IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 13, 2020, accepted March 30, 2020, date of publication April 2, 2020, date of current version April 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985211

Positioning Algorithm Based on the Fingerprint Database by Twice-Fuzzy Clustering in the High-Speed Railway Scenario

BAOFENG DUAN, CUIRAN LI, (Member, IEEE), JIANLI XIE, DONGMEI ZHOU, AND WEI WU

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Corresponding author: Cuiran Li (licr@mail.lzjtu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61661025 and Grant 61661026, and in part by the Foundation of A Hundred Youth Talents Training Program of Lanzhou Jiaotong University under Grant 152022.

ABSTRACT When the wireless communication network is optimized for the high-speed railway scenario (HSRS), GPS connections are prone to frequent interruptions. The causes of this phenomenon are analyzed after collecting a large quantity of measured data. A positioning algorithm based on the fingerprint database by twice-fuzzy clustering is proposed to obtain the locations of the terminal inside the carriage of high-speed train (HST) in real time. After collecting more than 300,000 sampled data of both network characteristics and location information, the database of original fingerprints has been constructed. The identification and elimination of abnormal fingerprints are helpful to improve the quality of the fingerprint database. The longitude and latitude of the terminal, which is losing the signals of GPS, can be calculated by setting the fingerprint integral counter, threshold value and weighted measurement eigenvalues and by constructing a matrix of dissimilarity. The experimental results show that the proportion of similar fingerprints between 15 and 200 by once-fuzzy clustering is as high as 90.81%; Additionally, the number of over 95.57% of the similar fingerprints is between 1-20 by twice-fuzzy clustering. The proportion of samples with positioning accuracy less than 10 m is 63.33%, and those less than 5 m account for 41.67%. The average positioning accuracy of the proposed algorithm is 9.02 m, which is suitable for acquiring location information when the signals of GPS are losing in HSRS.

INDEX TERMS Clustering algorithms, dissimilarity matrix, fingerprint database, fuzzy set theory, location information.

I. INTRODUCTION

With the continuous increase of railway operating mileage, high-speed train (HST) has become the main form of passenger transport because of the high speed, low energy consumption, large transportation capacity, safety and punctuality [1], [2]. In the networks of the long term evolution (LTE), how to improve the performance of a mobile communication system in the high-speed railway scenario (HSRS) has become an important research topic [3]. Due to both the large number of passengers and the high average revenue per user (ARPU), telecom operators regard high-speed railway as one of the most important scenarios to improve the quality of coverage for the wireless communication network [4]. The optimization of the wireless network can effectively eliminate network interference along the railway, avoid the existence of no-signal areas and improve the user's perception [5], [6]. Generally, the network performance data tested to optimize the wireless network of telecom operators in HSRS include the power and quality of the reference signal and other parameters. By matching these data with the loc0 ation information collected by GPS, the network performance can be analyzed, and the problems of weak coverage are about to be solved. However, the signals of GPS satellites have become very weak through the free-space loss [7], [8], the Doppler frequency shift [9], [10], and the penetration loss due to closed metal structure of HST [11]–[13], resulting in the frequent interruptions of GPS connections.

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

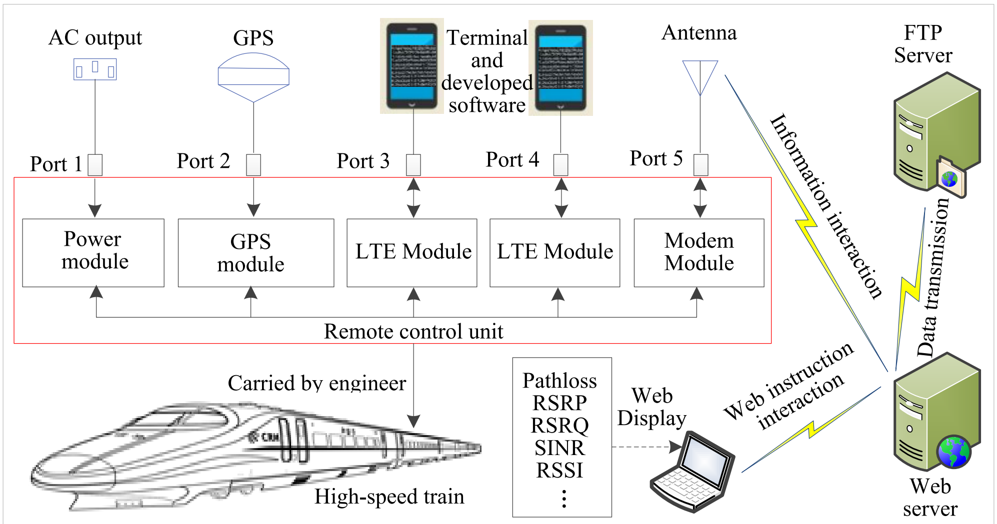


FIGURE 1. Principle of collecting massive original fingerprints.

In this case, how to obtain the locations of the terminal is a new challenge for telecom operators in wireless communication research, such as the analysis of measured data and the planning of the network optimization project in HSRS [14]. It is difficult to analyze the relationship between the network index and the geographic coordinates in real time because the terminal inside the carriage of HST cannot effectively obtain location information. Even if we know that the terminal has gone offline at some point, we cannot determine where it happened. Furthermore, both the project planning of the evolved universal terrestrial radio access network NodeBs (E-UTRAN NodeBs or eNBs) and the wireless optimization of 4G networks cannot obtain reliable support data for telecom operators.

Considering that multipath fading [15] of wireless signal has characteristic attributes at the same location, the method of constructing a fingerprint database with massive measured data can achieve terminal positioning. After two stages of offline collection and online positioning, a fingerprint database of received signal strength (RSS) is built to accomplish indoor terminal positioning [16]. In the wireless local area network (LAN) environment, the coordinates of indoor terminal are calculated based on a multidimensional fingerprint database of received signal strength indication (RSSI) [17]. Using indoor both receiver and transmitter to obtain a large quantity of original location data, a new mechanism of fingerprint identification is created [18]. A positioning algorithm of semi-supervised learning is proposed to form fingerprints by using RSSI spatial correlations of adjacent nodes for terminal locations [19]. A radio map is constructed by collecting signal strength samples in the location area, and the indoor terminal locations are determined by using the fingerprint identification technology [20]. A positioning algorithm utilizing a noise sensor based on a multi-objective evolutionary model is proposed. The noise covariance estimator is used to sense the noise covariance of RSSI and to find the optimal weight of the calibration point [21]. In [22], an acquisition and location system of double databases of fingerprints based on wireless sensor networks is designed according to the goal of low input and high efficiency. An improved fingerprint-based localization approach that adopts a path loss model for fingerprint creation and positioning is proposed [23]. In [24], the authors present an adaptive weighted positioning method of k-nearest neighbor (KNN) based on an omnidirectional fingerprint database (ODFD) and twice-affinity propagation clustering. To solve the challenges of low sampling efficiency and particle impoverishment, a time sequence Monte Carlo localization algorithm based on particle swarm optimization is proposed [25]. At present, however, the positioning algorithms of fingerprint database upon the loss of the signals of GPS are mainly for the scenarios of indoor, or low-speed mobile terminal, or non-railway. There is little research on terminal positioning inside the carriage of HST.

In HSRS, a great deal of experimental experience illustrates that terminal can also collect a large number of fingerprints with characteristic values, such as receiving power and signal quality. The relationship between these characteristic values and the location information creates the condition for positioning algorithm [26], [27]. As is known to all, there are a lot of positioning algorithms. However, the main research directions of these localization algorithms include two directions. On the one hand, some localization algorithms are mainly used for accurate and safe operation of trains. In order to achieve the positioning function, some of them add a relay device in the train to avoid the penetration loss of carriage, and some of them deploy navigation equipment and a large number of positioning sensors as prominent technologies to get the location information of the running train. On the other hand, some localization algorithms are mainly used in the

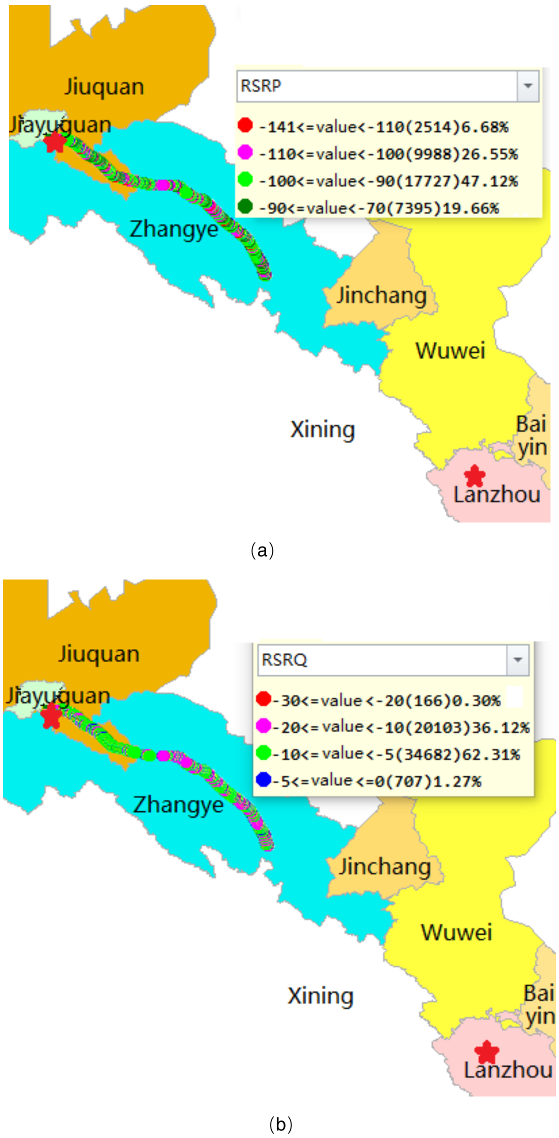


FIGURE 2. Network performance with different colors according to different grades in HSRS. (a) RSRP. (b) RSRQ.

static indoor environment. In other words, the terminal to be located is either stationary or moving at a very slow walking speed. Up to now, no researchers have been found to report a positioning algorithm for user terminals when the GPS signal is losing on HST without additional equipment. Considering this situation, this paper takes the positioning problem of a high-speed mobile terminal losing the signals of GPS as the research object to find an efficient and low-cost solution in HSRS.

In this paper, the positioning algorithm of terminal on HST is realized without modifying the test software and adding additional devices. The collection of original fingerprints with location information can be acquired through the Internet by remote control unit (RCU). It is picked up by the engineer on a train and composed of five modules, as shown in Fig. 1. Each module has a port, such as $1, 2, \ldots, 5$, to connect to other devices. Alternating current

TABLE 1. Train penetration loss.

Train level	Speed (km/h)	Material	Sealing characteristics	Average penetration loss (dB)
Local train	110	Iron, plastic	Fairly puny	6
Fast train	115	Alloy	Puny	8
Express train	150	Iron and steel alloy	Medium	12
CRH1	200	Stainless steel	Strong	24
CRH2	200	Aluminium alloy	Medium	16
CRH3	350	Aluminium alloy	Strong	20
CRH5	200	Aluminium alloy	Strong	24
CRH380A	350	Aluminium alloy	Very strong	28
CRH380B	300	Aluminium alloy	Strong	26
CRH380C	250	Aluminium alloy	Strong	24
CRH380D	250	Aluminium alloy	Strong	24
CRH6	180	Aluminium alloy	Medium	16
Fuxing train	350	Aluminium alloy	Very strong	28

(220 V) can be obtained from the train or from a portable power source. The terminals (HTC M8t), which have been installed with developed software to measure the networks of mobile communication, are connected to port 3 and port 4. The information, including control interactions and measurement interactions, is transferred between the antenna of the modem module and web server. Large quantity of measured data is stored on an FTP server. Logging into a specific website via a laptop can display some of the test results remotely, such as the path loss, the reference signal receiving power (RSRP), the reference signal receiving quality (RSRQ), the signal to interference plus noise ratio (SINR), RSSI, and other measured data.

Based on the analysis of the reasons for losing the signals of GPS, a novel model of positioning algorithm is proposed based on the fingerprint database by twice-fuzzy clustering. The locations of the high-speed mobile terminal are determined by removing abnormal fingerprints, calculating similar fuzzy clustering fingerprints and constructing a dissimilarity matrix. The research results provide robust technical support for the analysis of measured data and the optimization of wireless network. The rest of this paper is organized as follows. In Section II, the phenomenon of the loss of GPS signals in HSRS is presented through the measured data, and the causes are also analyzed. Section III elaborates the model of the positioning algorithm. Section IV presents the proposed algorithm for the fingerprint database by twice-fuzzy clustering in detail. The experimental results are presented to verify the effectiveness of the proposed algorithm in Section V. Finally, Section VI summarizes the whole article.

II. LOSS OF GPS SIGNALS IN HSRS

The frequent loss of GPS signals in HSRS causes the network performance indexes, such as the physical cell identifier (PCI), RSRP and RSRQ, to not be completely visualized on a map. Taking the section of a high-speed railway from Lanzhou station to Jiayuguan station in China as an example, the values of RSRP and RSRQ are colored by randomly measured data, and over half of the mileage had no location information, as shown in Fig. 2. In the legend of Fig. 2 (a), the unit of RSRP value is dBm. There are 2,514 sampling points, accounting for 6.68% of the sampling points, and their RSRP values range from −141 dBm to −110 dBm. They are marked red. In the legend of Fig. 2 (b), the unit of RSRQ value is dB. There are 166 sampling points, accounting for 0.3% of the sampling points, and their RSRQ values range from −30 dB to −20 dB. They are marked red. Similarly, the network performance values for the other grades are marked purple, green, and blue, respectively.

The phenomenon of the loss of GPS signals is closely related to the transmission loss of satellite signals, the speed of HST, the material used to construct the carriage of HST and the sealing characteristics. Through the analysis after collecting a large quantity of measured data, the average penetration losses under different train levels of running, different train materials and other factors are obtained, as listed in Table 1. The average penetration loss of an ordinary train is the lowest, and is approximately 6 dB. The average penetration loss of China railway high-speed (CRH) trains ranges from 16-28 dB, while the average penetration loss of the Fuxing bullet train is the greatest, as high as to 28 dB.

When HST runs within the coverage area of a cell, changes in the angle (called the incidence angle) between the eNB signal and the carriage cause different penetration losses, as illustrated in Fig. 3 (a) and (b): α and β correspond to the different angles of incidence. The smaller values of the incidence angle mean greater penetration losses. According to the wireless link budget theory, train penetration losses reduce the effective coverage of eNBs.

From the perspective of LTE network coverage optimization of telecom operators, in order to obtain better network coverage, the incidence angle must be increased through reducing the eNBs' spacing. Because when the materials of HST are the same, the larger angles of incidence are designed, the smaller the penetration loss is. From the above analysis, it can be observed that the loss of GPS signals in HSRS is quite serious and inevitable.

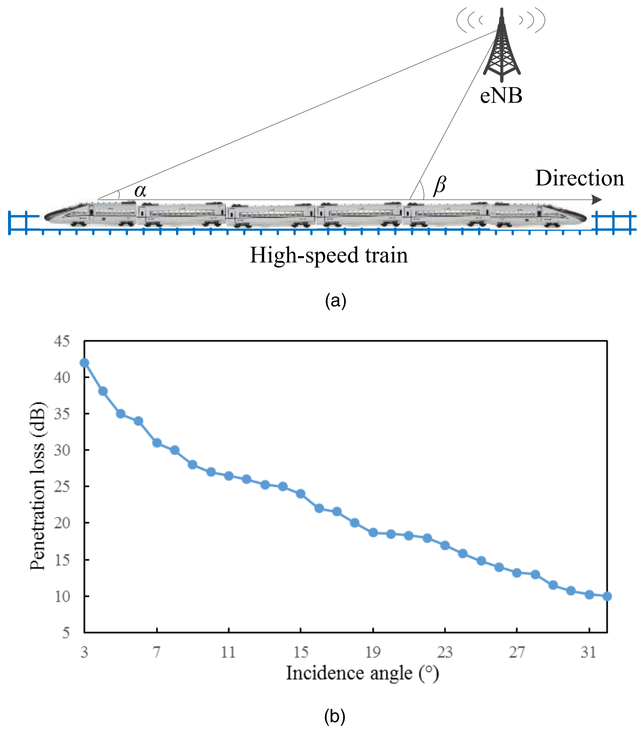


FIGURE 3. Incidence angle of the wireless signal and penetration loss of the train carriage. (a) Sketch of wireless signal the incidence angle. (b) Relationship between the incidence angle and penetration loss.

III. MODEL OF A POSITIONING ALGORITHM

In HSRS, eNBs are sparsely distributed in the plains and open sections, while relatively dense in special sections such as tunnels, ditches, viaducts and mountains. Normally, eNBs are built alternately on either side of the track. However, for non-linear track section, the eNB is usually placed on the insides of curved track to ensure the balance of the network coverage. As the train moves at high speed, the measured data of terminal changes continuously.

According to the positioning algorithm, the sampled data of one serving cell (SCell) and three optimal neighboring cells (NCell) are taken as original fingerprints. The fingerprint database consists of a large number of original fingerprints. The model of positioning algorithm based on the fingerprint database by twice-fuzzy clustering is established as shown in Fig. 4. According to the engineering parameters, the sampling data of attached location information for the SCell and NCell are assigned to the corresponding eNB in the ith sampling period ($i = 1, 2, \ldots, n$). Some sample points may exceed the theoretical threshold in the preliminarily original fingerprint database, which can be viewed as abnormal fingerprints and eliminated. The high-precision map of the high-speed railway is divided into grids with areas of 1×1.435 m^2 (1.435 m is the international standard gauge), and each grid contains measured data in the fingerprint database. When the signals of GPS are losing in HSRS, the integral value (C) is calculated according to the similarity of characteristics between measured data to be located and fingerprints. Then the fingerprints of the same type are filtered when the C of them are greater than the integral threshold value (λ). The process of calculating the location coordinates can be divided into two fundamental steps: the weighted measurement eigenvalue and construction of a dissimilarity matrix.

The C values of all fingerprints in the database are calculated according to the PCI matching degree. One of the design principles of the positioning algorithm is that the higher the PCI matching degree, the larger the C value. In addition, SCell and NCells are given different weights when calculating the C value. The essence of the integral threshold λ is to select the similar fingerprints after once-fuzzy cluster. The reasonable value of λ makes it easier for the similar fingerprints to be the result of once-fuzzy clustering.

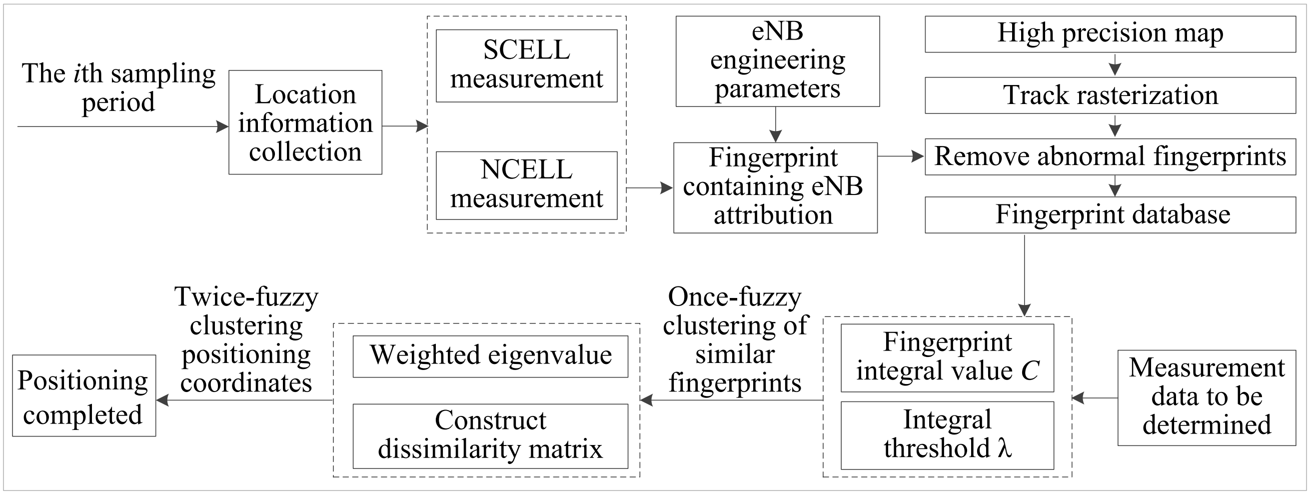


FIGURE 4. Model of positioning algorithm based on the fingerprint database by twice-fuzzy clustering.

In twice-fuzzy clustering, the measurement eigenvalues of fingerprints include the RSRP and RSRQ values of SCell and NCells. The differences of weighted eigenvalues are calculated between measurement sampling points to be located and similar fingerprints. First, the measured data of SCell and NCells are weighted in the process of calculating the C value. Then, the difference of between measured data to be located and fingerprints is calculated according to different weights.

The similarity of fingerprints can be quantified by constructing the dissimilarity matrix. On the one hand, dissimilarity matrix is the premise of positioning terminal coordinates by twice-fuzzy clustering. On the other hand, matrix computing improves the ability of data processing.

As shown in Fig. 5, a database for fingerprint positioning can be formed by the collection of measured data from the RCU. When the terminal loses the signals of GPS in tunnels, ditches or other special terrain, the characteristic values of the real-time measured data. We take eNB5 as an example. (RSRP_{i5}, RSRQ_{i5}) is a pair of measurement results of eNB 5 at data sequence $i, i = 1, 2, \ldots, N$. The final positioning is determined by the positioning algorithm based on the fingerprint database of twice-fuzzy clustering.

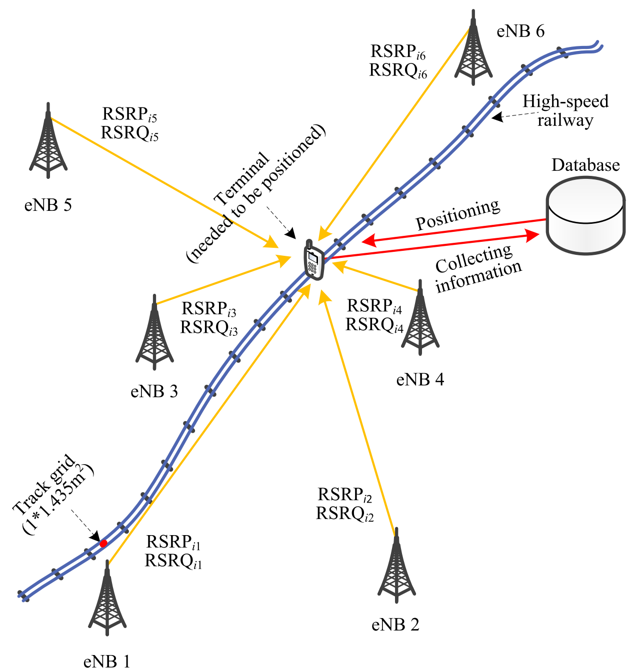


FIGURE 5. Sketch map of fingerprint-based positioning.

IV. POSITIONING ALGORITHM FOR THE TWICE-FUZZY CLUSTERING FINGERPRINT DATABASE

A. BUILDING A VALID FINGERPRINT DATABASE

Because the radio signal is affected by distance, terrain and obstacles, the signal characteristics caused by multipath fading have strong correlations with their location information [28], [29]. In each grid, the SCell and the 3 optimal measured data of NCells are selected to obtain 4 groups of PCI, RSRP and RSRQ sampling values, namely, ($\mathrm{PCI}_{i,m}$, $\mathrm{RSRP}_{i,m}$, $\mathrm{RSRQ}_{i,m}$), which constitute a fingerprint, $m = 1, 2, 3, 4$. Each grid contains at least one fingerprint, which has the location information.

When some segments in HSRS are covered by less than 3 NCells, there are losing elements in fingerprints. This is also characteristic information of the fingerprint database. This situation mainly occurs when HST pass through multiple eNBs in a short time, which causes frequent cell handover. To solve this problem, several adjacent physical cells are merged into one logical cell, so that there are no handover phenomena in one logical cell. Therefore, in HSRS there is a case where the same PCI represents multiple eNBs in the original fingerprint database, and its location coordinate set is recorded as $\{(X_l, Y_l)\}, l = 1, 2, \ldots, N$. Assuming that the location coordinate set of the terminal is recorded as (x_k, y_k),

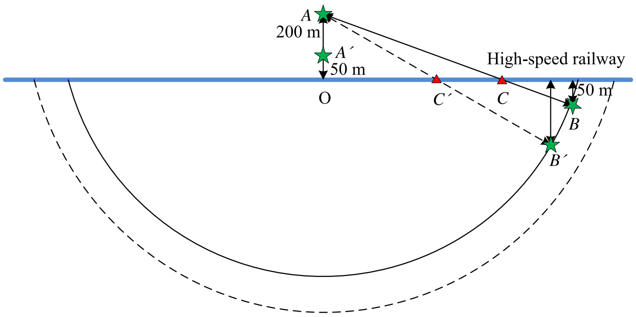


FIGURE 6. Distance between the terminal and eNBs (*A*, *A*', *B* and *B*' are respectively the locations of eNBs).

$k = 1, 2, \ldots, N$. The coordinate (X_j, Y_j) of the nearest eNB satisfies the following equation:

$$\sqrt{\left(X_j - x_k\right)^2 + \left(Y_j - y_k\right)^2} = \min_{1 \le j \le l} \left\{ \sqrt{(X_l - x_k)^2 + (Y_l - y_k)^2} \right\} \tag{1}$$

Due to the interference of network, the malfunction of eNB or the deviation of GPS, abnormal fingerprints, which affect the accuracy of fuzzy clustering, need to be eliminated from the original fingerprint database [30]. The validity of a fingerprint can be judged by the distance d, between the terminal and the nearest eNB, which can be calculated as follows:

$$d = \sqrt{\left(X_j - x_k\right)^2 + \left(Y_j - y_k\right)^2} \tag{2}$$

In HSRS, the distance between two adjacent eNBs ranges from 800 m to1,600 m, and the vertical distance between an eNB and the track ranges from 50 m to 200 m, as shown in Fig. 6. Suppose that point A and point A' represent the locations at the farthest (200 m) and closest (50 m) vertical distance between the eNB and track respectively. Draw an arc with point A as the center and 1,600 m as the radius. According to the layout principle of eNBs in HSRS, the closest point B and the farthest point B' perpendicular to the track are marked on the arc. AB and AB' intersect respectively the track at point C and point C'. When the distance d from the terminal to the nearest eNB satisfies equation (3), the corresponding sample is determined to be an effective fingerprint:

$$d_{\min} \le d \le d_{\max} \tag{3}$$

where, $d_{\min}$ and $d_{\max}$ are the nearest and farthest distances between the terminal and the adjacent eNB respectively. According to the calculation, $d_{\min} = OA' = 50$m, $d_{\max} = AC = 1{,}280$ m.

B. FUZZY CLUSTERING

1) ONCE-FUZZY CLUSTERING OF SIMILAR FINGERPRINTS

Assuming that the location information of the terminal is absent from measured data, the PCIs of SCell and NCell are expressed as PCI_{i1}', PCI_{i2}', …, PCI_{im}'. The calculation process for once-fuzzy clustering of similar fingerprints is as follows.

a: MATCHING DEGREE

Match the sampling value PCI_{i1}', PCI_{i2}', …, PCI_{im}' with the sampling value PCI_{i1}, PCI_{i2}, …, PCI_{im} in the fingerprint database. Define the variable g, which starts at 0. If a match occurs once, $g = g+1$, and so on; If an inequality occurs, the process of matching terminates. The larger the g value is, the higher the matching degree between the measured data and the fingerprints in the database, and vice versa.

b: INTEGRAL VALUE C

To treat the fingerprint with a higher matching degree in the previous step as a similar fingerprint by using a fuzzy clustering, an integral counter is set for sample ω_i, whose value is expressed as C, and the initial value is 0. If $g = 1$, $C = a$, a is assigned to C when the PCI_{i1} value of SCell matches. If $g >1$, for each match of PCI_{i2}, PCI_{i3}, …, PCI_{im} value, the value of b is assigned to C. That is:

$$C = a + b(g-1) \tag{4}$$

where, a and b, respectively, represent the cumulative values of the integral counter when the SCell and NCell PCI values match. Assume: a is a natural number, and $1 \le a \le 10$. By studying the relationships between C and a, considering the weight influences of the SCell and NCell in the process of evaluating the quality of network, and training lots of fingerprints, equation (5) is obtained:

$$b = \left\lceil \frac{10-a}{3} \right\rceil \tag{5}$$

Combining equations (4) and (5), the expression for C can be obtained as follows:

$$C = a + (g-1)\left\lceil \frac{10-a}{3} \right\rceil \tag{6}$$

c: COUNTER THRESHOLD

Set the threshold value λ of the counter, and the samples of $C > \lambda$ in the fingerprint database are similar fingerprints after once fuzzy clustering. In the section of high-speed railway with insufficient overlapping coverage (fewer than 3 NCells), the absent part of the elements in PCI_{i1}', PCI_{i2}', …, PCI_{im}', and the number of absent elements is expressed as t, namely, $0 \le t \le 3$. Referring to the expression principle for the C value in equation (6), the expression for λ is obtained:

$$\lambda = a + (3-t)\left\lceil \frac{10-a}{3} \right\rceil \tag{7}$$

2) TWICE-FUZZY CLUSTERING OF POSITIONING COORDINATES

Assuming that the RSRP of the SCell and NCell of measured data are RSRP_{i1}', RSRP_{i2}', …, RSRP_{im}', they are called p_{im}' for brevity, and RSRQ is RSRQ_{i1}', RSRQ$_{i2}$', …, RSRQ$_{im}$', or q_{im}' for brevity. The distance between these and the similar fingerprints p_{im} and q_{im} by once-fuzzy clustering are defined as:

$$d(p_i) = \alpha \times |p_{i1} - p_{i1'}| + (1-\alpha)\sum_{k=2}^{m} |p_{ik} - p_{ik'}| \tag{8}$$

$$d(q_i) = \alpha \times |q_{i1} - q_{i1'}| + (1-\alpha)\sum_{k=2}^{m} |q_{ik} - q_{ik'}| \tag{9}$$

Let $\alpha \in [0, 1]$ be the weighting coefficients of the measurement results of the SCell and NCell. Let $\beta \in [0, 1]$ be taken as the weighting coefficient of $d(p_i)$ and $d(q_i)$ for equation (6) and equation (7), that is, the weighting coefficients of the RSRP and RSRQ distances:

$$D = \begin{pmatrix} d_1 \\ d_2 \\ \cdots \\ d_n \end{pmatrix} = \begin{pmatrix} \beta d(p_1) + (1-\beta)d(q_1) \\ \beta d(p_2) + (1-\beta)d(q_2) \\ \cdots \\ \beta d(p_n) + (1-\beta)d(q_n) \end{pmatrix} \tag{10}$$

where, n represents the number of similar fingerprints by once-fuzzy clustering. The difference analysis method is used to determine the dissimilarity r_{uv} of d_u and d_v in equation (10) as follows:

$$r_{uv} = |d_u - d_v| \tag{11}$$

where, $1 \le u \le n$, and $1 \le v \le n$. The smaller the r_{uv} value is, the smaller the dissimilarity of the two elements. This means the higher similarity, and vice versa. Thus, the dissimilarity matrix R is constructed as follows:

$$M = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{pmatrix} \tag{12}$$

where, m represents the number of similar fingerprints by twice-fuzzy clustering. M is a symmetric matrix, that is, $r_{uv} = r_{vu}$, and $r_{uu} = 0$. The threshold value γ can be set according to the specific terrain of the high-speed railway. After twice-fuzzy clustering, fingerprints of the same type meet the following conditions:

$$r_{uv} \le \gamma \tag{13}$$

There are N_f fingerprints that satisfy the requirements of equation (13). According to equation (14), the average value of N_f fingerprints after twice-fuzzy clustering is calculated as the positioning result for the terminal as follows:

$$(x, y) = \frac{1}{N_f}\sum_{k=1}^{N_f}(x_k, y_k) \tag{14}$$

V. POSITIONING RESULTS

A. QUANTITATIVE ANALYSIS OF FUZZY CLUSTERING FINGERPRINTS

In once-fuzzy clustering of similar fingerprints, the value of t reflects the fingerprint feature dimension. This term shows

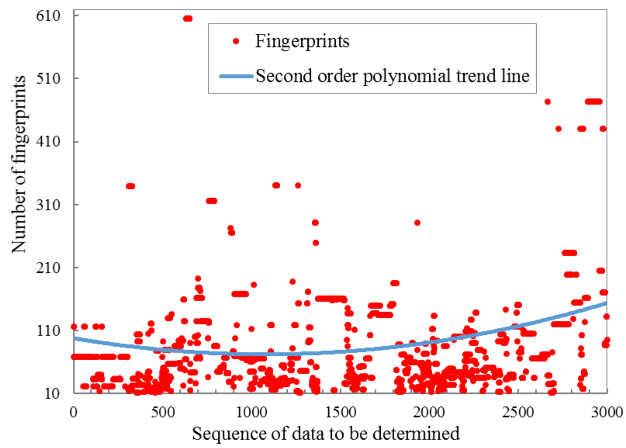


FIGURE 7. Once-fuzzy clustering of similar fingerprints.

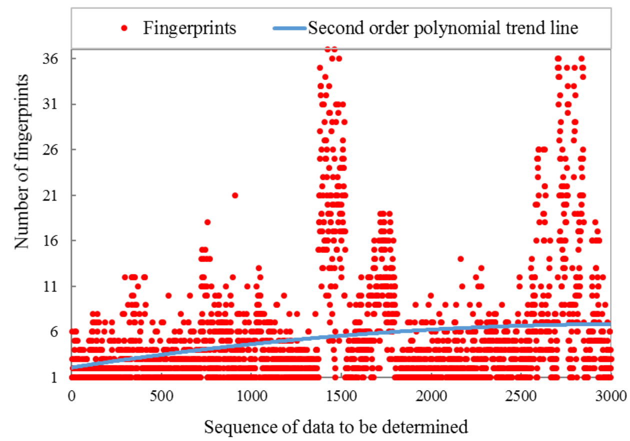


FIGURE 8. Twice-fuzzy clustering of fingerprints for locating coordinates.

the number of absent elements. If the C value is greater than the λ value, then the same type of fingerprint identification condition is experienced. When $a = 5$, the relationship between the fingerprint characteristics and the fingerprint identification listed in Table 2 indicates PCI numerical matching: "0" indicates no matching, "-" indicates elements losing, "$\surd$" indicates similar fingerprints, and "$\times$" indicates non-similar fingerprints.

Through the correlation of multi-feature of fingerprints, the range of similar fingerprint can be optimized, and fingerprints with greater similarities can be selected for fuzzy clustering. The measured data without the signals of GPS over one hour are randomly selected for positioning, and the number of similar fingerprints in once-fuzzy clustering is shown in Fig. 7. The number of fingerprints to be located in the fingerprint database is between 11-606. According to the second-order polynomial trend line, the average number of similar fingerprints is approximately 94. The proportion of similar fingerprints between 15 and 200 in once fuzzy clustering is as high as 90.81%.

TABLE 2. Relationship between fingerprint characteristics and fingerprint identification.

Category	Fingerprint features	Fingerprint identification															
1	$t = 3$	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	
	C	5	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	λ	3.25	3.25	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Is it similar	√	×	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	$t = 2$	11	10	01	00	-	-	-	-	-	-	-	-	-	-	-	-
	C	7	5	2	0	-	-	-	-	-	-	-	-	-	-	-	-
	λ	4.55	4.55	4.55	4.55	-	-	-	-	-	-	-	-	-	-	-	-
	Is it similar	√	√	×	×	-	-	-	-	-	-	-	-	-	-	-	-
3	$t = 1$	111	110	101	100	011	010	001	000	-	-	-	-	-	-	-	-
	C	9	7	7	5	4	2	2	0	-	-	-	-	-	-	-	-
	λ	5.58	5.58	5.58	5.58	5.58	5.58	5.58	5.58	-	-	-	-	-	-	-	-
	Is it similar	√	√	√	×	×	×	×	×	-	-	-	-	-	-	-	-
4	$t = 0$	1111	1110	1101	1011	1100	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001	0000
	C	11	9	9	9	7	7	7	5	6	4	4	2	4	2	2	0
	λ	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15	7.15
	Is it similar	√	√	√	√	×	×	×	×	×	×	×	×	×	×	×	×

TABLE 3. Statistics of the loss of GPS signals.

Starting station	Destination station	Number of RSRP	Number with GPS	Number without GPS	GPS loss rate (%)
Linze south	Gaotai south	873	646	227	26.00
Gaotai south	Qingshui north	916	699	217	23.69

The number of similar fingerprints by twice-fuzzy clustering is shown in Fig. 8. The number of cluster fingerprints for the data to be located ranges from 1 to 37. According to the second-order polynomial trend line, the average number of similar fingerprints in database remains approximately 6. The number of over 95.57% of the similar fingerprints is between 1-20 by twice-fuzzy clustering.

B. MEASUREMENT VERIFICATION

To verify the positioning algorithm based on the twice-fuzzy clustering fingerprint database proposed in this paper, a field test is carried out between the Linze South station and the Qingshui North station of the Lanzhou-Xinjiang high-speed railway. The effective test range is 95 km, and the disconnection of GPS data covered a distance of 48 km. The statistics of the losing network sampling points are listed in Table 3.

Simulation of the positioning algorithm is conducted by MATLAB. In Fig. 9, there are eleven segments (because the area of the image has been scaled, there are three segments joined together) of track with different lengths that lose the signals of GPS. Fortunately, the position of the terminal is calculated by the proposed positioning algorithm. Each section of the positioning curve can be smoothly connected with the GPS curve. Overall, the effect of position information can meet the requirement that the average positioning accuracy is less than 10 m).

The results of location information at different GPS miss rates are shown in Fig. 10. The red mark is the GPS test track,

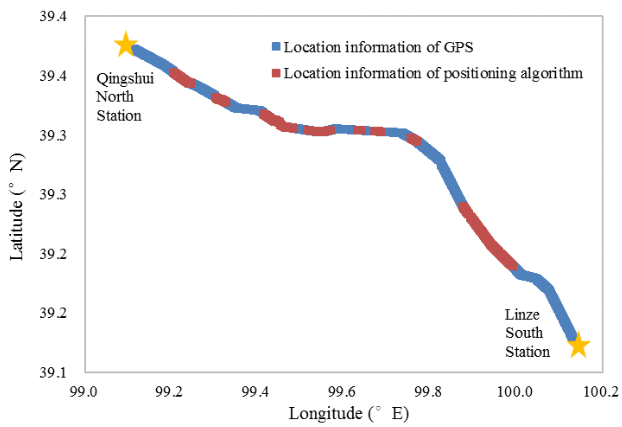
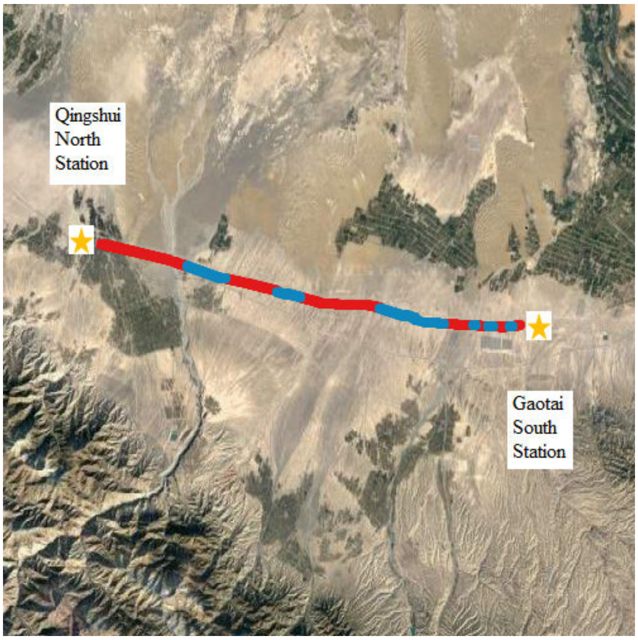


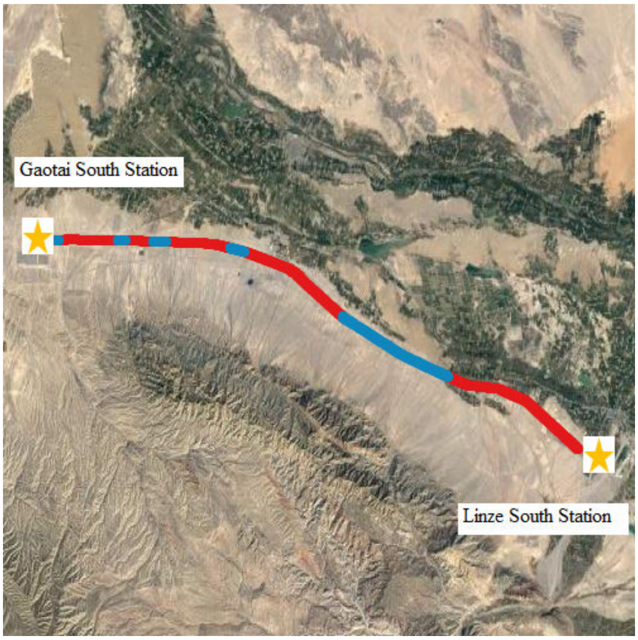
FIGURE 9. Compensation results of location information by the positioning algorithm.

and the blue mark is the compensation track of the proposed positioning algorithm. In Fig. 10(a), the GPS miss rate is 26.00%, and the terrain of most test sections without GPS location information is mainly plains. In Fig. 10(b), the loss rate of the signals of GPS is 23.69%. Because the speed of HST in the blue section is higher than that of the red section, the test track is longer when the number of sampling points is relatively small.

The positioning algorithm based on the twice-fuzzy clustering fingerprint database is compared with GPS for RSRP network coverage, as shown in Fig. 11. There are 5 colors that correspond to the RSRP ranges: blue-purple, grass-green, water-blue, yellow and red, which represent the network coverage qualities of the corresponding positions as excellent (−90dBm ∼63dBm), good (−95dBm ∼ −90dBm), general (−105dBm ∼ −95dBm), medium (−110dBm ∼ −105dBm) and poor (−121.12dBm ∼ −110dBm) respectively.

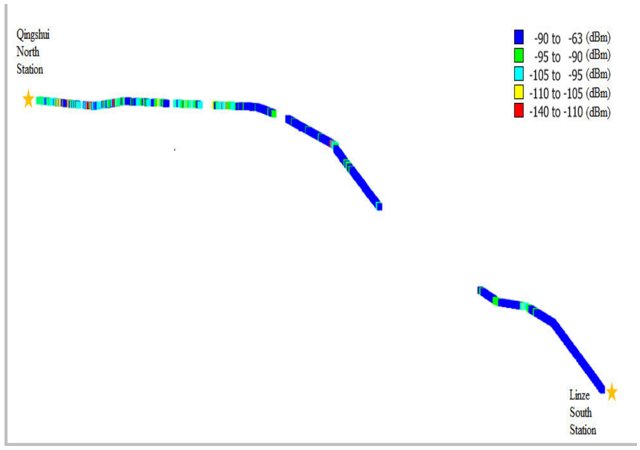


(a)

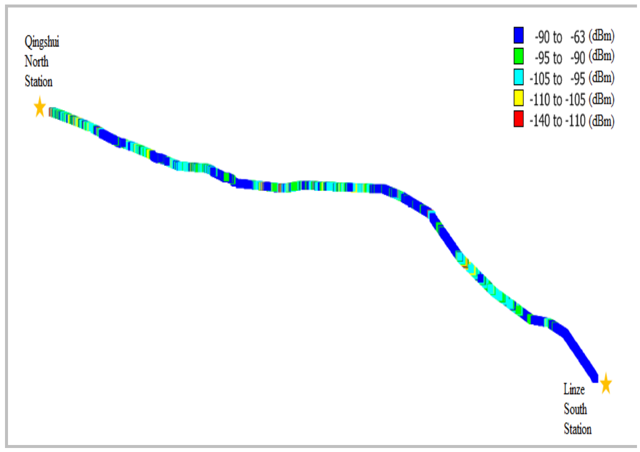


(b)

FIGURE 10. Compensation results of location information under different GPS miss rates. (a) Linze South station - Gaotai South station. (b) Gaotai South station - Qingshui North station.



(a)



(b)

FIGURE 11. Comparison of the positioning algorithm and GPS for RSRP network coverage. (a) GPS of application. (b) positioning algorithm based on the twice-fuzzy clustering fingerprint database of the application.

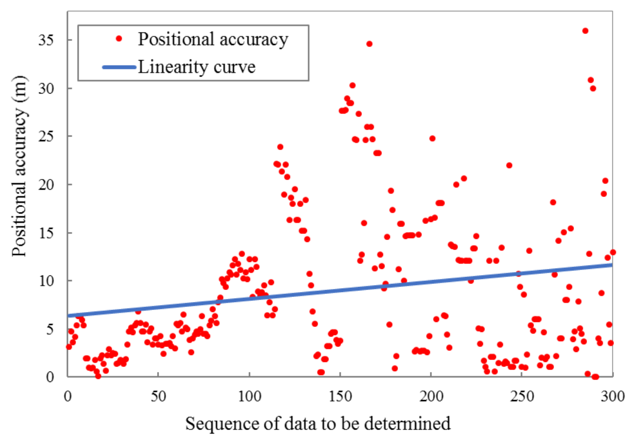


FIGURE 12. Accuracy of the positioning algorithm.

C. ACCURACY OF THE POSITIONING ALGORITHM

To evaluate the accuracy of the positioning algorithm, 300 consecutive data sequences are randomly selected for positioning. The positioning accuracy can be measured by using the mean value of *DIST* in equation (15), which is defined as the difference between the longitude and latitude (θ_c, ϕ_c) calculated by the proposed positioning algorithm and the longitude and latitude (θ_z, ϕ_z) collected by GPS. Its expression is as follows:

$$DIST = R \times \arccos\,(\sin\varphi_c \sin\varphi_z + \cos\varphi_c \cos\varphi_z \cos(\theta_c - \theta_z)) \times \pi/180 \tag{15}$$

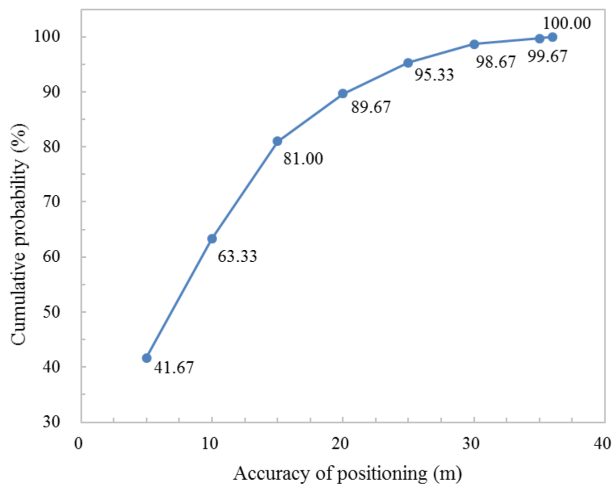


FIGURE 13. Accumulation probabilities of the positioning algorithm' s accuracy.

where, $c = 1, 2, \ldots, N$, and $z = 1, 2, \ldots, N$. R is the radius of the earth and the longitude and latitude are expressed in radians. As shown in Fig. 12, the average positioning accuracy of the positioning algorithm is 9.02 m. After comparing and analyzing the number of similar fingerprints with large error and small error respectively after clustering, it is concluded that the greater the positioning error, the fewer the number of similar fingerprints.

Both the distribution of positioning accuracy and the accumulation probabilities of the positioning algorithm are shown in Fig. 13. The proportion of samples with a positioning accuracy of less than 5 m is the highest, as high as 41.67%. The proportion of samples with a positioning accuracy of less than 10 m is 63.33%.

VI. CONCLUSION

Based on the fingerprint database constructed from over 300,000 samples, the measured data and localization results are analyzed. The signals of GPS are affected by the penetration loss and the incidence angle, and the degree of intensity attenuation is directly related to the train grade and train running direction. On the basis of a high-precision map, eNB engineering parameters and massive quantities of measured data, the proposed positioning algorithm model can clearly describe the positioning algorithm by the twice-fuzzy clustering fingerprint database. For the case of different loss rates of the signals of GPS, the longitude and latitude coordinates of the terminal can be supplied through the positioning algorithm based on the fingerprint database by twice-fuzzy clustering along the high-speed railway. The terminal location information with high accuracy by the positioning algorithm proposed can be applied to the association analysis of network quality. When the signals of GPS are losing in the measured data, positioning results can provide a reference for network coverage evaluation, base station planning and the network optimization for HSRS.

ACKNOWLEDGMENT

The authors are very grateful to the China Mobile Communications Group Gansu Company, Ltd. for the large amount of raw measured data.

REFERENCES

[1] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); UE Conformance Test Aspects-Performance Enhancements for High Speed Scenario in LTE (Release 14)*, document TR 36.878, 3GPP, 2017.

[2] C. Jian, Y. H. Zhou, X. Y. Zhou, S. Y. Wang, Z. G. Zhang, and Z. H. Liu, "Analysis on shock effect of China's high-speed railway on aviation transport," *Transp. Res. Pt. A-Policy Pract.*, vol. 108, pp. 35–44, Feb. 2018.

[3] F. Zhen, X. Cao, and J. Tang, "The role of access and egress in passenger overall satisfaction with high speed rail," *Transportation*, vol. 46, no. 6, pp. 2137–2150, Dec. 2019.

[4] Z. Hou, Y. Zhou, L. Tian, J. Shi, Y. Li, and B. Vucetic, "Radio environment map-aided Doppler shift estimation in LTE railway," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4462–4467, May 2017.

[5] L. Azpilicueta, J. J. Astrain, P. Lopez-Iturri, F. Granda, C. Vargas-Rosales, J. Villadangos, A. Perallos, A. Bahillo, and F. Falcone, "Optimization and design of wireless systems for the implementation of context aware scenarios in railway passenger vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2838–2850, Oct. 2017.

[6] B. Ai, R. He, G. Li, K. Guan, D. He, G. Shi, and Z. Zhong, "Determination of cell coverage area and its applications in high-speed railway environments," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 3515–3525, May 2017.

[7] A. Fedrizzi, R. Ursin, T. Herbst, M. Nespoli, R. Prevedel, T. Scheidl, F. Tiefenbacher, T. Jennewein, and A. Zeilinger, "High-fidelity transmission of entanglement over a high-loss free-space channel," *Nature Phys.*, vol. 5, no. 6, pp. 389–392, Jun. 2009.

[8] K. Haddadi and T. Lasri, "Geometrical optics-based model for dielectric constant and loss tangent free-space measurement," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 7, pp. 1818–1823, Jul. 2014.

[9] Q. Zhang, H. Sun, Z. Feng, H. Gao, and W. Li, "Data-aided Doppler frequency shift estimation and compensation for UAVs," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 400–415, Jan. 2020.

[10] C. Zhang, G. Wang, M. Jia, R. He, L. Zhou, and B. Ai, "Doppler shift estimation for millimeter-wave communication systems on high-speed railways," *IEEE Access*, vol. 7, pp. 40454–40462, 2019.

[11] M. Hayashi, Y. Tsunemitsu, and T. Maeyama, "Analysis of penetration loss of ultra high frequency band radio waves on trains," *IEICE Electron. Express*, vol. 12, no. 6, 2015, Art. no. 20150142.

[12] R. Raheem, A. Lasebae, M. Aiash, J. Loo, and R. H. Colson, "Mobile femtocell utilisation in LTE vehicular environment: Vehicular penetration loss elimination and performance enhancement," *Veh. Commun.*, vol. 9, pp. 31–42, Jul. 2017.

[13] L. Zhang, J. Rodríguez-Piñeiro, J. R. O. Fernández, J. A. García-Naya, D. W. Matolak, C. Briso, and L. Castedo, "Propagation modeling for outdoor-to-indoor and indoor-to-indoor wireless links in high-speed train," *Measurement*, vol. 110, pp. 43–52, Nov. 2017.

[14] R. Cheng, Y. Song, D. Chen, and X. Ma, "Intelligent positioning approach for high speed trains based on ant colony optimization and machine learning algorithms," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3737–3746, Oct. 2019.

[15] W. Wang, Y. Tian, and C.-J. Ahn, "Blind symbol timing synchronization scheme for real-time industrial wireless control network using high-speed preambleless OFDM systems over multipath fading channels," *Digit. Signal Process.*, vol. 62, pp. 30–37, Mar. 2017.

[16] B. Wang, S. Zhou, W. Liu, and Y. Mo, "Indoor localization based on curve fitting and location search using received signal strength," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 572–582, Jan. 2015.

[17] J. Yim, "Introducing a decision tree-based indoor positioning technique," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 1296–1302, Feb. 2008.

[18] Y.-W. Ma, J.-L. Chen, F.-S. Chang, and C.-L. Tang, "Novel fingerprinting mechanisms for indoor positioning," *Int. J. Commun. Syst.*, vol. 29, no. 3, pp. 638–656, Feb. 2016.

[19] S. Sorour, Y. Lostanlen, S. Valaee, and K. Majeed, "Joint indoor localization and radio map construction with limited deployment load," *IEEE Trans. Mobile Comput.*, vol. 14, no. 5, pp. 1031–1043, May 2015.

[20] A. K. M. M. Hossain and W.-S. Soh, "A survey of calibration-free indoor positioning systems," *Comput. Commun.*, vol. 66, pp. 1–13, Jul. 2015.

[21] X. Fang, L. Nan, Z. Jiang, and L. Chen, "Fingerprint localisation algorithm for noisy wireless sensor network based on multi-objective evolutionary model," *IET Commun.*, vol. 11, no. 8, pp. 1297–1304, Jun. 2017.

[22] Y. Chen, T. Du, S. Sun, and C. Jiang, "Indoor location of interference source with fuzzy c clustering cruising AP based on double fingerprint database," *J. Intell. Fuzzy Syst.*, vol. 37, no. 1, pp. 315–327, Jul. 2019.

[23] J. Zhang, G. Han, N. Sun, and L. Shu, "Path-loss-based fingerprint localization approach for location-based services in indoor environments," *IEEE Access*, vol. 5, pp. 13756–13769, 2017.

[24] J. Bi, Y. Wang, X. Li, H. Qi, H. Cao, and S. Xu, "An adaptive weighted KNN positioning method based on omnidirectional fingerprint database and twice affinity propagation clustering," *Sensors*, vol. 18, no. 8, pp. 2502–2518, 2018.

[25] C. Li, J. Xie, W. Wu, H. Tian, and Y. Liang, "Monte Carlo localization algorithm based on particle swarm optimization," *Automatika*, vol. 60, no. 4, pp. 451–461, Oct. 2019.

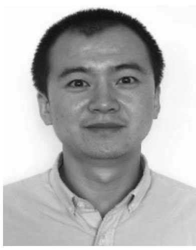
[26] M. Zhou, Y. Wei, Z. Tian, X. Yang, and L. Li, "Achieving cost-efficient indoor fingerprint localization on WLAN platform: A hypothetical test approach," *IEEE Access*, vol. 5, pp. 15865–15874, 2017.

[27] Z. Gu, Z. Chen, Y. Zhang, Y. Zhu, M. Lu, and A. Chen, "Reducing fingerprint collection for indoor localization," *Comput. Commun.*, vol. 83, pp. 56–63, Jun. 2016.

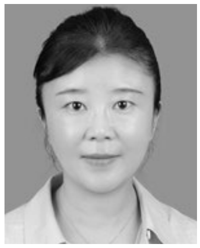
[28] T. M. Nguyen, K. Seung-Ho, and J. Y. Chung, "Characterization of multiple-input-multiple-output antennas under different multipath environments in a reverberation chamber," *Microw. Opt. Technol. Lett.*, vol. 60, no. 10, pp. 2437–2442, Sep. 2018.

[29] Y. Wu, Y. S. Zhu, S. H. Leung, W. K. Wong, and T. Wang, "Low-complexity code tracking loop with multipath diversity for GNSS over multipath fading channels," *Acta Astronautica*, vol. 115, pp. 367–375, Oct. 2015.

[30] J. Luo, X. Yin, Y. Zheng, and C. Wang, "Secure indoor localization based on extracting trusted fingerprint," *Sensors*, vol. 18, no. 2, pp. 469–491, 2018.



BAOFENG DUAN was born in Wugong, Shaanxi, China, in 1984. He received the master's degree in communication and information system from Lanzhou University, China, in 2010. He is currently pursuing the Ph.D. degree in major of railway transportation and communication engineering with Lanzhou Jiaotong University, China. He was certified as an Assistant Engineer and an Engineer, in 2013 and 2017, respectively. His research interests include railway wireless communications, wireless network optimization for high-speed rail scenarios, and positioning algorithms for high-speed mobile terminals. He won the China National Scholarship in 2018. His honor includes the Second Prize for the Science and Technology Progress Award from the Gansu Electronics Society of China.



CUIRAN LI (Member, IEEE) was born in Shanxi, China, in 1975. She received the bachelor's and master's degrees from the Department of Communication Engineering, Lanzhou Jiaotong University, in 1996 and 1999, respectively, and the Ph.D. degree from the School of Electronics and Information Engineering, Beijing Jiaotong University, in 2003. She is currently working as a Professor with Lanzhou Jiaotong University. She has published a book and over 40 scientific articles in her research area till now. Her research interests include railway mobile communication, wireless sensor networks, and cooperative communication technology.

JIANLI XIE was born in Gansu, China, in 1972. He received the B.Sc. degree from Sichuan University, in 1993, and the master's and Ph.D. degrees from the School of Electronics and Information Engineering, Lanzhou Jiaotong University, in 1999 and 2014, respectively. From 1993 to 1996, he worked as an Engineer with Lanzhou General Machinery Plant. From 1999 to 2007, he was a Senior Hardware Engineer with Beijing Great Dragon Company. He is currently working as a Professor with Lanzhou Jiaotong University. He has published over 20 scientific articles in his research area till now. His research interests include railway wireless communication and cognitive radio techniques.

DONGMEI ZHOU was born in Henan, China, in 1975. She received the bachelor's and master's degrees from the Department of Communication Engineering, Lanzhou Jiaotong University, in 1999 and 2010, respectively. She is currently working as an Associate Professor with Lanzhou Jiaotong University. She has published several scientific articles in her research area till now. Her research interests include mobile communication and cooperative communication technology.

WEI WU was born in Gansu, China, in 1974. He received the bachelor's and master's degrees from Lanzhou Jiaotong University, China, in 1997 and 2010, respectively, where he is currently pursuing the Ph.D. degree in major of railway transportation and communication engineering. He is currently working with Lanzhou Jiaotong University. He is also a member of a council of the Laboratory Management Branch, Chinese Association of Higher Education. His research interests include traffic information and control and GSM-R/LTE-R systems.