+^{\nu+}|c^{\nu-1:\nu})$ is determined by the state of the latest two control points $c^{\nu-1:\nu}$ by preserving the velocity at time instant $t^c_{\varepsilon+\nu-2}$, which gives $c_+^{\nu+} = 2c^\nu - c^{\nu-1}$. Assuming that the motion process noise is zero-mean Gaussian, the transition density is

$$g\left(c_+^{\nu+}|c^{\nu-1:\nu}\right) = \mathcal{N}\left(c_+^{\nu+}; \begin{bmatrix} -I_3 & 2I_3 \end{bmatrix} c^{\nu-1:\nu}, Q_{k+1}^c\right), \qquad (17)$$

where Q_{k+1}^c is the motion noise covariance.

The set of trajectories $\mathbf{C}_{k+1}$ at time step $k+1$ is the union of surviving trajectories, newborn trajectories, and trajectories spawned from existing trajectories. The set of newborn trajectories appears independently with a PPP with intensity

$$\lambda_{k+1}^B(\varepsilon, c^{1:\nu}|p_{k+1}) = \delta_{k+1}[\varepsilon]\delta_3[\nu]\lambda_{k+1}^B(c^{1:3}|p_{k+1}), \qquad (18)$$

where the Poisson intensity $\lambda_{k+1}^B(c^{1:3}|p_{k+1})$ depends on the camera's field-of-view. To model lane splitting, we assume that each trajectory $C = (\varepsilon, c^{1:\nu}) \in \mathbf{C}_k$ alive at time step k may spawn a new trajectory with probability $P_2^S(c^\nu)$. The set of trajectories spawned from trajectory C, where $\nu - \Xi_{k,\varepsilon}^{\tau,\tau^c} = 3$, at time step $k+1$ is then a Bernoulli process with density

$$g_{k+1}^S\left(\mathbf{C}_+|C\right)$$
$$= \begin{cases} \begin{aligned}&P_2^S(c^\nu)\delta_{k+1}[\varepsilon_+]\delta_3[\nu_+]\\ &\times \delta_{c^{\nu-1:\nu}}\left(c_+^{1:\nu_+-1}\right)\\ &\times g\left(c_+^{\nu+}|c^{\nu-1:\nu}\right),\end{aligned} & \mathbf{C}_+ = \left\{\left(\varepsilon_+, c_+^{1:\nu_+}\right)\right\},\\ 1 - P_2^S(c^\nu), & \mathbf{C}_+ = \emptyset,\\ 0, & \text{otherwise.} \end{cases} \qquad (19)$$

III. PROBLEM FORMULATION

With the system model introduced in Section II in place, we can now formulate the problem of simultaneous localization and multi-lane tracking. The objective is to recursively compute the joint posterior distribution of the vehicle trajectory $o_{0:k}$ and the set $\mathbf{C}_k$ of all B-spline trajectories given the SD map $\mathcal{G}$ and the measurements $\mathbf{w}_{1:k}^l$ and visual odometry $u_{1:k}^l$

up to and including time step k. The joint posterior distribution can be factorized as the product of the posterior distribution $f(o_{0:k}|\zeta_{1:k}, u_{1:k}^l, \mathcal{G})$ of the vehicle trajectory and the posterior distribution $f(\mathbf{C}_k|o_{0:k}, \mathbf{w}_{1:k}^l, \mathcal{G})$ of the set of all trajectories conditioned on the vehicle trajectory, i.e.,

$$f\left(o_{0:k}, \mathbf{C}_k|\zeta_{1:k}, u_{1:k}^l, \mathbf{w}_{1:k}^l, \mathcal{G}\right)$$
$$= f\left(o_{0:k}|\zeta_{1:k}, u_{1:k}^l, \mathbf{w}_{1:k}^l, \mathcal{G}\right) f\left(\mathbf{C}_k|o_{0:k}, \mathbf{w}_{1:k}^l, \mathcal{G}\right). \quad (20)$$

The computation of each posterior density involves a prediction step and an update step. For the vehicle motion model in Section II-B2, the predicted density of vehicle trajectory at time step k is

$$f\left(o_{0:k}|\zeta_{1:k-1}, u_{1:k}^l, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right) = g_k\left(o_k|o_{k-1}, u_k^l, \mathcal{G}\right)$$
$$\times f\left(o_{0:k-1}|\zeta_{1:k-1}, u_{1:k-1}^l, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right). \quad (21)$$

The predicted density of the set of all B-spline trajectories at time step k is

$$f\left(\mathbf{C}_k|o_{0:k}, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right) =$$
$$\int g_k\left(\mathbf{C}_k|\mathbf{C}_{k-1}\right) f\left(\mathbf{C}_{k-1}|o_{0:k}, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right)\delta\mathbf{C}_{k-1}, \quad (22)$$

where $g_k(\mathbf{C}_k|\mathbf{C}_{k-1})$ is the transition density of the set of all trajectories for the multi-lane dynamic model in Section II-C2, and the set integral $\int f(\mathbf{X})\delta\mathbf{X}$ is defined in [6].

The predicted density of the set of all B-spline trajectories at time step k is then updated using the multi-lane measurement model $\ell_k(\mathbf{w}_k^l|\mathbf{C}_k, o_k, \mathcal{G})$ in Section II-C1, which gives

$$f\left(\mathbf{C}_k|o_{0:k}, \mathbf{w}_{1:k}^l, \mathcal{G}\right)$$
$$= \frac{f\left(\mathbf{C}_k|o_{0:k}, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right)\ell_k\left(\mathbf{w}_k^l|\mathbf{C}_k, o_k, \mathcal{G}\right)}{\ell_k\left(\mathbf{w}_k^l|o_{0:k}, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right)}, \quad (23)$$

where $\ell_k(\mathbf{w}_k^l|o_{0:k}, \mathbf{w}_{1:k-1}^l, \mathcal{G})$ is a normalizing factor. Lastly, the predicted density of the vehicle trajectory at time step k is updated using the measurement model in Section II-B1, which gives

$$f\left(o_{0:k}|\zeta_{1:k}, u_{1:k}^l, \mathbf{w}_{1:k}^l, \mathcal{G}\right) = f\left(o_{0:k}|\zeta_{1:k-1}, u_{1:k}^l, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right)$$
$$\times \ell_k\left(\zeta_k, \mathcal{G}|o_k\right)\ell_k\left(\mathbf{w}_k^l|o_{0:k}, \mathbf{w}_{1:k-1}^l, \mathcal{G}\right)$$
$$/\ell_k\left(\zeta_k, \mathbf{w}_k^l, \mathcal{G}|\zeta_{1:k-1}, \mathbf{w}_{1:k-1}^l\right), \quad (24)$$

where $\ell_k(\zeta_k, \mathbf{w}_k^l, \mathcal{G}|\zeta_{1:k-1}, \mathbf{w}_{1:k-1}^l)$ is a normalizing factor.

In the next section, we will describe how to perform these prediction and update steps in a computationally tractable way. For notational simplicity, we will drop the explicit dependence in the predicted and posterior densities and use the shorthand notations, with $k' \in \{k-1, k\}$,

$$f\left(o_{0:k}|\zeta_{1:k'}, u_{1:k}^l, \mathbf{w}_{1:k'}^l, \mathcal{G}\right) = f_{k|k'}\left(o_{0:k}\right), \quad (25a)$$
$$f\left(\mathbf{C}_k|o_{0:k}, \mathbf{w}_{1:k'}^l, \mathcal{G}\right) = f_{k|k'}\left(\mathbf{C}_k|o_{0:k}\right). \quad (25b)$$

IV. TPMBM-BASED SIMULTANEOUS LOCALIZATION AND MULTI-LANE TRACKING

The main challenge in computing (20) lies in the multi-lane update step (23), which involves the dependence on the vehicle trajectory $o_{0:k}$, and the normalizing factor in (23) needs to be explicitly computed for updating the vehicle trajectory in (24). To address this challenge, we adopt a particle representation of the vehicle trajectory [21], [22]:

$$f_{k|k'}\left(o_{0:k}\right) = \sum_{n=1}^{N_p} \omega_{k|k'}^n \delta_{o_{0:k}^n}\left(o_{0:k}\right), \quad (26)$$

where N_p is the number of particles and $\omega_{k|k'}^n$ is the weight of particle $o_{0:k}^n$, with $\omega_{k|k'}^n \geq 0$ and $\sum_{n=1}^{N_p} \omega_{k|k'}^n = 1$.

With this particle representation, the multi-lane prediction (22) and update step (23) boil down to computing the predicted and posterior density of the set $\mathbf{C}_k$ of all trajectories for each particle $o_{0:k}^n$. In particular, given the multi-lane measurement and dynamic models in Section II-C1 and II-C2, the density $f_{k|k'}(\mathbf{C}_k|o_{0:k}^n)$ is of the form TPMBM with [4],

$$f_{k|k'}(\mathbf{C}_k|o_{0:k}^n) = \sum_{\mathbf{X} \uplus \mathbf{Y} = \mathbf{C}_k} f_{k|k'}^p(\mathbf{X}|o_{0:k}^n) f_{k|k'}^{mbm}(\mathbf{Y}|o_{0:k}^n), \quad (27)$$

$$f_{k|k'}^p(\mathbf{C}_k|o_{0:k}^n) = e^{-\int \lambda_{k|k'}(C|o_{0:k}^n)dC} \prod_{C \in \mathbf{C}_k} \lambda_{k|k'}(C|o_{0:k}^n), \quad (28)$$

$$f_{k|k'}^{mbm}(\mathbf{C}_k|o_{0:k}^n) = \sum_{a \in \mathcal{A}_{k|k'}^n} \omega_{k|k'}^a \sum_{\substack{\uplus_{l=1}^{I_{k|k'}^n} \mathbf{C}^l = \mathbf{C}_k}} \prod_{i=1}^{I_{k|k'}^n} f_{k|k'}^{i,a^i}\left(\mathbf{C}^i|o_{0:k}^n\right). \quad (29)$$

The TPMBM in (27) is the union of two independent sets: a trajectory PPP with density (28) that represents undetected trajectories that are hypothesized to exist but have never been detected, and a trajectory multi-Bernoulli mixture with density (29) that represents trajectories that have been detected at some point up to time step k. The intensity of the trajectory PPP is $\lambda_{k|k'}(\cdot)$. Each received measurement generates a Bernoulli component, and the number of Bernoulli components is $I_{k|k'}^n$. A global hypothesis is $a = (a^1, \ldots, a^{I_{k|k'}^n}) \in \mathcal{A}_{k|k'}^n$, where $a^i \in \{1, \ldots, h_{k|k'}^i\}$ is the index to the local hypothesis for the i-th Bernoulli and $h_{k|k'}^i$ is its number of local hypotheses.

We refer to the j-th measurement $w_k^{l,j}$ using the pair (k, j), and the set of all such measurement pairs up to (and including) time step k is denoted by $\mathcal{M}_k$. Then, a local hypothesis a^i for the i-th Bernoulli component has a set of measurement pairs denoted as $\mathcal{M}_k^{i,a^i} \subseteq \mathcal{M}_k$. The set $\mathcal{A}_{k|k'}^n$ of global hypotheses satisfies [4],

$$\mathcal{A}_{k|k'}^n = \Big\{ \left(a^1, \ldots, a^{I_{k|k'}^n}\right) : a^i \in \left\{1, \ldots, h_{k|k'}^i\right\} \forall i,$$
$$\uplus_{i=1}^{I_{k|k'}^n} \mathcal{M}_{k'}^{i,a^i} = \mathcal{M}_{k'} \Big\}. \quad (30)$$

Global hypothesis a has weight $\omega_{k|k'}^a \propto \prod_{i=1}^{I_{k|k'}^n} \omega_{k|k'}^{i,a^i}$, where $\omega_{k|k'}^{i,a^i}$ is the weight of local hypothesis a^i for the i-th Bernoulli component, and it should satisfy that $\sum_{a \in \mathcal{A}_{k|k'}^n} \omega_{k|k'}^a = 1$. The density of the i-th Bernoulli component with local hypothesis a^i is $f_{k|k'}^{i,a^i}(\cdot|o_{0:k}^n)$, parameterized by a probability of existence $r_{k|k'}^{i,a^i}$ and a single trajectory density $p_{k|k'}^{i,a^i}(\cdot|o_{0:k}^n)$.

We proceed to describe the prediction (21), (22) and update (23), (24) steps with the particle representation of the vehicle

trajectory (26) and the TPMBM density representation of the set of all trajectories conditioned on each particle (27). Due to page limits, we refer the readers to [4], [7] for the explicit expressions of the TPMBM prediction and update equations.

A. Vehicle State Prediction

Given a trajectory particle $o_{0:k-1}^n$, we first draw a sample of the vehicle pose at time step k via $p_k^n \sim g_k(p_k|p_{k-1}^n, u_k^l)$, and then we draw a sample of the vehicle's position within the SD map at time step k via $q_k^n \sim g_k(q_k|p_k^n, q_{k-1}^n, \mathcal{G})$. Lastly, the sampled vehicle state o_k^n is appended to $o_{0:k-1}^n$ to obtain the updated particle $o_{0:k}^n$. Since we directly draw samples from the vehicle motion transition density, the particle weights remain unchanged.

B. Multi-Lane Prediction

Given a TPMBM posterior density at time step $k-1$ and the multi-lane dynamic model in Section II-C2 without spawning, the predicted density would also be a TPMBM [4]. However, due to the dependencies between each alive and its spawned trajectory, the predicted density is no longer a TPMBM in its closed form[3]. As in [7], we approximate the predicted density as a TPMBM by discarding these dependencies. In addition, we assume that the PPP represents alive trajectories without spawning, and that each trajectory Bernoulli component has a deterministic start time. These approximations would considerably simplify the implementation of the multi-lane prediction step. Note that the TPMBM filter in [7] considers the sets of tree trajectories, which contain the genealogy information of each trajectory. However, in multi-lane tracking the genealogy information is not needed, and we only consider the traditional sets of trajectories formulation for notational simplicity [6].

The prediction of PPP and existing Bernoulli components are the same as their predictions in the TPMBM filter without spawning [4, Prop. 1]. For a spawned Bernoulli component, its probability of existence is jointly determined by the spawning probability $P_2^S(\cdot)$, the probability of existence and the probability of being alive at time step k of its parent trajectory. After the prediction step, the number of Bernoulli components (27) becomes $I_{k|k-1}^n = 2I_{k-1|k-1}^n$.

C. Multi-Lane Update

Given a TPMBM predicted density at time step k and the multi-lane measurement model in Section II-C1, the posterior density at time step k is also a TPMBM (27) [4, Prop. 2]. The complete TPMBM update consists of five steps:

1) The update of the PPP for undetected trajectories that remain undetected.
2) The update of the PPP for undetected trajectories that are detected for the first time. In this step, each received measurement generates a unique Bernoulli component, such that $I_{k|k}^n = I_{k|k-1}^n + |\mathbf{w}_k^l|$.

---

[3]For the standard dynamic and measurement models with multi-Bernoulli spawning, the posterior is a TPMBM density on the set of *tree* trajectories without approximations [7]. This modeling is not used in this work.

3) The update of local hypotheses corresponding to misdetection of existing Bernoulli components.
4) The update of local hypotheses corresponding to measurement update of existing Bernoulli components.
5) The update of global hypotheses.

The main challenge lies in the update of global hypotheses. Each global hypothesis corresponds to a unique partition of $\mathcal{M}_k$, and the number of global hypotheses is given by the Bell number of $|\mathcal{M}_k|$ [23]. Thus, it is computationally intractable to enumerate all the global hypotheses. To address this challenge, we can only consider global hypotheses with high likelihoods, using approaches that combine clustering and 2D assignments [23] or sampling-based methods [24].

D. Vehicle State Update

Computing the particle weight $\omega_{k|k}^n$ requires the computation of $\ell_k(\mathbf{w}_k^l|o_{0:k}^n, \mathbf{w}_{1:k-1}^l, \mathcal{G})$, which is the normalizing factor in computing the TPMBM posterior density (23) at time step k. It can be observed that this normalizing factor is the same as the normalizing factor in computing the normalized global hypothesis weight [21], i.e.,

$$\ell_k(\mathbf{w}_k^l|o_{0:k}^n, \mathbf{w}_{1:k-1}^l, \mathcal{G}) = \sum_{a \in \mathcal{A}_{k|k}^n} \prod_{i=1}^{I_{k|k}^n} \omega_{k|k}^{i,a^i}. \quad (31)$$

The updated particle weight is then given by

$$\omega_{k|k}^n \propto \omega_{k|k-1}^n \ell_k(\zeta_k|o_k) \ell_k(\mathcal{G}|o_k) \sum_{a \in \mathcal{A}_{k|k}^n} \prod_{i=1}^{I_{k|k}^n} \omega_{k|k}^{i,a^i}. \quad (32)$$

To avoid particle degeneracy, we resample the particles if the number of effective particles is below $N_p/4$.

V. EXPERIMENTAL RESULTS

In this section, we present the experimental results of our proposed solution for simultaneous localization and multi-lane tracking using onboard sensors and a SD map on real-world data. The data was collected by a test vehicle during a single traversal on a highway in Gothenburg, Sweden.

The vision data was obtained using a forward-facing, mono camera, operating at 15 Hz. Both the visual odometry and the lane marking edge detections were extracted from the camera images using Zenseact's in-house developed computer vision algorithms. The GNSS measurements were obtained using a single-antenna GPS receiver, operating at 1 Hz; the SD map is from the open-source OpenStreetMap.

To evaluate the localization performance of our proposed solution, the vehicle pose estimates reported by a high-precision localization system from Oxford Technical Solutions (OxTS) are used as ground truth. Due to the lack of the ground truth HD map, we evaluate the multi-lane estimation performance by overlaying the estimated traffic lines on the Google Earth satellite image.

The parameter setting used in the SLAMOT filter is specified as follows. In the vehicle measurement model, the GNSS measurement noise covariance Ω_k is a diagonal matrix. The

standard deviation of the latitudinal and longitudinal errors are provided by the GNSS, and we set the standard deviation of the altitude error to $3\,\mathrm{m}$. The concentration parameter κ in the von Mises distribution is set to 8. In the vehicle motion model, the motion noise covariance Q_k is a diagonal matrix with the standard deviation of the position error set to $0.05\,\mathrm{m}$ and orientation error set to $0.1°$. The standard deviation σ_k of motion noise of d_k is set to $0.2\epsilon_k$, where ϵ_k is defined in (5). For B-spline trajectory state modeling, the time interval for placing the control points is $\tau^c = 5\tau = 1/3\,\mathrm{s}$.

We consider lane marking detection points that are within $16\,\mathrm{m}$ distance of the vehicle and set $l_x = 3\,\mathrm{m}$ and $l_y = 0.05\,\mathrm{m}$ in the covariance matrix (14) that represents the lane extent. In addition to the 3D positional information, each lane marking edge detection point also contains the instance segmentation information provided by the lane marking edge detector. In this work, we have utilized the lane marking type information, such as solid or dashed, and the clustering information, i.e., to which lane marking cluster it belongs. Each traffic line is detected with probability $P_k^D = 0.95$ except for the following special case. If a traffic line was associated to a dashed lane marking detection point at time step $k - 1$ and is not detected at time step k, then its detection probability will be lowered to $P_k^D = 0.5$ until it is detected again. This setting helps to connect multiple dashed lane markings into a single traffic line. In addition, by using the clustering information, the extended object data association problem reduces to the point object data association problem. Each individual detection point w_k also has its own $\Omega_k^{w_k}$ provided by the lane marking detector. The clutter detections have uniform Poisson intensity 10^{-3}.

In the multi-lane dynamic model, the survival probability is $P_1^S = 0.99$ and the spawning probability is $P_2^S = 10^{-5}$. The motion noise covariance in the state transition density (17) is $Q_k^c = 0.36\tau^c I_3$. The Poisson birth intensity (18) is a single Gaussian defined in the vehicle coordinate, with weight 10^{-5}, mean $[6, 0, -0.5, 11, 0, -0.5, 16, 0, -0.5]^T$, and covariance $\mathrm{diag}(100, 400, 4, 100, 400, 4, 100, 400, 4)$. The weight at time step 0 is set to 10^{-3}.

The TPMBM update uses Murty's algorithm [25] to find the $M = \lceil 500\omega_{k|k}^a \rceil$-best global hypotheses for global hypothesis a with weight $\omega_{k|k}^a$. The maximum number of global hypothesis is 500, and we prune global hypotheses with weight smaller than 10^{-3}. We also prune Bernoulli components with existence probability smaller than 10^{-5} and Gaussian components in the Poisson intensity with weight smaller than 10^{-5}. In addition, we consider L-scan implementation with $L = 3$ [26], where control points before the last three time instants are considered independent and remain unchanged with new measurements. Furthermore, Bernoulli components with probability of being alive at the current time step smaller than 10^{-5} are considered dead, and they are no longer predicted or updated.

To report the traffic line estimates, we first find the TPMBM posterior density conditioned on the particle with the highest weight, and then we find the multi-Bernoulli component with the highest weight and its Bernoulli components with existence probability one. For each Bernoulli component, we extract the sequence of B-spline control points with the most likely length. Finally, we interpolate the control points to obtain continuous traffic line estimates. The vehicle pose estimate at each time step is obtained by taking the sample mean instead of choosing the particle with the highest weight at the last time step due to particle history degeneracy.

In the experiment, we use $N_p = 1000$ particles, initialized only using information provided by the GNSS measurement. The multi-lane estimation performance is evaluated by overlaying the estimated traffic lines on the Google Earth satellite image; screenshots are provided in Fig. 1. It can be observed from the results that the estimated traffic lines are continuous and smooth, and that they align well with the lane markings in the satellite image up to a lateral offset. This offset mainly depends on the lateral localization accuracy of the vehicle as in most cases, the offset between the estimated traffic lines and the lane markings matches the offset between the OxTS and the estimated vehicle trajectories. Lane splittings are also well estimated, with the only exception being at the bottom-left of Fig. 1, where a newborn traffic line is estimated to be spawned from an existing one.

The root-mean-square localization error in the x, y, and z directions are $0.97\,\mathrm{m}$, $1.10\,\mathrm{m}$, and $2.06\,\mathrm{m}$, respectively. The improvement with respect to using only visual odometry and GNSS is about 11%. Since lane marking detections can only provide accurate relative lateral information and GNSS measurement is of low accuracy, it is difficult to accurately localize the vehicle and track the lanes without using detections that have good longitudinal information, such as traffic signs/lights. Nevertheless, our proposed solution can provide a reasonable estimate of traffic lines with respect to the vehicle trajectory.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an effective solution for simultaneous localization and multi-lane tracking using onboard sensors and a SD map. The proposed solution is based on a TPMBM-based SLAMOT filter, where the estimation of traffic lines, in the form of B-spline trajectories, using lane marking detection points is formulated as a multiple extended object tracking problem. The efficacy of the proposed solution has been demonstrated on real-world data.

In the future, we plan to work on the batch formulation of the SLAMOT problem such that smoothed vehicle trajectory and traffic line estimates could be obtained by exploiting full information of the data. The mapping of static landmarks, such as traffic signs and traffic lights, can also be incorporated into the solution to further improve the localization accuracy, and hence also the multi-lane estimation performance.

REFERENCES

[1] G. Elghazaly, R. Frank, S. Harvey, and S. Safko, "High-definition maps: Comprehensive survey, challenges and future perspectives," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[2] J. Zhou, Y. Guo, Y. Bian, Y. Huang, and B. Li, "Lane information extraction for high definition maps using crowdsourced data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7780–7790, 2023.
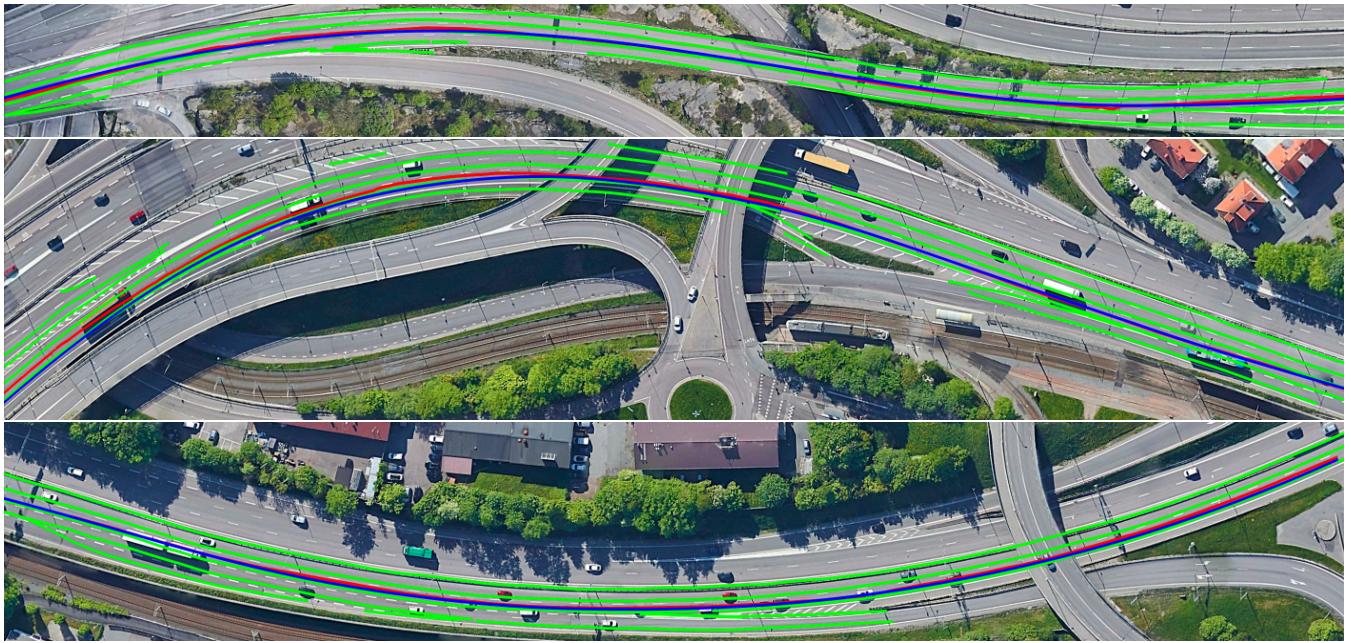
Fig. 1. OxTS vehicle trajectory (blue), estimated vehicle trajectory (red), estimated traffic lines (green) overlaid with Google Earth satellite image. The top figure illustrates the best traffic line estimation performance, followed by the bottom figure and the mid-figure.

[3] Y. Xia, K. Granström, L. Svensson, Á. F. García-Fernández, and J. L. Williams, "Extended target Poisson multi-Bernoulli mixture trackers based on sets of trajectories," in *22th International Conference on Information Fusion*. IEEE, 2019, pp. 1–8.

[4] Y. Xia, Á. F. García-Fernández, F. Meyer, J. L. Williams, K. Granström, and L. Svensson, "Trajectory PMB filters for extended object tracking using belief propagation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 6, pp. 9312–9331, 2023.

[5] K. Berntorp, M. Greiff, S. Di Cairano, and P. Miraldo, "Bayesian sensor fusion for joint vehicle localization and road mapping using onboard sensors," in *26th International Conference on Information Fusion (FUSION)*. IEEE, 2023, pp. 1–8.

[6] Á. F. García-Fernández, L. Svensson, and M. R. Morelande, "Multiple target tracking based on sets of trajectories," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 3, pp. 1685–1707, 2019.

[7] Á. F. García-Fernández and L. Svensson, "Tracking multiple spawning targets using Poisson multi-Bernoulli mixtures on sets of tree trajectories," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1987–1999, 2022.

[8] K. Li, "On embedding B-splines in recursive state estimation," *arXiv preprint arXiv:2307.00637*, 2023.

[9] A. Abramov, C. Bayer, C. Heller, and C. Loy, "Multi-lane perception using feature fusion based on GraphSLAM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 3108–3115.

[10] C. Kim, S. Cho, M. Sunwoo, and K. Jo, "Crowd-sourced mapping of new feature layer for high-definition map," *Sensors*, vol. 18, no. 12, p. 4172, 2018.

[11] M. Liebner, D. Jain, J. Schauseil, D. Pannen, and A. Hackelöer, "Crowd-sourced HD map patches based on road model inference and graph-based SLAM," in *IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 1211–1218.

[12] P. Cudrano, B. Gallazzi, M. Frosi, S. Mentasti, and M. Matteucci, "Clothoid-based lane-level high-definition maps: Unifying sensing and control models," *IEEE Vehicular Technology Magazine*, vol. 17, no. 4, pp. 47–56, 2022.

[13] C. Lundquist, U. Orguner, and F. Gustafsson, "Extended target tracking using polynomials with applications to road-map estimation," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 15–26, 2010.

[14] B. Akbari and H. Zhu, "Tracking dependent extended targets using multi-output spatiotemporal Gaussian processes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 18 301–18 314, 2022.

[15] K. Zhao, M. Meuter, C. Nunn, D. Müller, S. Müller-Schneiders, and J. Pauli, "A novel multi-lane detection and tracking system," in *IEEE Intelligent Vehicles Symposium*. IEEE, 2012, pp. 1084–1089.

[16] Z. Qiao, Z. Yu, H. Yin, and S. Shen, "Online monocular lane mapping using Catmull-Rom spline," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 7179–7186.

[17] M. A. Brubaker, A. Geiger, and R. Urtasun, "Map-based probabilistic visual self-localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 652–665, 2015.

[18] P. Merriaux, Y. Dupuis, P. Vasseur, and X. Savatier, "Fast and robust vehicle positioning on graph-based representation of drivable maps," in *IEEE International Conference on Robotics and Automation*. IEEE, 2015, pp. 2787–2793.

[19] S. Yang and M. Baum, "Tracking the orientation and axes lengths of an elliptical extended object," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4720–4729, 2019.

[20] D. Spinello and D. J. Stilwell, "Nonlinear estimation with state-dependent Gaussian observation noise," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1358–1366, 2010.

[21] H. Kim, K. Granström, L. Svensson, S. Kim, and H. Wymeersch, "PMBM-based SLAM filters in 5G mmwave vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 8646–8661, 2022.

[22] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit *et al.*, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," *Aaai/iaai*, vol. 593598, 2002.

[23] K. Granström, M. Fatemi, and L. Svensson, "Poisson multi-Bernoulli mixture conjugate prior for multiple extended target filtering," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 1, pp. 208–225, 2020.

[24] K. Granström, L. Svensson, S. Reuter, Y. Xia, and M. Fatemi, "Likelihood-based data association for extended object tracking using sampling methods," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 30–45, 2018.

[25] D. F. Crouse, "On implementing 2D rectangular assignment algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1679–1696, 2016.

[26] Á. F. García-Fernández and L. Svensson, "Trajectory PHD and CPHD filters," *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5702–5714, 2019.