Received December 23, 2019, accepted December 31, 2019, date of publication January 6, 2020, date of current version January 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964099

High-Accuracy Off-Line Map-Matching of Trajectory Network Division Based on Weight Adaptation HMM

YAN XIE[1,2], KAI ZHOU[3,4], FANG MIAO[5], AND QIAN ZHANG[4]

[1]College of Geophysics, Chengdu University of Technology, Chengdu 610059, China
[2]Geological Team 103, Guizhou Bureau of Geology Mineral Exploration Development, Tongren 554300, China
[3]College of Computer Science, Sichuan University, Chengdu 610065, China
[4]Science and Technology Information Department, Sichuan Public Security Department, Chengdu 610041, China
[5]Big Data Research Institute, Chengdu University, Chengdu 610106, China

Corresponding author: Kai Zhou (zhoukai0918@qq.com)

This work was supported in part by the Key Research and Development Task of Sichuan Science and Technology Planning Project under Grant 2019YFS0067, in part by the Science and Technology Innovation Talents Team of Manganese Ore Resources Forecast Evaluation in Guizhou under Grant QKHPTRC [2018] 5618, in part by the Science and Technology Plan Project in Guizhou under Grant QKHZC [2019] No. 2868, and in part by the Guizhou Bureau of Geology and Mineral Resources Exploration Geological Scientific Research Project under Grant QDKKH [2017] No. 2.

ABSTRACT In this paper, an accurate off-line map-matching (OM2) system is designed for complex trajectory networks. It is difficult to complex trajectories input into the hidden Markov model (HMM) directly. OM2 includes three key modules that are pre-processing, map-matching based on weight adaptation HMM (WA-HMM), and post-processing. The pre-processing module divides complex multi-trajectory into single-trajectory sets based on the self-defined trajectory division model (TDM) of crossroads. Another core module is the WA-HMM based on Boxplot, which is used to balance efficiency and accuracy of off-line map-matching. The post-processing is used to map the points of the crossroads so as to further improve the map-matching accuracy. OM2 employs the actual GPS trajectories of the internet company and road map of Sichuan province police GIS (PGIS). Our evaluation results show that the accuracy is about 98%, which is suitable for off-line map-matching and solves the problem of complex trajectory network matching being difficult and time-consuming.

INDEX TERMS Map-matching, off-line, complex trajectory division, hidden Markov model, weight adaptation.

I. INTRODUCTION

Map-matching is the process of mapping a series of GPS points with spatio-temporal information and precision loss onto the actual road, to solve related problems in urban computing such as construction of road networks [1], intelligent transportation [2], [3], user travel [4]–[6], trajectory depth understanding [7], [8] and other location-based services [9]–[11]. Matching of trajectories and road networks is divided into on-line and off-line matching. Off-line matching is map-matching after trajectory measurement. One of the main differences between off-line and on-line is whether the entire trajectory is included in the analysis, which provides additional information for selecting the correct segment. Our task is off-line map-matching. Because off-line map-matching is not useful for real-time navigation, it receives less attention than real-time on-line map-matching. The research results regarding improving the accuracy of off-line algorithms are less available than research on on-line algorithms, directly affecting location-based data services such as logistics and supply chain management, vehicle trajectory analysis, etc.

In recent years, with the rapid development of wireless networks [12]–[16] and the ''Internet + sharing economy'', the on-line travel service scene has turned to the mobile terminal, and GPS devices loaded on mobile terminals collect a large number of moving position sequences every day. The positioning accuracy is lower, and the deviation between the data collected by GPS and the actual road is larger (from several meters to tens of meters) due to the positioning and sampling

The associate editor coordinating the review of this manuscript and approving it for publication was Ning Zhang.

errors [17] of the sensor itself, as well as the influence of ambient noise. Due to low-frequency sampling, especially in the complex urban road network, details between any two sampling points are easily lost in the case of higher speed, shorter block, and larger error [18].

However, because the GPS data is taken from a third-party Internet company and there is no perfect solution for secure communication (especially wireless secure communication) [19], [20] and other factors, and the original data cannot be obtained comprehensively, which makes it impossible to make full use of the original data attributes to create a map-matching model. At the same time, for the complexity of the road network structure, the off-line map-matching has more time-consuming, which requires an effective balance of precision and time.

In this paper, we aim to simplify the map-matching model avoiding strong dependent on data attributes, and effectively solve the matching quality of short trajectories and complex trajectories and enhance the accuracy of off-line map-matching. Therefore, we combine the hidden Markov model (HMM) and the Viterbi algorithm to design an optimized method for off-line map-matching.

The specific contributions of this paper are as follows. (1) We designed the architecture of off-line map-matching (OM2), including the three key modules: re-processing, map-matching, and post-processing. (2) We defined the trajectory division model of crossroads. The model addresses a difficult problem that complex trajectory networks input into the HMM directly. (3) We proposed an HMM with weight adaptation (WA-HMM) based on Boxplot to balance efficiency and accuracy of off-line map-matching. (4) We implemented the OM2 and evaluated the performance using the actual GPS trajectories. The experimental accuracy reache about 98%.

The rest of the paper are organized as follows. In Section II, related work is introduced. In Section III, we describe the OM2 system in detail. We present the experimental results in Section IV. In Section V, we discuss the interrupt processing. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

Map-matching algorithms can be divided into many classes based on different dimensions. For one thing, according to the sampling point range, algorithms can be divided into local/incremental [21]–[24] and global [25]–[27]. The local algorithm presents the phenomenon of ''arc jump'', which leads to a significant decrease in accuracy [23]. This method only considers the current position, and the correlation between adjacent points is ignored. The results are greatly affected by measurement error, and the accuracy is generally low. But it is useful for on-line matching because it is fast and real-time. And the global algorithm takes the whole sampling trajectory into account and shows greater robustness to the reduction of the sampling rate, and is more suitable for off-line matching tasks. From another point of view, the algorithms can be further divided into geometric matching algorithms [28]–[30], topological relation algorithms [31], [32], probabilistic statistical algorithms [33]–[36], and advanced matching algorithms [37]–[43].

Some researchers [44]–[46] have developed approximate techniques to generate an unknown potential map or perform map-matching without referencing a known map topology, but by observing clusters of trajectories. Rappos et al. [47] used physical mechanics and borrowed from the field of force-directed graph drawing to employ off-line map-matching, which is weak on solving the problem of matching short trajectories and complex trajectories. Advanced matching algorithms are widely used because they comprehensively consider the geometric information, topological relations, probability, and other information involved in traditional methods, which greatly improves the accuracy and performance.

HMM, an advanced map matching algorithm, was released by Microsoft in 2009 [48]. The HMM model is a Markov process statistical model with hidden unknown parameters. The basic idea is to disassemble the probability of GPS points for candidate roads into the combination observation probability and transition probability. HMM has achieved good results in map-matching of many scenes, and is becoming more and more popular. Sharath proposed a dynamic two-dimensional method to map match by incorporating road width and dynamic weight coefficients [49]. For enhancing the efficiency of map matching, Fiedler *et al.* [50] prosed a scalable map matching algorithm based on Dijkstra's shortest path and Yan *et al.* [51] put forward the optimal path searching via finding the key points in the discrete trajectory. In addition, many scholars have developed the integration methods of HMM and multi-dimensional information. The authors proposed the junction decision domain model that includes the road network accuracy, the GPS accuracy, the road segment width, and the angle between two road segments, which can decrease the error rate of junction matching [43]. The turn restrictions were applied in the sparse map-matching result in the fast running time [52].

Those algorithms based on HMM achieved good results in computational efficiency and accuracy. However, those methods depend on the extra data information or introduce others knowledge. It is difficult to our data characteristics that are diverse data sources, insufficient data attributes, and complex structure of road network. A variety of factors mentioned have introduced great challenges to off-line map-matching.

Therefore, our method comprehensively considers the factors above. We achieve the off-line map-matching based on trajectory network division and weight adaptation HMM (WA-HMM). The trajectory division model is simple and have little dependence on multi-dimensional data attributes. WA-HMM can effectively balance the off-line map-matching accuracy of complex trajectory and running time.

III. OM2 SYSTEM

Map-matching is the process of mapping GPS points onto the actual road segments. The architecture of OM2 is shown in Fig. 1. The input data is the trajectory network, and the

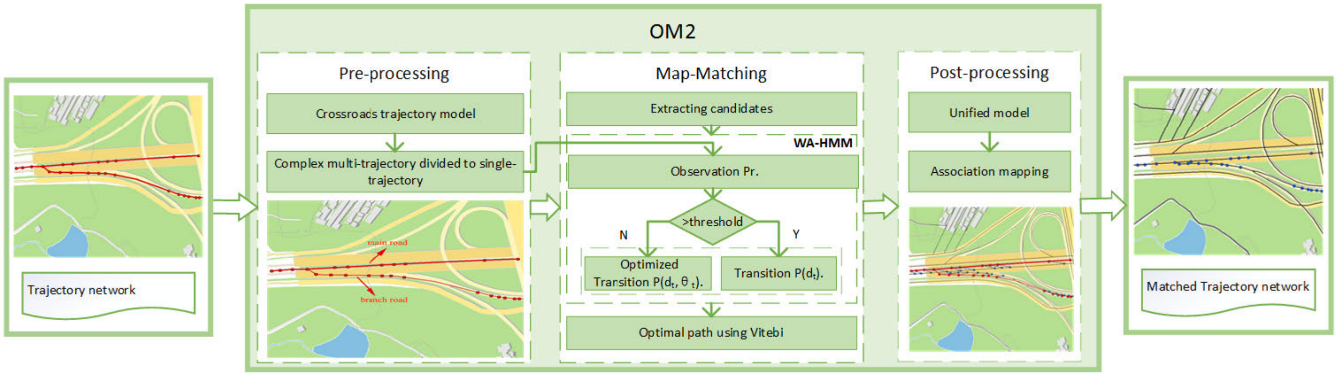


FIGURE 1. OM2 architecture with three key modules.

output data is the matched trajectory data mapped to the road. The middle is the three modules: pre-processing, map-matching, and post-processing.

The detail descriptions of the three modules are as follows. (1) The pre-processing module is used to divide the trajectory network (please refer to Definition 1) into single-trajectory (please refer to Definition 2) sets based on the self-defined trajectory division model of crossroads. The module addresses a difficult problem: the complex trajectories input into the HMM directly. (2) Another core of OM2 is an adaptive HMM with weight adaptation (WA-HMM) to balance efficiency and accuracy. (3) The post-processing is based on the united trajectory division model of GPS and the road, and maps the points of the crossroad to the actual road segment's position so as to further improve accuracy.

Definition 1 (Trajectory Network): A collection of all trajectories on a map, which is represented as undirected graph $G = \langle V, E \rangle$. E is the set of edges between two nodes, also known as the trajectory segments set. V is the set of all the nodes. According to the definition of node degree in the undirected graph, the node is subdivided into endpoints (starting and ending points) with degree 1, ordinary point with degree 2, and intersection point with any degree greater than or equal to 3. The degree of each node is denoted as D_1, D_2, D_{3+}.

Definition 2 (Single-Trajectory): A series of GPS timing point sets that are obtained by the trajectory network division, which can be expressed as $Tr = \{(p_1, t_1, D_1), \ldots, (p_i, t_i, D_2), \ldots, (p_n, t_n, D_1)\}$, $p_i = (x_i, y_i)$. Each GPS point contains latitude x_i, longitude y_i, time identification t_i, degree D $(D \leq 2)$.

A. SELF-DEFINED TRAJECTORY DIVISION MODEL FOR THE CROSSROADS

1) ASSUMPTION OF THE CROSSROADS TRAJECTORY DIVISION

An intersection is the key road structure of trajectory division. Its main characteristics are the connection relationship and connection mode between segments. The connecting relationship expresses which sections are passable, and which can be obtained directly by the vector trajectory network.

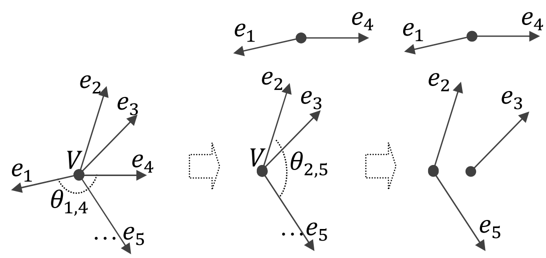


FIGURE 2. Diagram of the crossroads trajectory division model.

The angle between segments is used to describe the connection mode, and the angle can be calculated by the cosine theorem. Based on people's behavior habits, the actual situation of the road network and road traffic planning, we make the following assumptions about the intersection trajectory division. On the one hand, at the intersection, drivers are more willing to choose the path with less change in direction. On another hand, in the real road network, roads are divided into the main road, branch road, secondary branch road. According to relevant laws and regulations and industrial standards, the main road is generally straighter than the branch road and secondary branch road, which means the inner angle of the main road is larger than the non-main road.

2) SELF-DEFINED TRAJECTORY DIVISION MODEL OF THE CROSSROADS

We first identify the crossroad as the place of intersection of two or more roads (main road, branch road and secondary branch, etc.). Namely, according to definition 1 and graph theory, V is the node of the intersection ($degree\,(V) \geq 3$). As shown in Fig. 2, the self-defined trajectory division model (TDM) of crossroads is defined as follows:

$$TDM = \begin{cases} I.get\left(max\left(\theta_{i,j}\right)\right), & other \\ None, & max\left(\theta_{i,j}\right) < \dfrac{\pi}{3} \end{cases} \tag{1}$$

where crossroads consist of V and $E = \{e_1, e_2, e_3, \ldots, e_n\}$, $n = degree(V)$. The angle between any two edges is $\theta_{i,j} = \arccos \dfrac{\overrightarrow{e_i} \cdot \overrightarrow{e_j}}{\left|\overrightarrow{e_i}\right|\left|\overrightarrow{e_j}\right|}, i, j = 1, 2, 3, \cdots, n, i \neq j$. $I = \left\{\theta_{i,j} : \{e_i, e_j\}\right\}$ denotes the intersection of roads.

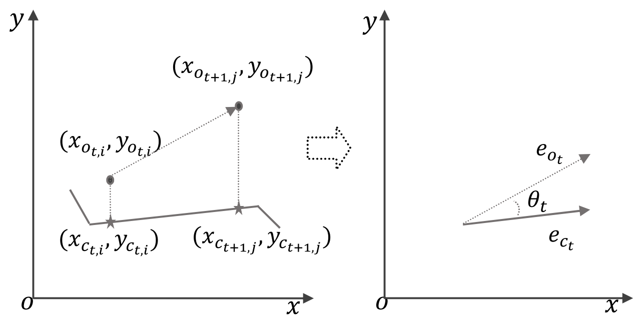


FIGURE 3. Computing the angular deviation using the cosine theorem.

Algorithm 1 Trajectory Division

Input: GPS trajectory network
Output: *trajectory_list*

Generate dictionaries E_{normal} and $E_{intersection}$ based on degree of nodes, dictionary is defined as $\{point : \{edge, \ldots\}, \ldots\}$
for each $Intersection \in E_{intersection}$ **do**
  $point \leftarrow Intersection.keys$
  $edge\_list \leftarrow Intersection.values$
  while len($edge\_list$) ≥ 2 **do**
    $edge\_pair \leftarrow$ combinations($edge\_list$,2)
    $\theta_{i,j} \leftarrow \arccos \frac{\overrightarrow{e_i} \cdot \overrightarrow{e_j}}{\left|\overrightarrow{e_i}\right|\left|\overrightarrow{e_j}\right|}$
    $I \leftarrow \{\theta_{i,j} : \{e_i, e_j\}\}$
    $\theta_{i,j} \leftarrow \max(\theta_{i,j})$
    if $\theta_{i,j} > \pi/3$ **then**
      $\{e_i, e_j\} \leftarrow I.get(\max(\theta_{i,j}))$ //Get the main road
      $E_{normal}.append(\{point : \{e_i, e_j\}\})$
    else
      Split $\{e_i, e_j\}$ to $\{e_i\}$ and $\{e_j\}$
      $E_{normal}.append(\{point : \{e_i\}\})$
      $E_{normal}.append(\{point : \{e_j\}\})$
    end if
    $edge\_list.remove(e_i)$
    $edge\_list.remove(e_j)$
    if len($edge\_list$) $= 1$ **then**
      $E_{normal}.append(\{point : \{edge\_list\}\})$
    end if
  end while
end for
Connect all edges of E_{normal} according to topology
Generate singletrajectory set and store it in the *trajectory_list*
return *trajectory_list*

According to the trajectory division model (TDM) and the diagram (Fig. 3), we can obtain the three new trajectory segment sets $[\{e_1, e_4\}, \{e_2, e_5\}, \{e_3\}]$ used for producing the single-trajectory. The trajectory division algorithm is implemented in Algorithm 1, in which, the points stored in Enormal contain endpoints and ordinary points.

TABLE 1. Rule definition with prior knowledge.

Probability	Rules
Initial Pr.	The probability that the moving object is located on a certain road segment at the beginning is calculated by using the observation probability.
Observation Pr.	The closer the trajectory point is to the candidate road segment, the higher the probability of belonging to this road segment.
Transition Pr.	The closer the two trajectories are, the higher the probability of state transition. The closer the path distance between two points on the real road segment and the distance between two track points, the higher the probability of state transition.

B. WEIGHT ADAPTATION HMM FOR OFF-LINE MAP-MATCHING

The connectivity model of roads is considered by HMM. Meanwhile, HMM considers many different path assumptions to solve the map-matching problem. Lamb and Thiebaux [53] first used HMM for map-matching, using a combination of the Kalman filter and HMM. Several Kalman filters traced the vehicle along different hypothetical paths, and HMM chose between them. Hummel [54] and Krumm [55] used HMM to balance the measured noise and path probability, and the effect was useful. This paper combines map-matching with HMM and the Viterbi algorithm, determines the hidden actual position sequence through the observable GPS position sequence, and calculates the initial state probability matrix, observed state probability matrix, and state transition probability matrix. At the same time, certain rules are used to fit the real historical data. Among them, the definition of rules should satisfy people's prior knowledge (details please refer to Table 1).

1) OBSERVATION PROBABILITY

$$p\left(o_{t,i} \mid c_{t,i}\right) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-0.5\left(\frac{d_{t,i}-u}{\sigma_z}\right)^2} \tag{2}$$

where $d_{t,i} = dist\,(o_i, c_i)$ is the Euclidean distance between $o_{t,i}$ and $c_{t,i}$, with mean $u = 0$ and standard deviation $\sigma = 7$ m based on empirical findings and statistical analysis of the data.

2) TRANSITION PROBABILITY

The transition probability is defined as the probability that the shortest path from a given candidate point $c_{t,i}$ to the next candidate point $c_{t+1,j}$ is the correct path mapping from $o_{t,i}$ to $o_{t+1,j}$. The transmission probability $p\,(d_t)$ is given by the following expressions:

$$p\,(d_t) = \frac{1}{\beta} e^{-d_t/\beta} \tag{3}$$

where,

$$d_t = \left| \|o_t - o_{t+1}\|_{great\_circle} - \left\|c_{t,i^*} - c_{t+1,j^*}\right\|_{route} \right| \tag{4}$$

where the shortest route distance denotes $\| \; c_{t,i^*} - c_{t+1,j^*} \; \|_{route}$ and the greatest circle distance between the measured points is $\| o_{t,i^*} - o_{t+1,j^*} \|_{great\_circle}$. We estimate the value of β based on (5) suggested by Gather and Schulte [56]. Note that in (5), we use c_{t,i^*} and c_{t+1,j^*} as the ground truth road candidate points corresponding to the starting index i^* and ending index j^*.

$$\beta = \frac{1}{\ln(2)} median_t \left(\left| \|o_t - o_{t+1}\|_{great\_circle} - \left\|c_{t,i^*} - c_{t+1,j^*}\right\|_{route} \right| \right) \quad (5)$$

3) OPTIMIZED TRANSITION PROBABILITY

All the branch/secondary branch roads obtained by trajectory division have changes in direction, which are characterized by relatively large changes in angles geometrically.

To improve the matching accuracy, the emission probability is calculated by adding the direction factor, which can be formulated as:

$$p(\theta_t) = e^{-\lambda\theta_t} \quad (6)$$

where according to the test data and our experience, the accuracy is good when setting $\lambda \in (3, 5)$.

The transition probability is optimized as (7):

$$p(d_t, \theta_t) = p(d_t) \cdot p(\theta_t) \quad (7)$$

where because of the complex road network, especially in urban areas, there are many candidate roads segments within the buffer zone radius. It is not necessary to calculate all the differences between each observation point $o_{t,i}$ and the angle of all candidate points c_{t,i^*}, so we have improved the angular deviation θ_t.

$$\theta_t : \varphi_{o_t} \to \varphi_{c_{t,i^*}} \quad (8)$$

with $\varphi_{o_t} \to \varphi_{c_{t,i^*}}$ as the change angle between observation point $o_{t,i}$ and all of candidate points c_{t,i^*} on the road network are matched. The angle deviation $\varphi_t : \varphi_{o_t} \to \varphi_{c_{t,i^*}}$ is calculated by the cosine theorem, shown as (9):

$$\theta_t = arccos \frac{\overrightarrow{e_{o_t}} \cdot \overrightarrow{e_{c_t}}}{\left|\overrightarrow{e_{o_t}}\right| \left|\overrightarrow{e_{c_t}}\right|} \quad (9)$$

where since the road data has no direction attribute, it is impossible to directly determine whether the included angle is θ or $\pi - \theta$. Therefore, the included angle is uniformly normalized as $\theta_t \in [0, \pi/2]$:

$$\theta_t = \begin{cases} \theta_t, & \theta_t \le \dfrac{\pi}{2} \\ \pi - \theta_t, & \theta_t > \dfrac{\pi}{2} \end{cases} \quad (10)$$

where the coordinate of observation point $o_{t,i}$ refers to $(x_{o_{t,i}}, y_{o_{t,i}})$, and $(x_{c_{t,i}}, y_{c_{t,i}})$ denotes the coordinate of candidate points $c_{t,i}$. The coordinate of observation point $o_{t+1,j}$ refers to $(x_{o_{t+1,j}}, y_{o_{t+1,j}})$, and $(x_{c_{t+1,j}}, y_{c_{t+1,j}})$ denotes the coordinate of candidate points $c_{t+1,j}$. We use the starting and ending coordinates of the GPS trajectory $(x_{o_{t,i}}, y_{o_{t,i}})$ and $(x_{o_{t+1,j}}, y_{o_{t+1,j}})$ to construct the GPS trajectory vector e_{o_t}.

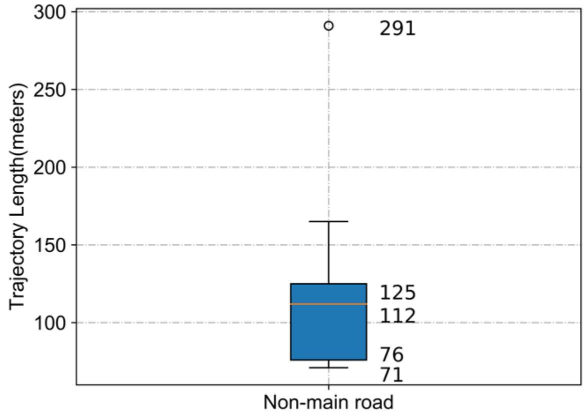


FIGURE 4. Choice threshold D based on data analysis using Boxplot.

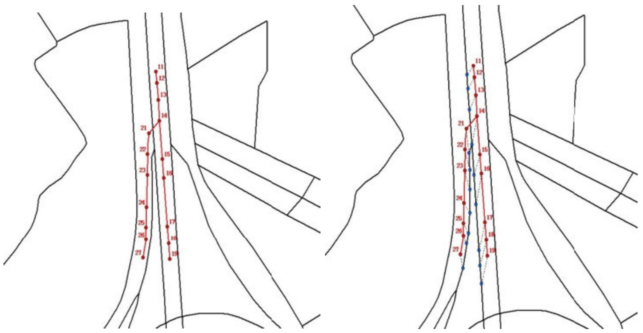
Algorithm 2 Weight Adaptation HMM for Off-Line Map-Matching

Input: *trajectory_list, road network*
Output: *trajectory_matchied*
: Calculate the threshold of the GPS short trajectory based on the Boxplot
for each *trajectory* $\in$ *trajectory_list* **do**
  Calculate candidate segments for each GPS point of the trajectory
  $p\left(o_{t,i}|c_{t,i}\right) \leftarrow \frac{1}{\sqrt{2\pi}\sigma_z} e^{-0.5\left(\frac{d_{t,i}-u}{\sigma_z}\right)^2}$
  $p(d_t) \leftarrow \frac{1}{\beta} e^{-d_t/\beta}$
  if trajectory.geolength < threshold **then**
    $p(\theta_t) \leftarrow e^{-\lambda\theta_t}$
    $p(tr_t) \leftarrow p(d_t) \cdot p(\theta_t)$
  else
    $p(tr_t) \leftarrow p(d_t)$
  end if
  Viterbi decoding
  Add matched results to trajectory_matched
end for
return *trajectory_matched*

Similarly, we use the mapping coordinates on the road from the start and end of the GPS trajectory $(x_{c_{t,i}}, y_{c_{t,i}})$ and $(x_{c_{t+1,j}}, y_{c_{t+1,j}})$ to construct the road vector e_{c_t} (as shown in Fig. 3).

4) WEIGHT ADAPTATION BASED ON BOXPLOT

The real track of the vehicle cannot be restored due to the off-line track. The direction is important track information that needs to be considered for a driving track containing branch roads. Specially, when the GPS track of a non-main road is matched, adding the direction factor to calculate the observation probability can effectively improve the matching accuracy. However, if the direction factor of each point is considered when calculating the observation probability of HMM, a large amount of time will be consumed. Therefore, it is necessary to choose a reasonable trajectory



(a) (b)

FIGURE 5. Precise mappings of intersection points. (a) input; (b) the exact matching of the point level. Red points indict GPS points. The black line is the digital map of PGIS. Blue points denote the exact mapping points via post-processing of the point level.

length threshold. In this paper, Boxplot is used to calculate this critical threshold.

For a set of discrete data, Boxplot can intuitively show the discrete degree of the data, and set its lower quartile, median and upper quartile as Q_1, Q_2, Q_3 respectively. According to Tukey's test outlier estimation method, let us make the following definitions.

The minimum estimated value is:

$$LowerLimit = Q_1 - k\,(Q_3 - Q_1) \tag{11}$$

The maximum estimated value is:

$$UpperLimit = Q_3 + k\,(Q_3 - Q_1) \tag{12}$$

where k is used to determine the degree of abnormality. Generally, $k = 1.5$ means moderate abnormality, and $k = 3$ means extreme abnormality.

According to the actual data of this paper, Boxplot (as shown in Fig. 4) was used to calculate the threshold value D of non-main road (the branch road and secondary branch, etc.) lengths. Fig. 4 shows us the parameters of Boxplot. We counted the length of non-main roads to obtain the threshold D. The quartiles are $Q_1 = 76$, $Q_2 = 112$, and $Q_3 = 125$, respectively. In this paper, we employed the moderately abnormal $k = 1.5$. According to testing data, we calculated $LowerLimit = 2.5$ and $UpperLimit = 198.5$. Therefore, in our system, the value D was 198 meters ($D \in (2.5, 198.5)$).

The whole off-line map-matching algorithm of weight adaptation HMM is shown in Algorithm 2.

C. POST-PROCESSING

In map matching, line segment level accurate matching is firstly satisfied. Namely, the current segment of the road is determined. Then, the exact matching of the point level is satisfied. That is, the specific position in the current road segment is determined. The specific position point determined is usually located in the crossroads, road junction, or complex intersection. The post-processing uses the united trajectory division model (TDM) to map the points of the crossroad to the actual road so as to further improve accuracy.

As shown in Fig. 5(b), a few local adjacent points around No. 14 point need to finish exact matching of the point level

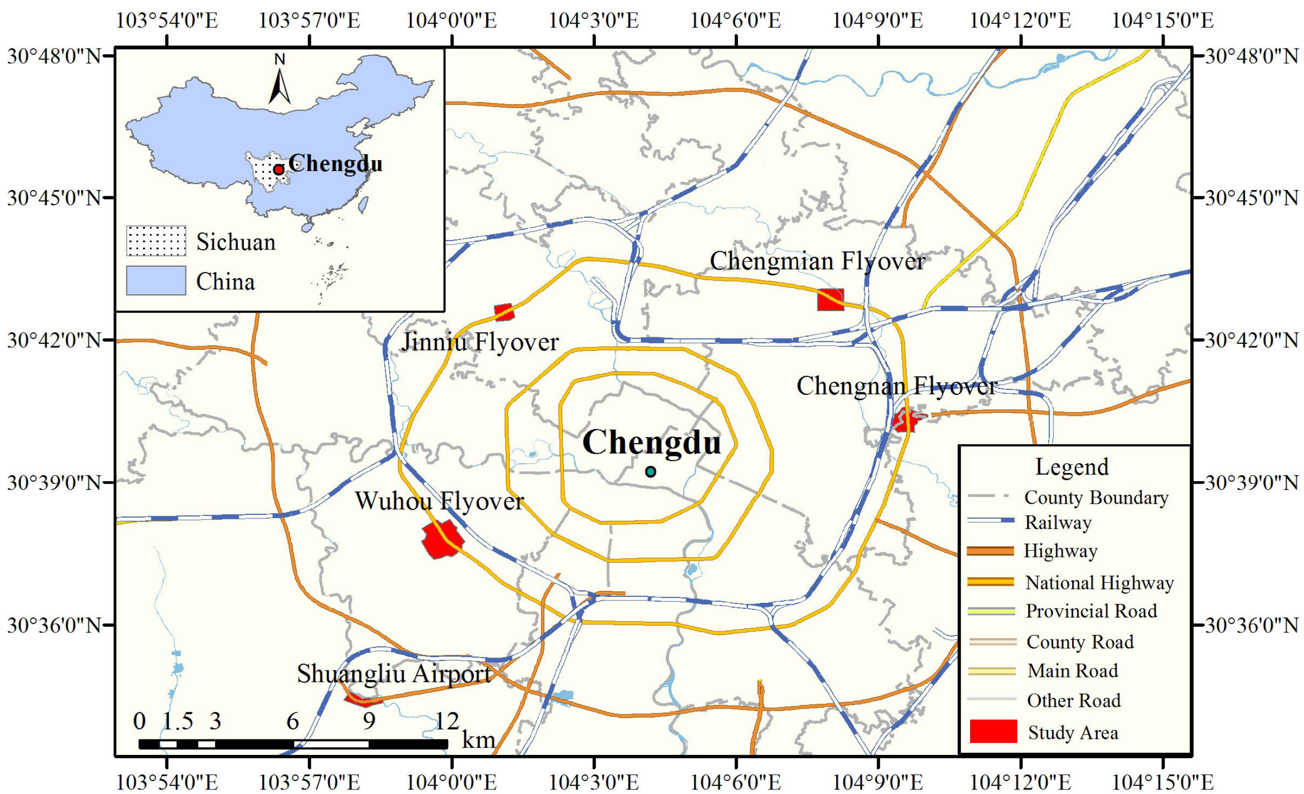


FIGURE 6. The location of the five testing areas located in Chengdu, Sichuan province.

FIGURE 7. GPS data used for testing in Chengdu, Sichuan, China. This consists of five areas, including Chengmian Flyover (a), Jinniu Flyover (b), Wuhou Flyover (c), Chengnan Flyover (d), and Shuangliu Airport (e). The red dots represent GPS track points and the blue lines represent GPS track lines, which are superimposed on the road map.

TABLE 2. The testing data including the road area (km²), trajectories length (km) and points of GPS.

	road area	GPS trajectories length	GPS points
Chengmian Flyover	0.76	16.10	683
Jinniu Flyover	0.37	8.42	170
Wuhou Flyover	0.82	21.92	1105
Chengnan Flyover	0.76	15.80	605
Shuangliu Airport	0.40	33.80	757
Total	3.11	96.04	3320

exactly associated based on the unified trajectory division model of GPS and road networks, namely the self-defined TDM using (1). In the second step, the adjacent points around No. 14 point are mapped by using the method of equal proportion mapping. At the same time, other distant points in the example can be mapped by using the method of vertical projection. Of course, we would like to employ an interactive post-processing method, such as the visual interactive map-matching proposed by Kruger *et al.* [57].

IV. EXPERIMENTAL RESULTS

A. DATA

The GPS trajectory data is provided by the third-party Internet company, and the road network is from the Sichuan police GIS (PGIS). The data covers the whole Sichuan Province for LBS from the practical requirements of the PGIS. The road network data is the basic data of the PGIS digital map. GPS trajectory data is non-real-time historical data and has complex tracks, is low-frequency, and has coarse-grained network positioning information. The statistical results of a large number of data show that the deviation between the GPS data and PGIS digital map is large (about 7.2 m), which increases the difficulty of actual map matching. Sichuan Province has an administrative area of 486,000 km^2, the total length of GPS navigation trajectories is 344,000 km, and the total length of the road network is 470,000 km. Fig. 6 shows the location of the study area. Based on the actual situation, we selected five representative regions in Chengdu, Sichuan for testing, and comprehensively considered the complexity and importance. Fig. 7 is the GPS testing data of five areas: Chengmian Flyover, Jinniu Flyover, Wuhou Flyover, Shuangliu Airport and Chengnan Flyover, which consists of 3320 points from the 96.04 km trajectories length, and covers about 3.11 km^2 of road area (as shown in Table 2).

B. RESULTS AND EVOLUTION

Map-matching is not easy to evaluate. In this paper, a common assessment method, the correct matching percentage (CMP), can be formulated as (13).

$$CMP = \frac{N_{sr}}{N_{sn}} * 100\% \tag{13}$$

where N_{sr} is the number of samples matched correctly, and N_{sn} denotes the number of samples that were mismatched.

According to (13), we can obtain the test results (as shown in Fig. 8). We observe that our model has the highest average accuracy (more than 98%), and each result of the five groups of tests is the optimal compared with HMM [48] and optimized HMM (the probability of all GPS points calculated by adding the angle factor). We can also observe that the performance of our model is relatively stable.

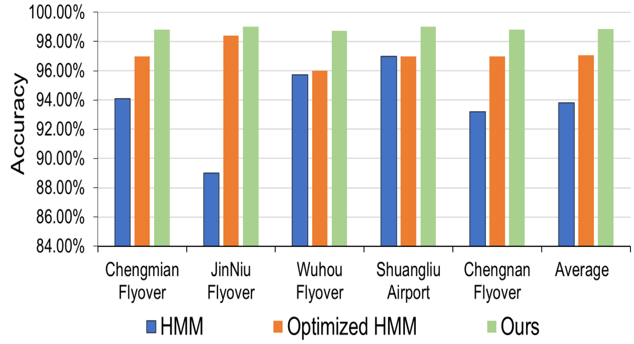


FIGURE 8. Results of HMM, Optimized HMM and our model.

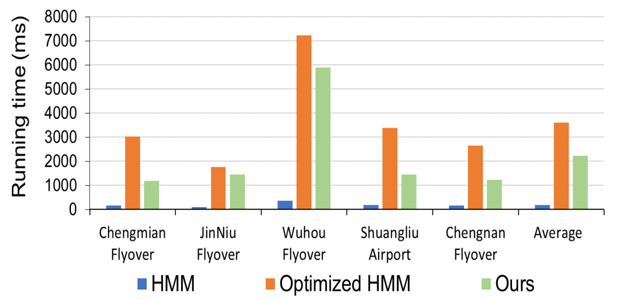


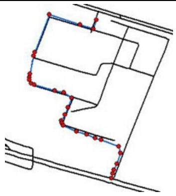
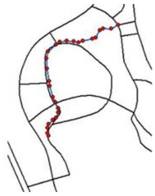
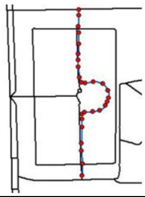
FIGURE 9. The running time of HMM, optimized HMM and ours.

The result of the running time (Fig. 9) shows that our model takes slightly more time than optimized HMM because of mapping post-processing of intersections in the crossroads, in order to further improve the accuracy. However, we have used Boxplot optimization to reduce the running time. An off-line map matching task is different from a real-time on-line task. Instead of the running time, it is preferable to guarantee accuracy first. Experimental results show that the running time is acceptable. The results are trained on a PC with E5-2630 CPU at 2.4GHz with 16GB RAM and CentOS 7 operating system.

V. DISCUSSION ON INTERRUP PROCESSING

In this paper, we assumed that the road information was complete in the process of off-line map-matching. However, there are often incomplete problematic segments such as the topology error, missing road, and road changes. The problematic segments lead to discontinuity and exceptions in the map-matching process using HMM, and cause matching interruptions, as shown in Table 3. We monitor interruptions in the map-matching process and deal with them as they occur. When an interruption is found, the program automatically breaks and records the interruption information to minimize the impact of the interruption on the map-matching output. Interruption handling is another issue we will continue to address in the future work. We reversely focused on the correlation between the interruptions and the problematic segments, so as to solve errors of the matching

TABLE 3. Classification, analysis, and illustration of incomplete road network.

Classification	Analysis	Illustration
Topology Error	There are topology errors in the road network data. Therefore, the transition probability between the corresponding candidate road segment and the following possible road segments is zero or very small.	
Missing	The track data passes through a missing road not recorded by road network data. The observation probability from the track point to candidate road segment is extremely low, or the transition probability of two consecutive trajectories is very small or zero.	
Changing	The track data goes through a changing road that is not recorded by road network data. The observation probability from the track point to candidate road segment is extremely low.	

results caused by the problem segments, improve the matching accuracy, and support the task of updating the road network.

VI. CONCLUSION

In this paper, we proposed the OM2 system, which provides accurate off-line map-matching for complex trajectory networks. We designed the architecture of OM2 system, including three core modules: pre-processing, map-matching of WA-HMM, and post-processing. We present the pre-processing module of OM2 to divide complex multi-trajectories into single-trajectory sets based on the self-defined trajectory division model of crossroads. The pre-processing makes the complex trajectory network inputting into the HMM directly easily. Another core of OM2 is an adaptive HMM with weight adaptation (WA-HMM) based on Boxplot in order to balance efficiency and accuracy. The definite united model of GPS and the road network is a key element of post-processing, which is used to map the points of the crossroad so as to further improve accuracy. Finally, OM2 employs the actual GPS trajectories of the internet company and road map of the Sichuan province. The evaluation shows us an accuracy of about 98%. Our is effective for solving the problems of map-matching difficult and time-consuming for complex multi-trajectory network.

In the future, we will continue to monitor and deal with program interruptions caused by incomplete roads such as topology errors, missing roads, and road changes, so that further improve the off-line map-matching accuracy and update the map.

REFERENCES

[1] W. Yang, "Automated construction of road networks from GPS tracks," in *Proc. 11th Int. Conf. Adv. Geographic Inf. Syst., Appl., Services, GEOProcessing*, Feb. 2019, pp. 35–41.

[2] C. Feijoo, J. Ramos, and F. Perez, "A system for fleet management using differential GPS and VHF data transmission mobile networks," in *Proc. Vehicle Navigat. Inf. Syst. Conf.*, Oct. 1983, pp. 445–448.

[3] R. R. Joshi, "A new approach to map matching for in-vehicle navigation systems: The rotational variation metric," in *Proc. ITS*, Aug. 2001, pp. 33–38.

[4] H. Gonzalez, J. Han, X. Li, M. Myslinska, and J. Sondag, "Adaptive fastest path computation on a road network: A traffic mining approach," in *Proc. Int. Conf. Very Large Data Bases VLDB Endowment*, Sep. 2007, pp. 794–805.

[5] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining-KDD*, Aug. 2012, pp. 195–203.

[6] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-Drive: Enhancing driving directions with taxi drivers' intelligence," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 220–232, Jan. 2013.

[7] Q. Yuan, G. Cong, Z. Y. Ma, A. X. Sun, and M. T. Nadia, "Who, where, when and what: Discover spatio-temporal topics for Twitter users," in *Proc. SIGKDD*, Aug. 2013, pp. 605–613.

[8] X. L. Long, J. Lei, and J. James, "Exploring trajectory-driven local geographic topics in foursquare," in *Proc. Int. Workshop Location-Based Social Netw.*, Sep. 2012, pp. 927–934.

[9] M. Li, Y. Zhang, and W. Wang, "Analysis of congestion points based on probe car data," in *Proc. 12th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2009, pp. 1–5.

[10] M. A. Quddus, R. B. Noland, and W. Y. Ochieng, "A high accuracy fuzzy logic based map matching algorithm for road transport," *J. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 103–115, Sep. 2006.

[11] I. Rauschert, P. Agrawal, R. Sharma, S. Fuhrmann, I. Brewer, and A. Maceachren, "Designing a human-centered, multimodal GIS interface to support emergency management," in *Proc. 10th ACM Int. Symp. Adv. Geographic Inf. Syst.-GIS*, Nov. 2002, pp. 119–124.

[12] L. Ale, N. Zhang, H. Wu, D. Chen, and T. Han, "Online proactive caching in mobile edge computing using bidirectional deep recurrent neural network," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5520–5530, Jun. 2019.

[13] N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen, "Synergy of big data and 5G wireless networks: Opportunities, approaches, and challenges," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 12–18, Feb. 2018.

[14] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. S. Shen, "Software defined space-air-ground integrated vehicular networks: Challenges and solutions," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 101–109, Jul. 2017.

[15] P. Yang, N. Zhang, S. Zhang, L. Yu, J. Zhang, and X. Shen, "Content popularity prediction towards location-aware mobile edge caching," *IEEE Trans. Multimedia*, vol. 21, no. 4, pp. 915–929, Apr. 2019.

[16] Q. Ye, W. Zhuang, S. Zhang, A.-L. Jin, X. Shen, and X. Li, "Dynamic radio resource slicing for a two-tier heterogeneous wireless network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9896–9910, Oct. 2018.

[17] D. Pfoser and C. Jensen, "Capturing the uncertainty of moving-object representations," in *Advances in Spatial Databases* (Lecture Notes in Computer Science), vol. 1651. Berlin, Germany: Springer, Jun. 1999, pp. 111–131.

[18] T. Miwa, D. Kiuchi, T. Yamamoto, and T. Morikawa, "Development of map matching algorithm for low frequency probe data," *Transp. Res. C, Emerg. Technol.*, vol. 22, pp. 132–145, Jun. 2012.

[19] D. Chen, N. Zhang, R. Lu, N. Cheng, K. Zhang, and Z. Qin, "Channel precoding based message authentication in wireless networks: Challenges and solutions," *IEEE Netw.*, vol. 33, no. 1, pp. 99–105, Jan. 2019.

[20] D. Chen, N. Zhang, N. Cheng, K. Zhang, Z. Qin, and X. S. Shen, "Physical layer based message authentication with secure channel codes," *IEEE Trans. Dependable Secure Comput.*, early access, 2019, doi: 10.1109/TDSC.2018.2846258.

[21] S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk, "On map-matching vehicle tracking data," in *Proc. VLDB*, 2005, pp. 853–864.

[22] S. S. Chawathe, "Segment-based map matching," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2007, pp. 1190–1197.

[23] J. Greenfeld, "Matching GPS observations to locations on a digital map," in *Proc. TRB*, 2002.

[24] C. Wenk, R. Salas, and D. Pfoser, "Addressing the need for map-matching speed: Localizing globalb curve–matching algorithms," in *Proc. 18th Int. Conf. Sci. Stat. Database Manage. (SSDBM)*, Aug. 2006, pp. 379–388.

[25] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, "Map-matching for low-sampling-rate GPS trajectories," in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.-GIS*, Seattle, WA, USA, 2009, p. 352.

[26] H. Yin and O. Wolfson, "A weight-based map matching method in moving objects databases," in *Proc. 16th Int. Conf. Sci. Stat. Database Manage.*, vol. 16, Nov. 2004, pp. 437–438.
[27] K. Zheng, Y. Zheng, X. Xie, and X. Zhou, "Reducing uncertainty of low-sampling-rate trajectories," in *Proc. IEEE 28th Int. Conf. Data Eng. (ICDE)*, Apr. 2012, pp. 1144–1155.
[28] D. Bernstein and A. Kornhauser, "An introduction to map matching for personal navigation assistants," *Geometric Distrib.*, vol. 122, no. 7, pp. 1082–1083, 1996.
[29] G. Taylor, G. Blewitt, D. Steup, S. Corbett, and A. Car, "Road reduction filtering for GPS–GIS navigation," *Trans. GIS*, vol. 5, no. 3, pp. 193–207, Jun. 2001.
[30] C. E. White, D. Bernstein, and A. L. Kornhauser, "Some map matching algorithms for personal navigation assistants," *Transp. Res. C, Emerg. Technol.*, vol. 8, nos. 1–6, pp. 91–108, Feb. 2000.
[31] H. Alt, A. Efrat, G. Rote, and C. Wenk, "Matching planar maps," *J. Algorithms*, vol. 49, no. 2, pp. 262–283, Nov. 2003.
[32] M. A. Quddus, W. Y. Ochieng, L. Zhao, and R. B. Noland, "A general map matching algorithm for transport telematics applications," *GPS Solutions*, vol. 7, no. 3, pp. 157–167, Dec. 2003.
[33] M. Abbour, P. Bonnifait, and V. Cherfaoui, "Map matching integrity using multi-sensor fusion and multi-hypothesis road tracking," *J. Intell. Transp. Syst. Technol. Planllingand Oper.*, vol. 6, no. 4, pp. 189–201, 2008.
[34] S. Nadine and W. Kay, "Map-matching of GPS traces on high-resolution navigation networks using the multiple hypothesis technique (MHT)," *Arbeitsberichte Verkehrsund Raumplanun*, vol. 568, pp. 1–22, Oct. 2009.
[35] W. Ochieng, M. Quddus, and R. Noland, "Map-matching in complex urban road networks," *Revista Brasileira De Cartografia*, vol. 55, no. 2, pp. 1–14, 2009.
[36] M. Quddus, R. Noland, and W. Ochieng, "Validation of map matching algorithm using high precision positioning with GPS," *J. Navigat.*, vol. 58, no. 2, pp. 257–271, 2004.
[37] C. Smaili, M. E. El Najjar, and F. Charpillet, "A road matching method for precise vehicle localization using hybrid Bayesian network," *J. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 176–188, Nov. 2008.
[38] W. Kim, G.-I. Jee, and J. Lee, "Efficient use of digital road map in various positioning for ITS," in *Proc. IEEE Position Location Navigat. Symp.*, Mar. 2000, pp. 170–176.
[39] Z. Li and W. Chen, "A new approach to map-matching and parameter correcting for vehicle navigation system in the area of shadow of GPS signal," in *Proc. IEEE Intell. Transp. Syst. (ITSC)*, vol. 5, Oct. 2005, pp. 425–430.
[40] D. Obradovic, H. Lenz, and M. Schupfner, "Fusion of map and sensor data in a modern car navigation system," *J. Signal Process. Syst.*, vol. 45, no. 1, pp. 111–122, 2006.
[41] J. Pyo, D. Shin, and T. Sung, "Development of a map matching method using the multiple hypothesis technique," in *Proc. Intell. Transp. Syst.*, 2001, pp. 23–27.
[42] H. Xu, H. Liu, C. Tan, and Y. Bao, "Development and application of an enhanced kalman filter and global positioning system error-correction approach for improved map matching," *J. Intell. Transp. Syst., Technol., Planning Oper.*, vol. 14, no. 1, pp. 27–36, Feb. 2010.
[43] H. Qi, X. Di, and J. Li, "Map-matching algorithm based on the junction decision domain and the hidden Markov model," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0216476.
[44] K. Buchin, "Clustering trajectories for map construction," Presented at the 25th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL), New York, NY, USA, 2017.
[45] Y. Li, Y. Li, D. Gunopulos, and L. Guibas, "Knowledge-based trajectory completion from sparse GPS samples," Presented at the 24th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS), New York, NY, USA, 2016.
[46] S. Wang, Y. Wang, and Y. Li, "Efficient map reconstruction and augmentation via topological methods," Presented at the 23rd SIGSPA- TIAL Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL), 2015.
[47] E. Rappos, S. Robert, and P. Cudré-Mauroux, "A force-directed approach for offline GPS trajectory map matching," in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.-SIGSPATIAL*, vol. 18, 2018, pp. 319–328.
[48] P. Newson and J. Krumm, "Hidden Markov map matching through noise and sparseness," in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.-GIS*, Seattle, WA, USA, 2009, p. 336.
[49] M. Sharath, N. R. Velaga, and M. A. Quddus, "A dynamic two-dimensional (D2D) weight-based map-matching algorithm," *Transp. Res. C, Emerg. Technol.*, vol. 98, pp. 409–432, Jan. 2019.
[50] D. Fiedler, M. Čáp, J. Nykl, P. Žilecký, and M. Schaefer, "Map matching algorithm for large-scale datasets," Sep. 2019, *arXiv:1910.05312*. [Online]. Available: https://arxiv.org/abs/1910.05312
[51] S. Yan, J. Yu, and H. Zhou, "Fast map-matching based on hidden Markov model," in *Mobile Computing, Applications, and Services*. Cham, Switzerland: Springer, 2019, pp. 85–95.
[52] H. Bast and P. Brosi, "Sparse map-matching in public transit networks with turn restrictions," in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.-SIGSPATIAL*, Seattle, WA, USA, 2018, pp. 480–483.
[53] P. Lamb and S. Thiebaux, "Avoiding explicit map-matching in vehicle location," in *Proc. ITS*, 1999, pp. 8–12.
[54] B. Hummel, "Map matching for vehicle guidance," in *Dynamic and Mobile GIS: Investigating Changes in Space and Time*. Boca Raton, FL, USA: Taylor & Francis, 2006.
[55] J. Krumm, E. Horvitz, and J. Letchner, "Map matching with travel time constraints," Presented at the SAE World Congr. Exhib., May 2007.
[56] U. Gather and V. Schultze, "Robust estimation of scale of an exponential distribution," *Statistica Neerlandica*, vol. 53, no. 3, pp. 327–341, Nov. 1999.
[57] R. Kruger, G. Simeonov, F. Beck, and T. Ertl, "Visual interactive map matching," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 6, pp. 1881–1892, Jun. 2018.

YAN XIE was born in Jilin. She received the bachelor's degree in computer science and technology and the master's degree in computer application technology from the Chengdu University of Technology, Sichuan, in 2010 and 2013, respectively. She is currently pursuing the Ph.D. degree with the Chengdu University of Technology.

She is currently working with the Geological Team 103, Guizhou Bureau of Geology Mineral Exploration Development. She is also a Researcher with the Key Laboratory of Earth Exploration and Information Techniques, Ministry of Education, Chengdu University of Technology. She has published a few articles in SCI, EI, and Chinese core articles. She holds five invention patents. She is interested in artificial intelligence algorithms and big data.

Dr. Xie won the Third Prize of Scientific and Technological Innovation issued by the Bureau of Geology, and the Mineral Exploration and Development of Guizhou Province, in 2015, and the First Prize of Geological Science and Technology Achievements issued by the Bureau of Geology and Mineral Exploration and Development of Guizhou Province, in 2018.

KAI ZHOU was born in Xichang, China, in 1984. He received the bachelor's degree in information management and information system from Shenyang Aerospace University, in 2006, and the master's degree in software engineering from Sichuan University, in 2015, where he is currently pursuing the Ph.D. degree in electronics and information with the College of Computer Science.

He is currently working with the Science and Technology Information Department, Sichuan Public Security Department. Since 2011, he has been a Main Researcher of Sichuan police geographic information system (PGIS). He is interested in artificial intelligence algorithms and big data. He holds one invention patent.

Dr. Zhou won the Third Prize of S&T Progress Award from the Sichuan Society of Surveying, Mapping, and Geoinformation, in 2016, the Third Prize of S&T Progress Award from the Chinese Society for Geodesy Photogrammetry and Cartography, in 2017, and the First Prize of S&T Progress Award from Sichuan Province, in 2018.

FANG MIAO was born in Beijing, China, in 1958. He received the B.S., M.S., and Ph.D. degrees in geophysics from the Chengdu University of Technology, Sichuan, China, in 1982, 1985, and 1990, respectively.

He was a Visiting Scholar with the Bullard Laboratory, Cambridge University, U.K., from March 1994 to May 1995. Since 1995, he has been a Professor with the Department of Computer, College of Information Engineering, Chengdu University of Technology. He is the author of seven books and more than 200 articles. He holds more than ten inventions. His research interests include spatial information technology and its applications, data processing, and data foundation studies. His main contribution was the proposal of the theory and the practice of data-oriented architecture (DOA).

Dr. Miao is the Standing Director of the Sichuan Computer Society. He was awarded the title of Sichuan Ph.D. outstanding contribution, in 1991. He has received the Second Prize of Researching Category of Fund from the Fok Ying Tung Education Foundation, in 1992, enjoyed state department special allowance, in 1998. He was selected as the Hundred Academic and Managerial Leader of China's Information, in 2006. He was a recipient of the Science Chinese Person of the Year 2017.

QIAN ZHANG received the bachelor's degree in international economics and trade from the College of Arts and Sciences, Sichuan Normal University, in 2017.

She is currently working with the Science and Technology Information Department, Sichuan Public Security Department. Since her graduation in 2017, she has been a Researcher with the Data Center of Chengdu Success Territory Technology Company, Ltd. She has participated in the development and application of police GIS system for a long time. She is interested in the research of cross-disciplinary computer applications. In 2018, she has participated in the research of major science and technology planning project of Sichuan Province such as construction of collaborative system for Sichuan Police Spatio-Temporal Big Data Integration and Emergency Disposal Intelligent Command Demonstration Application Project. She has participated in the establishment of the domain adaptation network segmentation and extraction model of the target objects, such as buildings, roads, and water systems, to improve the segmentation and extraction accuracy.