WILEY | Hindawi

Research Article

An Unscented Kalman Filter-Based Method for Reconstructing Vehicle Trajectories at Signalized Intersections

Jiantao Mu ⓘ**, Yin Han, Cheng Zhang** ⓘ**, Jiao Yao** ⓘ**, and Jing Zhao** ⓘ

Department of Traffic Engineering, University of Shanghai for Science and Technology, 516 Jungong Road, Shanghai 200093, China

Correspondence should be addressed to Jing Zhao; jing_zhao_traffic@163.com

On-board data of detected vehicles play a critical role in the management of urban road traffic operation and the estimation of traffic status. Unfortunately, due to limitations of technology and privacy issues, the sampling frequency of the detected vehicle data is low and the coverage is also limited. Continuous vehicle trajectories cannot be obtained. To overcome the above problems, this paper proposes an unscented Kalman filter (UKF)-based method to reconstruct the trajectories at signalized intersections using sparse probe data of vehicles. We first divide the intersection into multiple road sections and use a quadratic programming problem to estimate the travel time of each section. The weight of each initial possible trajectory is calculated separately, and the trajectory is updated using the unscented Kalman filter (UKF); then, the trajectory between two updates is also obtained accordingly. Finally, the method is applied to the actual scenario provided by the NGSIM data and compared with the real trajectory. The mean absolute error (MAE) is adopted to evaluate the accuracy of the proposed trajectory reconstruction. Sensitivity analysis is provided in order to provide the requirement of sampling frequency to obtain highly accurate reconstructed vehicle trajectories under this method. The results demonstrate the applicability of the technique to the signalized intersection. Therefore, the method enables us to obtain richer and more accurate trajectory data information, providing a strong prior basis for future urban road traffic management and scholars using trajectory data for various studies.

1. Introduction

Accurate and reliable vehicle trajectory data are significant in the intelligent transportation system and urban traffic management [1]. Trajectory data provide a rich source of information for many application areas, such as getting on-site speed, queue, delay, acceleration, and driving time [2]. The extracted trajectory can not only reflect the traffic flow characteristics of a specific road section at the micro level but also explain the traffic demand and temporal and spatial distribution characteristics of the entire transportation network at the macro level [3]. It can also be used to measure arterial performance, estimated queue length [4], travel time [5, 6], and vehicle energy and emissions [7]. Therefore, vehicle trajectory data have high research value, and their importance to urban traffic operation and management is self-evident.

So far, traffic data can be divided into two major categories: (i) traditional fixed-location sensor data, e.g., ring detector, video, automatic vehicle identification (AVI), and automatic license plate recognition (ALPR) [8], and (ii) mobile sensor data, e.g., global positioning system (GPS), floating car, and smartphone [9]. Fixed-location sensors can provide information such as driving time, average speed, and traffic volume within 30 seconds to 5 minutes at the installation location [10]. However, the collected data cover a limited area due to the funding constraints and the low coverage of sensors. With the development of technologies such as wireless communication, mobile positioning, and vehicle-road collaboration, massive traffic data are constantly being generated. Vehicle trajectory data have a wide range of applications in transportation due to their rich temporal and spatial information [11, 12]. Unfortunately, due to the limitation of the existing technology, it is still

difficult to directly obtain complete and high sampling rate vehicle trajectory data within a specific space-time range [13]. The raw probe data with incomplete trajectories usually cannot provide the actual traffic conditions. Even shortly, we can successfully implement to obtain high-frequency data, but it will lead to privacy and security problems [14, 15]. At the same time, the data usually contain a large number of measurement errors and outliers [16]. To more accurately evaluate traffic conditions and solve traffic problems, we focus on trajectory reconstruction between two sparse updates.

So far, trajectory reconstruction has received a lot of attention from all over the world, and many scholars have done in-depth research on vehicle trajectory reconstruction [17–19]. Coifman [20] proposed an estimation method based on traffic flow theory (TFT) to reconstruct the vehicle trajectory on the expressway section, using the traffic data from a single dual loop detector which is extrapolated to the extended link. Van Lint and Hoogendoom [21] integrated the upstream and downstream data and used a multiloop sensor to reconstruct the vehicle trajectory on a route. However, this method is not suitable for urban road trajectory reconstruction. Punzo et al. [22] proposed a method to inspect trajectory data accuracy, based on jerks' analysis, consistency analysis, and spectral analysis, which is successively applied to datasets. Based on this research, Montanino and Punzo [23] presented a multistep filtering method, which aimed to eliminate outliers leading to nonphysical acceleration values by local reconstruction of the vehicle trajectory, as well as cutting off residual random disturbances from the signal while maintaining driving dynamics.

Meanwhile, compared with the methods of vehicle trajectory reconstruction on highways, more and more researchers focus on signal-controlled urban roads. For example, Mehran et al. [24] combined fixed-location sensor data with signal timing information and used variational theory (VT) to solve the problem of vehicle trajectory reconstruction. Sun and Ban [25] proposed the optimization-based method and the delay-based method to estimate the shockwave boundaries and then used the variation formulation (VF) method to reconstruct vehicle trajectories. This method is macroscopic. Hao et al. [26] focused on finding the most likely driving mode sequences (deceleration, idling, acceleration, and cruising) to forecast the vehicle's trajectory. But, different from the approach of Sun and Ban [25], this is a microscopic-based approach. Chen et al. [10] developed a hybrid method combining Kalman filtering (KF) and variational theory (VT) to reconstruct the vehicle trajectory of signalized intersection. The VT was used to establish an extended variational network, and the KF was used to characterize the dynamic and random characteristics of queue boundary curves.

Because travel time is a crucial traffic index, many studies have focused on estimating travel time to reconstruct the vehicle trajectory. Shan et al. [27] proposed a probabilistic model that used sparse data to reconstruct vehicle trajectories. Wan et al. [28] divided the intersection into many short sections. The section travel time was estimated by the expectation maximization (EM) method, presenting situations when spanning single or multiple intersections. Then, they employed the maximum likelihood estimation (MLE) approach to generate the most likely trajectory between two consecutive sampling times. The effectiveness of the method has been proved by using transit buses data of San Francisco. Hao et al. [29] proposed a stochastic model to estimate the second-by-second vehicle velocity trajectory by examining all possible sequences of modal activity between sparse locations and continuous data points of velocity measurements. The vehicle dynamic probability is then expressed as the product of multiple independent event probabilities. That is, the detailed vehicle velocity trajectory is reconstructed from the modal activity sequences with maximum likelihood. Fard et al. [30] proposed a two-step technique based on wavelet analysis, primarily by using a wavelet transform with unique processing methods to identify and modify outliers. Next, the noise in the trajectory data is eliminated by applying a wavelet-based filter, which is used in order to filter errors and reconstruct the trajectory data. Venthuruthiyil and Chunchu [31] improved the locally weighted polynomial regression (LWPR) by proposing a method for trajectory reconstruction by determining the optimal window size and polynomial order and taking into account the dynamics of individual vehicles. Wang et al. [32] proposed a segmented deceleration and acceleration model developed to calculate acceleration for different driving modes in a trajectory. A new expectation maximization (EM) algorithm was then used to estimate each driving mode's acceleration data distribution parameters. The acceleration statistics are then used to reconstruct the corresponding part of the trajectory. Similar to Wan et al. [28], Wei et al. [33] adopted a particle filter (PF)-based approach to reconstruct the missing trajectory between the continuous updates of vehicles on the main road of single and multiple intersections.

Because there are outliers and errors in the traffic data, there will be deviations in the actual results. In the specific research methods, most scholars use filtering algorithms to solve the trouble. Common filtering techniques are the moving average method, the local polynomial method, wavelet analysis, the Kalman filter (KF), etc. It is noteworthy that many studies have applied the Kalman filter (KF) to reconstruct the vehicle trajectory. The KF can estimate the state of the dynamic system from a series of data with measurement noise when the measurement variance is known. In a dynamic system with uncertain information, the KF can make informed speculation about what the system will do next. Even if there is noise information interference, we can usually find out what happened and find the invisible correlation between images. These have demonstrated the superiority of the KF. However, the KF is an equation of state that transforms the problem into a linear Gaussian model, which has many limitations in the processing of trajectory. Subsequently, some studies used the particle filter (PF) to reconstruct the vehicle trajectory at signalized intersections. The PF enables vehicle trajectory reconstruction by updating the state-space equation based on the sampling of trajectory candidates and selects the particle with the highest weight by

displaying the weight difference of each candidate, which can be directly used to solve the problem of sparse trajectory reconstruction. Xie et al. [34] had done pioneering research in this field, and Wei et al. [33] solved some problems on this basis and obtained more accurate results. These prove that the particle filter is an excellent method to solve vehicle trajectory reconstruction. Although the PF can be used as an effective means to solve nonlinear problems, it needs a large number of sample points, resulting in a very complex calculation. Otherwise, it cannot get a good a posteriori probability density. These questions are worth our attention and study. Motivated by the above issues, this study presents an unscented Kalman filter (UKF)-based method to reconstruct the vehicle trajectory for the signalized intersection. The UKF is also a good method to solve nonlinear problems, which has high calculation accuracy. The generated sigma particles are selective instead of a large number of points, and the weighted mean and variance of particles converge to the real mean and variance [35, 36]. There are few studies to date that have applied the UKF to the signalized arterial in order to reconstruct the vehicle trajectory. The main contributions of this study are briefed as follows:

(1) By resorting to the advantage of UKF, this study uses sparse probe data to reconstruct the vehicle trajectory of a signalized intersection. It is beneficial to solve the practical problems encountered in complex traffic operations.

(2) This study optimizes the previous models and uses the constrained quadratic programming problem to reduce the error of time allocation. It helps make the initial trajectories more accurate and contributes to the computational efficiency and accuracy of trajectory reconstruction.

(3) In this study, a new method is proposed to test the applicability of the method by changing the proportion of known points to all points instead of changing the sampling frequency (the points are selected uniformly) to obtain the best applicability of the method within a specific range and thus achieve the optimal accuracy of reconstructed trajectories. It has important guiding significance in practical applications.

Specifically, referring to the structure and vehicle operation characteristics of the signalized intersection, with time as the reference quantity, previous methods are optimized. The trajectory between two consecutive update points was reconstructed based on the UKF method. In addition, through preprocessing the NGSIM data [37], the method is operated in Lankershim Boulevard, Los Angeles. The results show that the reconstructed trajectory has higher accuracy compared with the actual situation. Finally, the applicability of the UKF-based method is investigated through sensitivity analysis.

In the rest of this paper, firstly, the framework and a brief description of the theory of the article are explained. Next, the estimation of segmented travel time statistics is explained. After considering the various operation modes of

vehicles at signalized intersections, a method for UKF-based trajectory reconstruction is proposed. Furthermore, the proposed method is applied and demonstrated by NGSIM data of Lankershim Boulevard Street in the US, compared with the actual trajectory on the ground. The sensitivity analysis is carried out to find the range of known point proportions for which the method has optimal applicability. Finally, the conclusions and future research directions are discussed.

It is worth noting that, in this study, we ignore the length of vehicles and the behaviors such as lane changes caused by vehicle-to-vehicle interactions during the travel time. We consider the multiple trajectories obtained through the model as initial sigma points. Each trajectory represents an initial point labeled with a number, and their weights are calculated separately.

The rest of this paper is organized as follows: Section 2 describes the framework of the method, introduces the unscented Kalman filter (UKF), and explains the method of segmental travel time estimation and the method for reconstructing the most likely trajectory of a vehicle including two cases where vehicles have one-stop and no stop at an intersection. Section 3 uses the NGSIM data for case analysis to obtain the results for comparison with the actual. Section 4 presents a sensitivity analysis to seek the optimal sampling frequency under this method. And the conclusions are presented in Section 5.

2. Methodology

2.1. Framework of the Proposed Method. This research focuses on vehicle trajectory reconstruction at signalized intersections. The flowchart of this method is shown in Figure 1. The method consists of three parts. It can be divided into two cases: no stop and stop, and estimate the travel time to get the initial trajectories which can be seen as initial sigma points. Then, it can calculate the proportion of the multiple potential trajectories. The one with the highest weight can reconstruct the vehicle trajectory. Finally, perform sensitivity analysis so that we should reduce the time interval between two updates, reconstruct the trajectory in segments, and then combine the trajectories.

2.2. Unscented Kalman Filter. The UKF is the combination of unscented transform (UT) and standard Kalman filter system. The nonlinear system equation is suitable for the standard Kalman system under the linear assumption through unscented transformation. This technique mainly uses n points collected in the prior distribution, which we call sigma points, uses a group of sigma points to describe the Gaussian distribution of random variables, and then uses the weighted statistical linear regression technique to approximate the posterior mean and variance of nonlinear functions through the transmission of nonlinear functions.

The UKF is mainly divided into prediction and update. The calculation accuracy of nonlinear distribution statistics is at a higher level. Three significant steps are involved, i.e.,
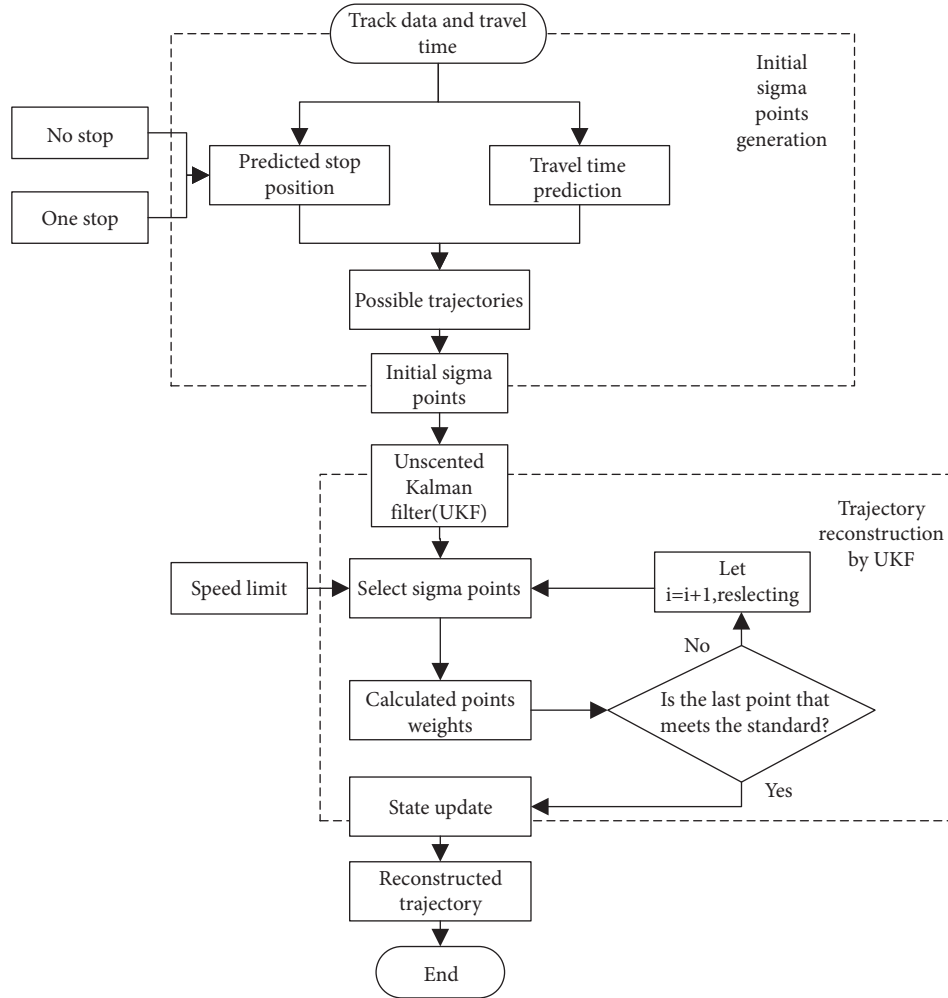
Figure 1: Flowchart of the method of vehicle trajectory reconstruction.

unscented transform, generating sigma points, and state prediction and update.

2.2.1. Unscented Transform.

$y = f(x)$ is a nonlinear function, $\overline{x}$ is an n-dimensional random variable, its mean is x, and its variance is P_x. UT obtains a group of sigma sampling points $\{\chi_i\}$, $i = 0, 1, \ldots, L$, according to a certain sampling strategy and sets the mean weight W_i^m and variance weight W_i^c to approximate the a posteriori mean and variance of the nonlinear function. Using the selected sigma sampling point set $\{\chi_i\}$, $i = 0, 1, \ldots, L$, $y_i = f(\chi_i)$ can be obtained by nonlinear function transfer, where y_i is the corresponding point of sigma sampling after nonlinear function transfer. According to the weighted statistical linear regression technique, the statistical characteristics of y can be approximated:

$$
\begin{aligned}
\overline{y} &= \sum_{i=0}^{L} W_i^m y_i, \\
P_{yy} &= \sum_{i=0}^{L} W_i^c (y_i - \overline{y})(y_i - \overline{y})^T, \\
P_{xy} &= \sum_{i=0}^{L} W_i^C (\chi_i - \overline{x})(y_i - \overline{y})^T.
\end{aligned}
\tag{1}
$$

2.2.2. Generating Sigma Points: The Selection of Sampling Points Shall Meet.

$$
g\left[\{\chi_i, W_i^m, W_i^c\}, L, P_x(x)\right] = 0, \tag{2}
$$

where $\{\chi_i, W_i^m, W_i^c\}$ is the set of sigma sampling points, mean weight, and variance weight, L is the number of

sampling points, $P_x(x)$ is the density function of random variable x, and $g[.]$ determines the relevant information of x.

Assuming $L = 2n$, the number of sigma sampling points is $2n + 1$. We can see that

$$
\begin{cases}
\chi_0 = \overline{x}, \\
\chi_i = \overline{x} + \left(\sqrt{(n+\lambda)P_x}\right)_i, \quad i = 1, 2, \ldots, n, \\
\chi_{i+n} = \overline{x} - \left(\sqrt{(n+\lambda)p_x}\right)_i.
\end{cases}
\tag{3}
$$

The corresponding weight

$$
W_i^m =
\begin{cases}
\dfrac{\lambda}{(n+\lambda)}, & i = 0, \\[3mm]
\dfrac{1}{2(n+\lambda)}, & i \neq 0,
\end{cases}
\tag{4}
$$

$$
W_i^C =
\begin{cases}
\dfrac{\lambda}{(n+\lambda)} + 1 + \beta - \alpha^2, & i = 0, \\[3mm]
\dfrac{1}{2(n+\lambda)}, & i \neq 0.
\end{cases}
$$

It is optimal when $\beta = 2$ and λ and α are scaling factors, where $\alpha \in [0, 1]$.

Therefore, the formula for calculating the sigma point weights is

$$
W_i =
\begin{cases}
\dfrac{\lambda}{(n+\lambda)}, & i = 0, \\[3mm]
\dfrac{1}{2(n+\lambda)}, & i \neq 0.
\end{cases}
\tag{5}
$$

The sum of all sigma point weights is equal to 1.

2.2.3. State Prediction: Initial State Estimation and Estimated Variance

$$
\begin{cases}
\widehat{x}_{0|0} = E(x_0), \\
P_{0|0} = E\left(x_0 - \widehat{x}_{0|0}\right)\left(x_0 - \widehat{x}_{0|0}\right)^T.
\end{cases}
\tag{6}
$$

Time update: for the state estimation $\widehat{x}_{k|k}$ and the estimation variance $P_{k|k}$ at time k, $2n + 1$ sigma sampling points χ_i and corresponding weights W_i^m and W_i^c can be obtained.

Nonlinear state transfer of sampling points:

$$
\gamma_{k+1|k}^i = f(\chi_i).
\tag{7}
$$

Updated mean and variance:

$$
\begin{cases}
\widehat{x}_{k+1|k} = \displaystyle\sum_{i=0}^{2n} W_i^m \gamma_{k+1|k}^i, \\[3mm]
P_{k+1|k} = \displaystyle\sum_{i=0}^{2n} W_i^c \left(\gamma_{k+1|k}^i - \widehat{x}_{k+1|k}\right)\left(\gamma_{k+1|k}^i - \widehat{x}_{k+1|k}\right)^T + Q_k,
\end{cases}
\tag{8}
$$

where $\widehat{x}_{k+1|k}$ is the state valuation, $P_{k+1|k}$ is the variance of the state prediction, and Q_k means the covariance.

The particles are obtained by a given sampling method, and the weight of the particles updates the measured value and state of the particles. The variance also changes. When the variance distribution is minimized, it can represent the real state distribution.

2.3. Estimation of the Road Section Travel Time.

To reconstruct the vehicle trajectory of signalized intersection based on the UKF, it is necessary to generate appropriate initial particles. The potential trajectory of the detection vehicle between two successive updates can be defined as the initial particle. Based on the method proposed by Wan et al., the front and back of the intersection are divided into multiple segments to estimate the travel time of each road section. Based on this method, the initial sigma particles with the actual situation can be obtained. As shown in Figure 2, the whole intersection was divided into several road units with a certain length.

The roads in the front and rear of the intersection are evenly divided into sections, represented by $\{C_1, C_2, \ldots, C_n\}$. The number of passing vehicles is n, denoted by $\{1, 2, \ldots, n\}$. The sum of the travel time allocated to each section is equal to the total observed time:

$$
t_{\{C_1, C_n\}}^j = \sum_{i=1}^{N} t_{C_i}^j.
\tag{9}
$$

After allocating time to each unit, it is redistributed through the maximum likelihood function, and the expectation and variance of each unit can be obtained.

Distribute the initial time evenly:

$$
t_{C_i}^j = \frac{1}{C_N - C_1 + 1}\left(t_{\{C_1, C_N\}}^j\right).
\tag{10}
$$

Assume that time is normally distributed:

$$
P\left(t_{C_i}^j\right) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}} e^{-\left(\left(t_{C_i}^j - \mu_{C_i}\right)^2 / 2\sigma_{C_i}^2\right)}.
\tag{11}
$$

Write it as a log-likelihood function:

$$
\log\left[P\left(t_{C_i}^j | \mu_{C_i}, \sigma_{C_i}\right)\right] = -\frac{\left(t_{C_i}^j - \mu_{C_i}\right)^2}{2\sigma_{C_i}^2} - \log\left(\sqrt{2\pi}\,\sigma_{C_i}\right).
\tag{12}
$$
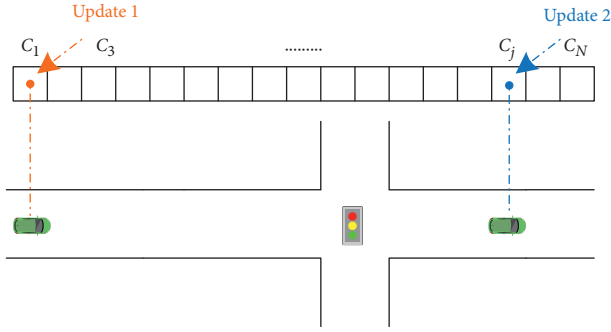
FIGURE 2: Assignment of the road.

Analyze all sections and redistribute time:

$$\log\left[P\left(t_X^j|\mu_X, \sum X\right)\right] = \sum_{i=1}^{N} \log\left[P\left(t_{C_i}^j|\mu_{C_i}, \sigma_{C_i}\right)\right], \quad (13)$$

where $X = [C_1, C_2, \ldots, C_N]^T$, $\mu_X = [\mu_{C_1}, \mu_{C_2}, \ldots, \mu_{C_N}]^T$,

and $\sum X = \begin{bmatrix} \sigma_{C_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{C_N}^2 \end{bmatrix}$.

Solving the maximum likelihood function, transform it into a quadratic programming problem:

$$\arg_{t_{C_i}^j} \min y = \frac{1}{2}x^T H x + f^T x$$

$$\text{s.t.} \sum_{i=1}^{N} t_{C_i}^j = t_{\{C_1, C_N\}}^j, \quad (14)$$

where $x = [t_{C_1}^j, t_{C_2}^j, \ldots, t_{C_N}^j]^T$, $H = \begin{pmatrix} 1/\sigma_{C_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_{C_N}^2 \end{pmatrix}$,

and $f = [\mu_{C_1}/\sigma_{C_1}^2, \ldots, \mu_{C_N}/\sigma_{C_N}^2]^T$.

Make $j = j + 1$ repeat the above steps for multiple iterations until the difference between the two iterations is less than the threshold and stop the iteration.

2.4. Trajectory Estimation of the Signalized Intersection.
In urban road networks, due to signal control, each vehicle's arrival time and queue length are different. There are some uncertain factors near the intersection, so we should find the difference instead of simply averaging them. We generally divide the trajectory at the intersection into two categories, i.e., no stop and with stop (we only consider the case of one-stop). For vehicles with a stop, there is a possibility of multiple trajectories between two updates, which is affected by the difference in parking locations, so we have to keep updating and predicting possible parking locations. These are shown in Figure 3.

If there is a queue of vehicles or if the signal is a red phase, vehicles need to stop at the signalized intersection to queue. Shockwave theory has been widely used to describe queuing dynamics, and we can also obtain various traffic parameters [38]. As shown in Figure 3(b), A is the start of the red light. When the red light is on, the vehicles start to queue at the

intersection. B is the start of the green light. When the green light is on, the vehicles begin to start through the intersection. According to the dataset, we can estimate the farthest queue point C. The slope of AC is the queuing rate, and the slope of BC is the discharge rate. Then, we use the actual data to analyze all the possible stopping positions and the corresponding waiting times between A and C. We can get different trajectories, respectively. The specific research methods are as follows.

2.4.1. Estimation of Trajectories with No Stop

Step 1: estimating travel time and comparing with actual time.

For nonstop trajectories, the estimated travel time can be calculated by

$$\widehat{T}(C_N) = T(C_1) + \sum_{i=1}^{N} \mu_{C_i}, \quad (15)$$

where μ_{C_i} is the average time of road section, which can be obtained in Section 2.3, and C_1 and C_N represent the first and last sections which we divided. We define the difference between the actual and the expected travel time as

$$\Delta t_{\{C_1, C_N\}} = T(C_N) - \widehat{T}(C_N). \quad (16)$$

Step 2: allocating the error into each section.

We can get

$$\Delta t_{\{C_1, C_N\}} = \sum_{i=1}^{N} \Delta t_{C_i}. \quad (17)$$

Write the error distribution as a normal distribution:

$$P\left(\Delta t_{c_i}|0, \sigma_{C_i}\right) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}} \cdot e^{-\left(\left(\Delta t_{C_i}\right)^2/2\sigma_{C_i}^2\right)}. \quad (18)$$

Establish a quadratic programming problem:

$$\arg_{\Delta t_{C_i}} \min y = \frac{1}{2}x^T H x + f^T x$$

$$\text{s.t.} \sum_{i=1}^{N} \Delta t_{C_i} = \Delta t_{\{C_1, C_N\}}, \quad (19)$$

where $x = [\Delta t_{C_1}, \Delta t_{C_2}, \ldots, \Delta t_{C_N}]^T$, $H = \begin{bmatrix} 1/\sigma_{C_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_{C_N}^2 \end{bmatrix}$,

and $f = [0, 0, \ldots, 0]^T$.

An optimal time error allocation can be obtained. Then, the real time is

$$\widehat{t_{C_i}^j} = t_{C_i}^j + \widehat{\Delta t_{C_i}^j}. \quad (20)$$

2.4.2. Estimation of Trajectories with One Stop

Step 1: assuming the stop position.

Assuming that the parking position is C_z, the stop time is $\text{time}(C_z)$, and the waiting time is tw_{C_z}.
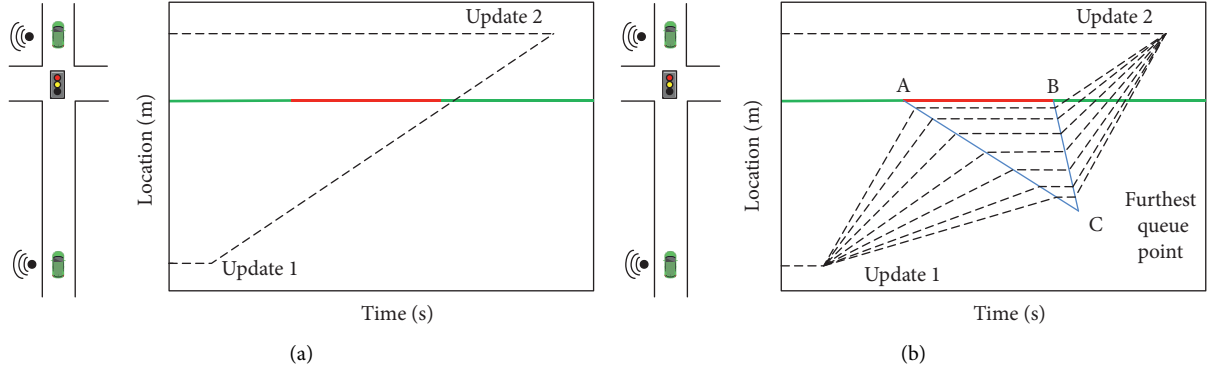
FIGURE 3: Trajectories of two cases at the signalized intersection. (a) Nonstop trajectory. (b) Trajectories with one stop.

Step 2: calculating the error of expected and real time.

We divide the whole process into preparking and postparking:

$$\Delta t^{j}_{d\{C_1,C_{z-1}\}} = \text{time}(C_z) - \left[\text{time}(C_1) + \sum_{i=1}^{z-1} t_{C_i}\right],$$

$$\Delta t^{j}_{d\{C_{z+1},C_N\}} = \text{time}(C_N) - \left[\text{time}(C_z) + tw_{C_z} + \sum_{i=z+1}^{N} t_{C_i}\right].$$

(21)

Step 3: allocating the error into each section.

Let us still write the question as a quadratic programming problem:

$$\arg_{\Delta t_{d,C_i}} \min y = \frac{1}{2}x^T H x + f^T x$$

$$\text{s.t.} \sum_{i=1}^{z-1} \Delta t_{C_i} = \Delta t_{d\{C_1,C_{z-1}\}},$$

(22)

$$\sum_{i=z+1}^{N} \Delta t_{C_i} = \Delta t_{d\{C_{z+1},C_N\}},$$

where $\quad x = [\Delta t_{d,C_1}, \Delta t_{d,C_2}, \ldots, \Delta t_{d,C_N}]^T,$

$$H = \begin{bmatrix} 1/\sigma^2_{C_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma^2_{C_N} \end{bmatrix}, \text{ and } f = [0, 0, \ldots, 0]^T.$$

The optimal error distribution is obtained:

$$\widehat{t^{j}_{C_i}} = t^{j}_{C_i} + \widehat{\Delta t^{j}_{d,C_i}}.$$

(23)

We can find different parking positions between the farthest point and the nearest point.

Then, the vehicle trajectory reconstruction with one-stop can be divided into three parts: the first part is the trajectory from the first update to the predicted stopping position, the second part is the trajectory between the predicted stopping position and the predicted starting position after queuing (assuming it is stationary and the trajectory is a straight line), the third part is the trajectory between the starting position and the second update, and

then the combination of these three parts is the entire reconstructed trajectory.

3. Case Analysis

3.1. Description of the Dataset. In this paper, we use NGSIM data which are public. The data contain global time, longitude and latitude, frame, global location, and several other attributes of vehicles. The data come from the Next Generation Simulation project, which was started by the Federal Highway Administration (FHWA) of transportation in 2002. The data of NGSIM include the driving conditions of all vehicles on US-101, I-80, and other roads in a period. They are obtained by cameras and then processed into track point records one by one [37].

To verify the applicability of our proposed method, we select an intersection of Lankershim Boulevard in Los Angeles, California, for example analysis. NGSIM data were collected on June 16, 2005. The study area was approximately 500 meters (1600 feet) in length, the entire datasets from 8:30 a.m. to 8:45 a.m. were included, and the trajectory data were updated every 0.1 seconds.

3.2. Trajectory Reconstruction at Intersection and Comparison with the Truth Data. After preprocessing the NGSIM data and obtaining some parameters we needed, the trajectories of 50 vehicles were selected randomly as the reference data. Next, we select the trajectory points at the same time interval and create a new sparse trajectory dataset for follow-up research.

The research area of this paper is a single intersection of Lankershim Boulevard Street. We select two roads before and after the signal light with a total length of 120 meters. The whole road is divided into 60 sections, and each cell is 2 meters long. In this paper, we set the average speed of the selected points to exceed the threshold of 10.5 m/s. Based on these data, we can use the method mentioned in Section 2 to estimate the average μ_{C_i} and variance $\sigma^2_{C_i}$ of the travel time for each road section C_i. Based on this information, we can reconstruct the vehicle trajectory between two updates.

Firstly, we take the vehicle with ID 2 as an example. We select the points of its two consecutive updates, which fell into road sections C_1 and C_{60}, respectively. The time interval

between the two updates is 11 s. We will construct potential trajectories between two update points that are no stop and one-stop, respectively, according to the method that we proposed. We can get 24 possible trajectories between the two updates through this method, which means we get 24 initial particles. Then, according to the method in the UKF, we calculate the weight of the initial sigma points that we filtered and select the particle with the highest weight. Figure 4(a) shows the weight of each particle. We can see that the weight of particle 1 is 96%, so the most likely track between the two updates is the first trajectory which is a track without stopping, and then the comparison with the actual trajectory is shown in Figure 4(b).

To analyze another case, we take the vehicle with ID 5, and the process is similar to the first. However, the time interval between the two updates became 53 s. The stopped and nonstopped trajectories are reconstructed according to the above method, and 28 initial sigma points can be obtained. After calculating the weight of the points, the difference from the first case, we get multiple particles with high weight proportion, as shown in Figure 5(a). We choose the trajectory of particle 6, which has the highest weight, as the most likely trajectory to reconstruct the trajectory of the vehicle with ID 5. The comparison with the real trajectory is shown in Figure 5(b).

The results prove that the trajectory reconstruction method based on the UKF proposed in this paper has good accuracy.

3.3. Comparison of Methods. To further prove the superiority of the proposed method, we choose the method based on a new expectation maximization (EM) algorithm for comparison in this paper. The method proposed a new approach to reconstruct vehicle trajectories using low sampling frequency data, which does not require additional information. Segmental deceleration and acceleration models were developed for calculating the acceleration in the trajectory for different driving modes. The acceleration data distribution parameters for each driving mode were then estimated using a new expectation maximization (EM) algorithm. The acceleration statistics are then used to reconstruct the corresponding part of the trajectory. Vehicle IDs 2 and 5 are still selected as the study objects, and the comparison results are shown in Figure 6.

It can be intuitively seen from Figure 6 that the method proposed in this study has higher accuracy and relatively high agreement with the actual trajectory comparison. In order to make the results more convincing, we quantitatively analyze the results and calculate the mean absolute error (MAE) between the two methods and the actual ground trajectory. The definition of MAE is

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \widehat{L}_i - L_i \right|, \tag{24}$$

where $\widehat{L}_i$ represents the estimated position and L_i is the actual position at time i.

For the accuracy of the results, two more vehicles were randomly selected from the 50 previously selected vehicles, and the results are shown in Table 1.

We can clearly see the proposed method in this paper has better applicability. Moreover, the MAE of vehicles 2, 5, 8, and 19 is 2.91 m, 3.81 m, 3.32 m, and 4.67 m. Compared with the EM method, the accuracy has a range of ten to forty percent improvement, and the trajectory reconstruction in both stop and nonstop cases achieves good accuracy, which also proves the applicability of the method under different driving conditions.

4. Sensitivity Analysis

According to the conclusion in Section 3.2, we can know that the time interval between two successive updates has an impact on the weight of the initial sigma points. When calculating the particle weight, there will be multiple initial points with weight greater than 0, and some of the points have nearly the same weight, which can have an effect on the trajectory reconstruction. Therefore, the sampling frequency is an important factor that will affect trajectory reconstruction.

In the analysis, if we use the update frequency as a variable, it is difficult for us to control the time interval increment, which may impact the conclusion. In this paper, we propose a new idea to consider the points of our collated dataset as all the points. That is, the trajectory obtained by these points is the ground truth trajectory. We want to control the frequency of updates by controlling the number of known points as a proportion of the total number of points in the dataset. It is important to note that the selection of known points is not random but uniform because we want to ensure that the time interval between two different updates is the same in a fixed time. In that case, when the number of known points is increasing, the update frequency will also be increasing. Then, we analyze the applicability of the model by controlling the ratio of the number of known points to all points, instead of controlling the update frequency. The trajectory between every two update points is reconstructed, and they are stitched together to form a complete trajectory.

In this study, since the reconstructed trajectory accuracy with no stop has reached a good level, we focus our study on vehicles with one-stop, and a vehicle without stopping is selected as a reference. The MAE value with only two updates is considered the vertical coordinate of the zero point. We choose ten percent as the proportional increment of the known number of particles to study the change of MAE. Taking vehicles v_5, v_{19}, and v_{24} as an example and v_2 as a reference, a sensitivity analysis was performed. As shown in Figure 7, the MAEs of the three vehicles were varied following the change of the percentage of the known points of the total.

Through Figure 7, we can see that all three curves show a decreasing trend, which means that when there are more known points that can represent the more frequent updates at a fixed time, we can get the higher accuracy of the reconstructed trajectory, and this proves the applicability of
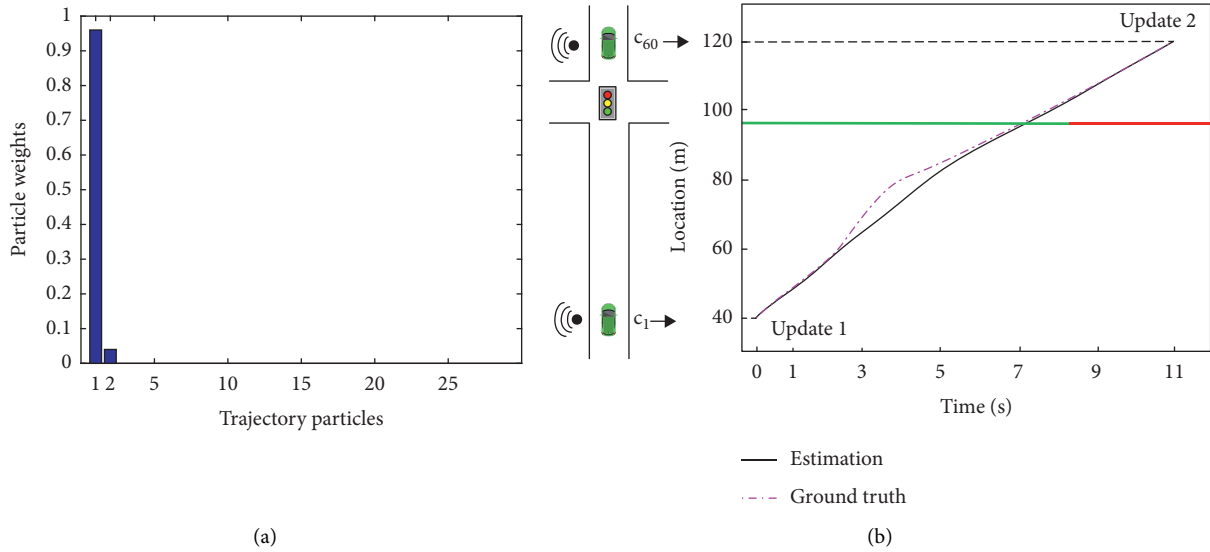
(a)

(b)

FIGURE 4: Reconstructing the trajectory of v_2. (a) Particle weights of v_2. (b) Comparison between the estimation and the truth.
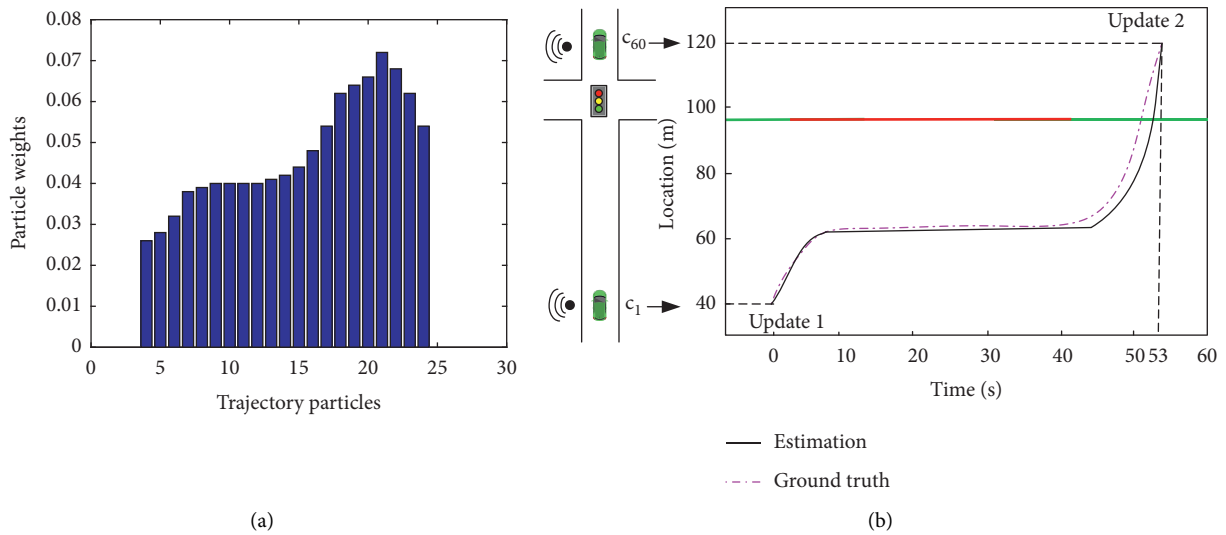

(a)

(b)

FIGURE 5: Reconstructing the trajectory of v_5. (a) Particle weights of v_5. (b) Comparison between the estimation and the truth.

the model. One of the curves for vehicle v_5, which starts with a clear tendency to fall and then rise, is different from the remaining two curves and may be a result of the actual conditions encountered by the vehicle on the road, thus influencing the results. In the range of 50 to 70 percent of the total number of known points, we can see that the three curves have a significant decline in the slope of slowing down the situation, which means that the rate of change of MAE is not obvious; then in the range of 50 to 70 percent of the total number of known points, the most favorable update

frequency has been reached, which proves that the applicability of the model has reached a good effect. The decline of the curve of vehicle v_2, unlike the other three curves, is relatively smooth, but within the first twenty percent of the known point scale, the decline of the curve is faster than the latter. The results demonstrate that the method can have good applicability in the case of the vehicles without stopping, the update frequency is at a relatively lower condition compared to the case of the vehicle with one-stop, and the overall change is relatively smooth.
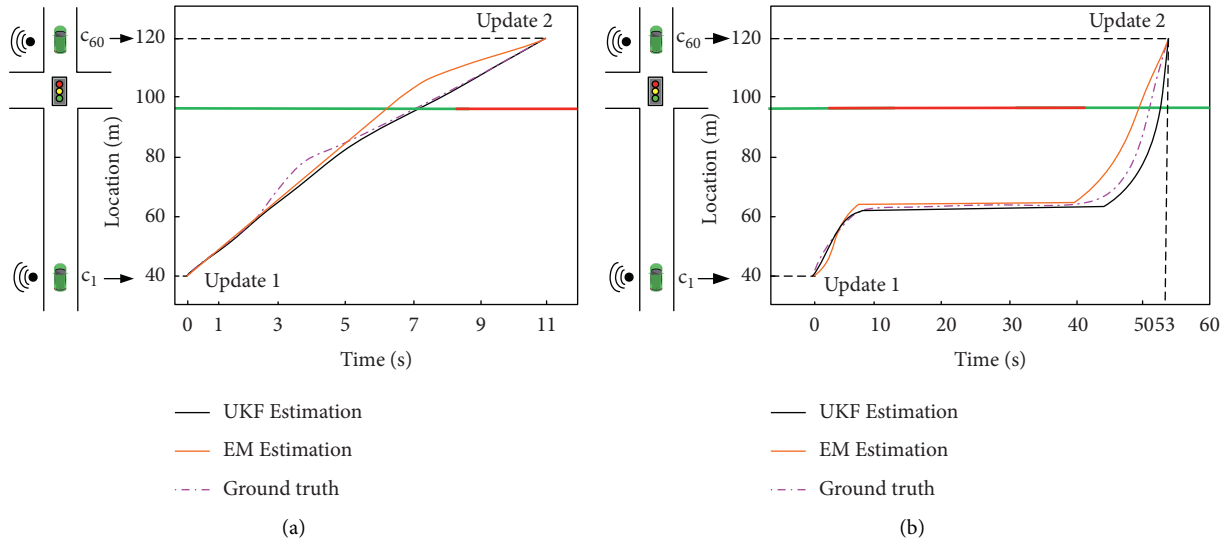
(a)

(b)

Figure 6: Comparison of methods. (a) Vehicle v_2. (b) Vehicle v_5.

Table 1: Comparison of results of trajectory reconstruction methods.

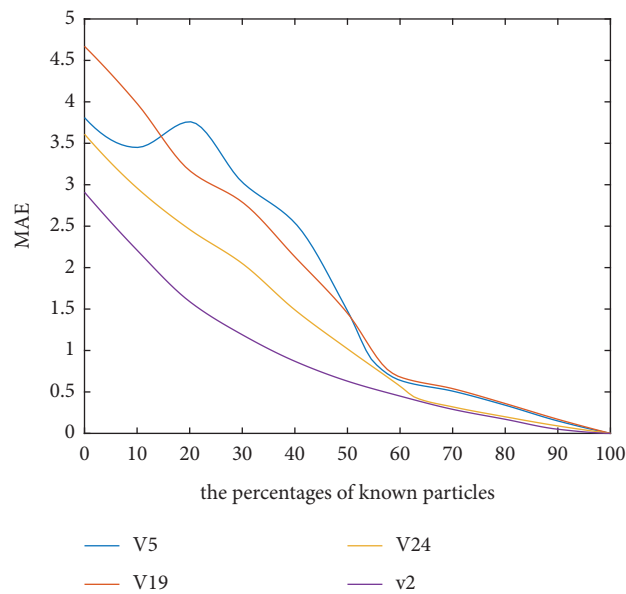| Vehicle ID	Number of stops	Method	MAE (m)
2	0	EM	5.62
		Proposed in this paper	2.91
5	1	EM	4.27
		Proposed in this paper	3.81
8	0	EM	6.41
		Proposed in this paper	3.32
19	1	EM	5.53
		Proposed in this paper	4.67

Figure 7: MAE follows the proportional change of known points.

5. Conclusions

Due to the limitation of many factors, the trajectory data we often obtain are relatively low in sampling frequency, and the update in time and position is sparse. And in the previous research, the trajectory data obtained from the detection vehicle will be studied as the real trajectory, but the obtained trajectory and the actual are erroneous and are not the real trajectory. Based on this problem, in order to get a more accurate vehicle trajectory reconstruction method for sparse detection data, we propose a UKF-based method. Based on the NGSIM database, after preprocessing, a new dataset is constructed, and we divide the signal-controlled intersection into multiple road sections and estimate the travel time of each road section; then, by assigning time errors to each road section, the vehicle trajectories with no stop and one-stop are reconstructed (where the case of one-stop requires multiple updates of the stopping position for calculation). Example analysis was performed on a single signalized intersection.

In this paper, we selected Lankershim Boulevard Street for the case study, and we selected the intersection front and rear length of 120 m as the study area. Compared with the actual trajectory, the MAE values by the proposed method range from 2.91 m to 4.67 m. In contrast to the existing method, i.e., the EM-based method, the MAE values range from 4.27 m to 6.41 m. The comparison results prove that the trajectory reconstruction based on the UKF method has good accuracy. Since the update frequency of trajectory points affects the trajectory reconstruction, we conducted a sensitivity analysis in order to verify the applicability of the model and find the suitable update frequency. The results show that the proposed method is highly applicable when the number of known trajectory points accounts for 50% to 70% of the total trajectory points or a higher percentage.

The presented work provided a new approach to vehicle trajectory reconstruction that can be used to solve a series of traffic problems. In order to improve our approach to making it more generalizable, future research is planned by considering the case of oversaturated signalized intersections and the case of urban arterial roads with multiple intersections, which can also incorporate traffic shock wave theory for the trajectories of vehicle queues.

Data Availability

The dataset used to support the findings of this study has come from the Federal Highway Administration's (FHWA) Next Generation Simulation project, which began in 2002. It is a publicly available dataset. It can be downloaded and accessed on the website of https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] X. Qi, Y. Ji, W. Li, and S. Zhang, "Vehicle trajectory reconstruction on urban traffic network using automatic license plate recognition data," *IEEE Access*, vol. 9, Article ID 49110, 2021.

[2] J. Zheng, S. Weili, H. Shihong et al., "Traffic signal optimization using crowdsourced vehicle trajectory data," in *Proceedings of the Transportation Research Board 97th Annual Meeting*, Washington DC, US, July 2018.

[3] Z. Liu, Y. Liu, Q. Meng, and Q. Cheng, "A tailored machine learning approach for urban transport network flow estimation," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 130–150, 2019.

[4] H. Zhang, H. X. Liu, P. Chen, G. Yu, and Y. Wang, "Cycle-based end of queue estimation at signalized intersections using low-penetration-rate vehicle trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3257–3272, 2019.

[5] T. Vajakas, J. Vajakas, and R. Lillemets, "Trajectory reconstruction from mobile positioning data using cell-to-cell travel time information," *International Journal of Geographical Information Science*, vol. 29, no. 11, pp. 1941–1954, 2015.

[6] Z. Li, R. Kluger, X. Hu, Y.-J. Wu, and X. Zhu, "Reconstructing vehicle trajectories to support travel time estimation," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 42, pp. 148–158, 2018.

[7] T. Vieira da Rocha, L. Leclercq, M. Montanino et al., "Does traffic-related calibration of car-following models provide accurate estimations of vehicle emissions?" *Transportation Research Part D: Transport and Environment*, vol. 34, pp. 267–280, 2015.

[8] M. Chen, G. Yu, P. Chen, and Y. Wang, "A copula-based approach for estimating the travel time reliability of urban arterial," *Transportation Research Part C: Emerging Technologies*, vol. 82, pp. 1–23, 2017.

[9] Z. Zhang, Y. Wang, P. Chen, Z. He, and G. Yu, "Probe data-driven travel time forecasting for urban expressways by matching similar spatiotemporal traffic patterns," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 476–493, 2017.

[10] P. Chen, L. Wei, F. Meng, and N. Zheng, "Vehicle trajectory reconstruction for signalized intersections: a hybrid approach integrating Kalman Filtering and variational theory," *Transportation Business: Transport Dynamics*, vol. 9, no. 1, pp. 22–41, 2021.

[11] C. Zhang, J. Wang, J. Lai, X. Yang, Y. Su, and Z. Dong, "Extracting origin-destination with vehicle trajectory data and applying to coordinated ramp metering," *Journal of Advanced Transportation*, vol. 2019, Article ID 8469316, 15 pages, 2019.

[12] C. Zhang, B. Y. Chen, W. H. K. Lam et al., "Vehicle Re-identification for lane-level travel time estimations on congested urban road networks using video images," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[13] J. Zhao, V. L. Knoop, and M. Wang, "Two-dimensional vehicular movement modelling at intersections based on optimal control," *Transportation Research Part B: Methodological*, vol. 138, pp. 1–22, 2020.

[14] B. Hoh, T. Iwuchukwu, Q. Jacobson et al., "Enhancing privacy and accuracy in probe vehicle-based traffic monitoring via virtual trip lines," *IEEE Transactions on Mobile Computing*, vol. 11, no. 5, pp. 849–864, 2011.

[15] Y. Dai, J. Shao, C. Wei, D. Zhang, and H. T. Shen, "Personalized semantic trajectory privacy preservation through trajectory reconstruction," *World Wide Web*, vol. 21, no. 4, pp. 875–914, 2018.

[16] L. Du and H. Dao, "Information dissemination delay in vehicle-to-vehicle communication networks in a traffic stream," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 66–80, 2014.

[17] M. R. Islam, M. Hadiuzzaman, S. Barua, and T. H. Shimu, "Alternative approach for vehicle trajectory reconstruction under spatiotemporal side friction using lopsided network," *IET Intelligent Transport Systems*, vol. 13, no. 2, pp. 356–366, 2019.

[18] P.-C. Aubin-Frankowski, N. Petit, and Z. Szabó, "Kernel regression for vehicle trajectory reconstruction under speed and inter-vehicular distance constraints," *IFAC-PapersOnLine*, vol. 53, no. 2, Article ID 15084, 2020.

[19] A. Abdelhalim and M. Abbas, "Vt-lane: an exploratory study of an ad-hoc framework for real-time intersection turn count and trajectory reconstruction using nema phases-based virtual traffic lanes," in *Proceedings of the in 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, IEEE, Rhodes, Greece, September 2020.

[20] B. Coifman, "Estimating travel times and vehicle trajectories on freeways using dual loop detectors," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 4, pp. 351–364, 2002.

[21] J. W. C. Van Lint and S. P. Hoogendoorn, "A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 8, pp. 596–612, 2010.

[22] V. Punzo, M. T. Borzacchiello, and B. Ciuffo, "On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1243–1262, 2011.

[23] M. Montanino and V. Punzo, "Making NGSIM data usable for studies on traffic flow theory," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2390, no. 1, pp. 99–111, 2013.

[24] B. Mehran, M. Kuwahara, and F. Naznin, "Implementing kinematic wave theory to reconstruct vehicle trajectories from fixed and probe sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 20, no. 1, pp. 144–163, 2012.

[25] Z. Sun and X. Ban, "Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 268–283, 2013.

[26] P. Hao, K. Boriboonsomsin, G. Wu, and M. Barth, "Probabilistic model for estimating vehicle trajectories using sparse mobile sensor data," in *Proceedings of the in 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1363–1368, IEEE, Qingdao, China, October 2014.

[27] X. Shan, P. Hao, X. Chen, K. Boriboonsomsin, G. Wu, and M. J. Barth, "Probabilistic model for vehicle trajectories reconstruction using sparse mobile sensor data on freeways," in *Proceedings of the in 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 689–694, IEEE, Rio de Janeiro, Brazil, November 2016.

[28] N. Wan, A. Vahidi, and A. Luckow, "Reconstructing maximum likelihood trajectory of probe vehicles between sparse updates," *Transportation Research Part C: Emerging Technologies*, vol. 65, pp. 16–30, 2016.

[29] P. Hao, K. Boriboonsomsin, G. Wu, and M. J. Barth, "Modal activity-based stochastic model for estimating vehicle trajectories from sparse mobile sensor data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 701–711, 2016.

[30] M. R. Fard, A. Shariat Mohaymany, and M. Shahri, "A new methodology for vehicle trajectory reconstruction based on wavelet analysis," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 150–167, 2017.

[31] S. P. Venthuruthiyil and M. Chunchu, "Trajectory reconstruction using locally weighted regression: a new methodology to identify the optimum window size and polynomial order," *Transportmetrica: Transportation Science*, vol. 14, no. 10, pp. 881–900, 2018.

[32] H. Wang, C. Gu, and W. Y. Ochieng, "Vehicle trajectory reconstruction for signalized intersections with low-frequency floating car data," *Journal of Advanced Transportation*, vol. 2019, Article ID 9417471, 14 pages, 2019.

[33] L. Wei, Y. Wang, and P. Chen, "A particle filter-based approach for vehicle trajectory reconstruction using sparse probe data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 2878–2890, 2020.

[34] X. Xie, H. van Lint, and A. Verbraeck, "A generic data assimilation framework for vehicle trajectory reconstruction on signalized urban arterials using particle filters," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 364–391, 2018.

[35] I. Ullah, Y. Shen, X. Su, C. Esposito, and C. Choi, "A localization based on unscented Kalman filter and particle filter localization algorithms," *IEEE Access*, vol. 8, pp. 2233–2246, 2019.

[36] X.-B. Jin, R. J. Robert Jeremiah, T.-L. Su, Y.-T. Bai, and J.-L. Kong, "The new trend of state estimation: from model-driven to hybrid-driven methods," *Sensors*, vol. 21, no. 6, p. 2085, 2021.

[37] B. Coifman and L. Li, "A critical evaluation of the Next Generation Simulation (NGSIM) vehicle trajectory dataset," *Transportation Research Part B: Methodological*, vol. 105, pp. 362–377, 2017.

[38] S. Wang, W. Huang, and H. K. Lo, "Traffic parameters estimation for signalized intersections based on combined shockwave analysis and Bayesian Network," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 22–37, 2019.