

# Fast Fingerprint Database Construction Method in Bluetooth Indoor Positioning System

Mu Zhou, Xiaoxiao Jin, Lingxia Li, Zengshan Tian, Haoliang Ren, Haifeng Cong  
zhoumu@cqupt.edu.cn, jinxiaoxiaosx@163.com, lilx@cqupt.edu.cn

Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts  
and Telecommunications, Chongqing 400065, China

**Abstract.** The fingerprint samples are collected point-by-point in traditional fingerprint database construction approach, which is inefficient. In response to this problem, this paper proposes a database construction system based on fast acquisition of fingerprint samples and the Radial Basis Function (RBF). Firstly, several linear paths are marked in the indoor environment according to the determined sampling interval and sampling number. Secondly, according to the Pedestrian Dead Reckoning (PDR) algorithm, the coordinates are assigned to the Received Signal Strength (RSS) to form a sparse fingerprint database. Thirdly, we select the RBF to extend sparse fingerprint database, with the purpose of constructing a database with higher fingerprint granularity. Experimental results show that our proposed approach can be applied to reduce the labor cost in database construction, as well as guaranteeing high-enough accuracy performance of fingerprint database.

**Key words:** Indoor positioning; Radial Basis Function; fingerprint database; PDR; Bluetooth

## 1 Introduction

Nowadays, more than 80% of human activities are concentrated in indoor environments, which leads to the popularity of wireless positioning technology. Although the fingerprint positioning system has been able to achieve high positioning accuracy, there are still many issues that needed to be focused [1]. Such as multi-path effect of wireless signal, the flow of people and the constant change of position of the device, which all lead to the decline of positioning accuracy. In addition, the inefficiency of traditional fingerprint database construction approach is not conducive to the development of fingerprint positioning technology.

The fingerprint positioning system includes offline and online phases [2]. In offline phase, multidirectional fingerprint samples are collected by professionals at each Reference Point (RP). In online phase, the user uploads the real-time RSS to the server, and the server searches the fingerprint database to give the locations of the user. Therefore, we can see from the above, the fingerprint samples needs to be collected point-by-point in traditional method. To solve this, we select a few linear paths and adopt dynamic walking way to collect fingerprints. Then, we review the interpolation algorithm [3] to expand the raw sparse fingerprint database. This system can not only improve the efficiency, but also ensure the validity of the fingerprint database.

The remainder of the paper is organized as follows. Section 2 reviews the approaches of constructing fingerprint database. In Section 3, the system block diagram and the content of each module is described. Section 4 shows the experimental results. Finally, the conclusion is provided in Section 5.

## 2 Related work

In recent years, experts and scholars have conducted a series of studies on how to construct fingerprint database efficiently. A crowdsourcing-based fingerprint database auto-building system is proposed in [4]. The server-side of the system modifies the user-uploaded crowdsourcing data to extract the fingerprint samples. However, the system does not filter those large error data, which is a unwise act. In [5], the compressed sensing theory is used to construct fingerprint database, including signal sparse representation, coding and reconstruction algorithm. It must be noted that, the system only works in a relatively narrow area. However, the propagation environment of signal in indoor is complicated, so it is still necessary to study the system in open space. The linear interpolation method is used for expansion in [6], but the simple linear interpolation algorithm can not well estimate the fingerprint information at RPs in complex environments. In view of the above problems, this paper proposes an efficient fingerprint database construction approach with dynamical collection approach combined with the RBF interpolation algorithm. Firstly, we collect fingerprint samples through dynamic walking way. Then, we get the coordinates of RPs by using the speed and heading angle. Next, to get a complete fingerprint database, we select the RBF interpolation. Finally, based on the Weighted K-nearest Neighbor (WKNN) algorithm, the locations of the user is outputted.

## 3 Algorithm description

### 3.1 Algorithm overview

The overall framework shown in **Figure 1**, which is composed of fast acquisition of fingerprint samples module, speed and heading calculation module, and expansion of sparse fingerprint database module. Firstly, according to the sampling interval, we mark trajectories in indoor environments and collect the fingerprint samples based on the sampling number. Secondly, through modifying the error of PDR trajectory, we update the coordinates. And then combined with RSS, we get a sparse fingerprint database. Lastly, the RBF is used to expand the sparse fingerprint database.

### 3.2 Fast acquisition of fingerprint data

The fingerprint position technology has been widely applied because of its high accuracy and no need for too many equipments. The volume of the database increases with the increase of environment area. The traditional fingerprint collection approach consumes a lot of manpower and time, that is not suitable for large-scale environments,

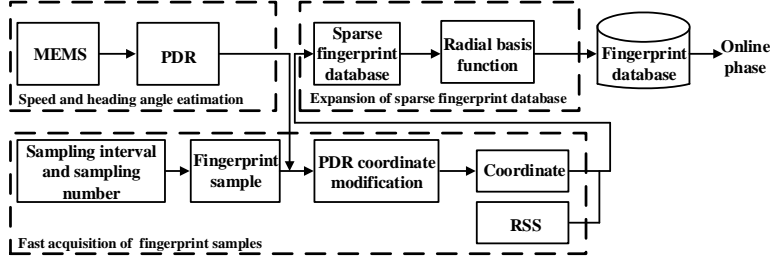


Fig. 1. Block diagram of fingerprint database construction system.

so we propose the fast acquisition of fingerprint samples method. In offline phase, the starting points of several paths are marked in the indoor environment, which cover the entire target area. Then the professional collects fingerprints on each path sequentially. Each time when we complete data collection, the professional stops moving. At the same time, the handheld terminal and the server side save the Micro Electromechanical System (MEMS) data and Bluetooth data separately. Thus each path data contains the coordinates of starting point and ending point, heading angle and speed.

This paper proposes a method for dynamic acquisition of fingerprint samples. Before this, we need to determine the locations and sampling number of each path. We study the change characteristic and stability of signal according to the real test. The number of paths should be moderate and the paths should be uniformly distributed in the indoor environment. In addition, the sampling number of each path should also be moderate. In conclusion, the fingerprint database constructed by fast acquisition approach not only saves time and cost, but also ensures positioning accuracy.

### 3.3 Speed and heading angle estimation

This system uses acceleration, magnetometer and gyroscope data of the terminal to estimate speed and heading angle. The gait detection on peak detection of acceleration. Besides, combined with the sensor data and quaternion [7], we estimate the heading angle.

**Speed estimation.** According to the accelerometer data, the gait of pedestrian can be judged [8]. In this paper, we use three-axis acceleration to calculate the average speed of pedestrian per second (referred to as second speed). In order to avoid additional error caused by the difference of equipments, a smoothing filter is first performed, the total acceleration is calculated as  $a_i^{total} = \sqrt{(a_i^x)^2 + (a_i^y)^2 + (a_i^z)^2}$ .  $(a_i^x, a_i^y, a_i^z)$  is the three-axis accelerometer data at each sampled point. From [9], the acceleration variable is sinusoidal. Therefore, we detect striding behavior by analyzing whether the total acceleration peak is greater than the threshold. The formula for velocity estimation is

$$v_k = \frac{L_k}{t_k} = \frac{L_k f_s}{\Delta N} \quad (1)$$

where the sampling frequency of the sensor is set as  $f_s$ , the step length of pedestrians at step  $k$  is  $L_k$ , the number of sampled points between the peaks is  $\Delta N$ , the time required for step  $k$  is  $t_k = \frac{\Delta N}{f_s}$ .

**Heading angle estimation.** Through the output data of sensors, we can calculate the attitude angle of the carrier. We set  $\theta, \gamma, \varphi$  as pitch angle, roll angle and heading angle respectively. This paper is based on the EKF fusion algorithm and quaternion. The gravitational acceleration measured by pedestrian at rest is  $a^b$ , the component of the known gravity acceleration in the North East Sky navigation coordinate system is  $a^n$ . According to the conversion relationship, the acceleration of gravity in the carrier coordinate system is expressed as

$$a^b = C_n^b a^n \quad (2)$$

According to the transformation matrix, the roll angle and pitch angle are obtained as follows

$$\begin{cases} \gamma = \arctan\left(\frac{a_x}{\sqrt{a_y^2 + a_z^2}}\right) \\ \theta = \arctan\left(-\frac{a_y}{a_z}\right) \end{cases} \quad (3)$$

Then we choose quaternion theory to solve the real-time attitude angle. According to the theory, the attitude angle of state equation is expressed as

$$\dot{X} = \frac{1}{2} \begin{bmatrix} 0 & -(\omega_x - w_x) & -(\omega_y - w_y) & -(\omega_z - w_z) \\ (\omega_x - w_x) & 0 & (\omega_z - w_z) & -(\omega_y - w_y) \\ (\omega_y - w_y) & -(\omega_z - w_z) & 0 & (\omega_x - w_x) \\ (\omega_z - w_z) & (\omega_y - w_y) & -(\omega_x - w_x) & 0 \end{bmatrix} X \quad (4)$$

where  $\omega_x, \omega_y, \omega_z$  are three-axis output values of gyroscopes,  $w_x, w_y, w_z$  are measurement error. The output of accelerometers and magnetometers are taken as observations. Based on EKF algorithm, the optimal four elements of state variables are calculated, and the heading angle is further obtained

$$\psi = \arctan\left(-\frac{2(q_1 q_2 + q_0 q_3)}{q_0^2 + q_1^2 - q_2^2 - q_3^2}\right) \quad (5)$$

### 3.4 Generation of sparse fingerprint database

The PDR algorithm relies on historical location information [10] to calculate current position, and there is accumulative error in it. It is necessary to modify the PDR prediction trajectory.

**Deducing the standard coordinates.** After obtaining the starting points, terminal coordinates and fingerprints of trajectories, the standard coordinates are generated

$$\begin{cases} X_i = x_0 + i * L_{step-x} \\ Y_i = y_0 + i * L_{step-y} \end{cases} \quad (i \in 1, 2, \dots, \frac{x_{end} - x_0}{L_{step-x}}) \quad (6)$$

where  $(x_0, y_0)$  and  $(x_{end}, y_{end})$  are the starting point and terminal coordinate,  $L_{step-x}$  and  $L_{step-y}$  are RP interval constant in standard fingerprint database.

**Correction of PDR trajectory.** The PDR trajectory contains cumulative error, it is necessary to add the error correction factor

$$\begin{cases} x_i = x_0 + \sum_{n=1}^i vx_n + \sum_{n=1}^i \varepsilon x_n \\ y_i = y_0 + \sum_{n=1}^i vy_n + \sum_{n=1}^i \varepsilon y_n \end{cases} \quad (7)$$

where  $(x_i, y_i)$  is the revised coordinate at  $RP_i$ ,  $vx_i$  and  $vy_i$  are the velocity components on the X and Y axis,  $\varepsilon x_i$  and  $\varepsilon y_i$  are the correction values on the X and Y axis. The correction value is

$$\begin{cases} \varepsilon x_i = \frac{vx_i}{\sum_{n=1}^{end} vx_n} (L_x - L_{xpd_r}) \\ \varepsilon y_i = \frac{vy_i}{\sum_{n=1}^{end} vy_n} (L_y - L_{ypdr}) \end{cases} \quad (8)$$

where  $L_x$  and  $L_y$  are the standard length of modified trajectory projection on the X and Y axis,  $L_{xpd_r} = \sum_{n=1}^{end} vx_n$  and  $L_{ypdr} = \sum_{n=1}^{end} vy_n$  are the PDR recurrence length of the modified trajectory on the X and Y axis.  $vx_i$  and  $vy_i$  are the projection components of the velocity on the two axes

$$\begin{cases} vx_i = v_i \sin(head_i) \\ vy_i = v_i \cos(head_i) \end{cases} \quad (9)$$

**Generation of sparse fingerprint database.** The fingerprint database coordinates are obtained from (7). Firstly, we match the estimated coordinates with the nearest coordinates in standard database. In addition, calculate the mean value of RSS at each second. The mean processing procedure for RSS is

$$RSSI_n = \frac{\sum_{i=1}^M rssi_i}{M} \quad (10)$$

where  $RSSI_n$  represents the RSS value at  $RP_n$ ,  $rssi_i$  is the RSS which matched to  $RP_n$  at the  $i^{\text{th}}$  time,  $M$  indicates the times that  $RP_n$  are stored into.

### 3.5 Expansion of sparse fingerprint database

From the last part, we get the sparse fingerprint database. However, the fingerprint granularity of the sparse database is small, we still need to expand the database. In this paper, we use the spatial correlation between sampled points to expand fingerprint database. By using the RBF, the RSS from the same anchor can be well estimated in a certain area. In this paper, we choose the RBF to estimate RSS.

**Radial basis function.** The RP that be sampled fingerprint samples is called the labeled sample. According to the RBF, the labeled sample is used as the feedback, and the point to be estimated is called non-feedback point. Delimit the cut-off zone at each non-feedback point, the schematic diagram is shown in **Figure 2**. The RBF is a monotone function, which represents the distance from any point to a certain center, where the distance refers to the Euclidean distance.

$$\varphi(x-a) = \varphi(\|x-a\|) \quad (11)$$

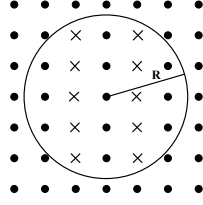


Fig. 2. The cut-off area.

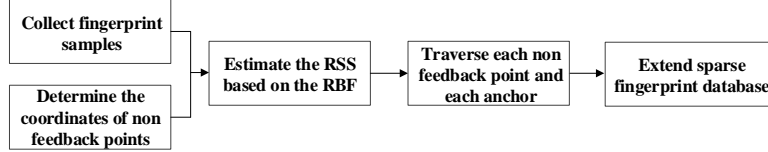


Fig. 3. The process of expanding the fingerprint database.

By constructing the RBF space, the object function can be approximated. Given function  $s(x) \in C[a, b]$ ,  $x_1, x_2, \dots, x_N \in [a, b]$ , then we look for the following form of function

$$f(x) = \sum_{i=1}^N a_i \varphi(x - x_i) \quad (12)$$

The interpolation condition is

$$Sf(x) = y \quad (13)$$

The coefficient matrix is expressed as

$$\begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix}^T \begin{pmatrix} \varphi(\|x_1 - x_1\|) \varphi(\|x_2 - x_1\|) \dots \varphi(\|x_n - x_1\|) \\ \varphi(\|x_1 - x_2\|) \varphi(\|x_2 - x_2\|) \dots \varphi(\|x_n - x_2\|) \\ \vdots \vdots \vdots \\ \varphi(\|x_1 - x_n\|) \varphi(\|x_2 - x_n\|) \dots \varphi(\|x_n - x_n\|) \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}^T \quad (14)$$

For  $x_i, x_j, i \neq j$ , the sufficient and necessary condition that the formula can have solutions is the symmetric matrix is nonsingular. By solving coefficient matrix and bringing it into the target function, the coefficient output value can be estimated. In addition, the Multi-Quadric is selected to be kernel function. In this paper, we apply the RBF to approximate the target function of estimating the RSS value of the non-feedback point. The process is shown in **Figure 3**. Firstly, the professional collects fingerprint samples of marked paths, then we determine the coordinates of non-feedback points. Next, with each non-feedback point as the center of designated the cut-off area, RSS of non-feedback point from different anchor RSS value is estimated. Each non-feedback point is traversed according to the above process. Finally, RSS sequences and the corresponding coordinates are combined to add into the sparse fingerprint database.

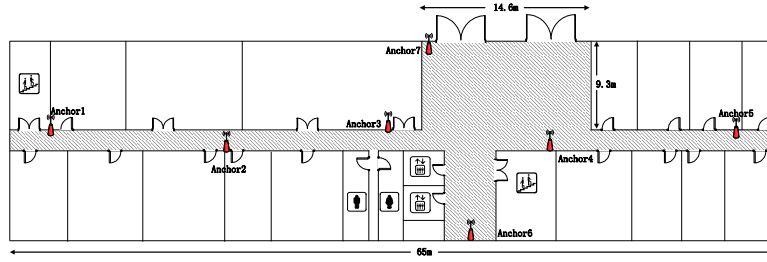


Fig. 4. Layout of target environmental.

## 4 Performance evaluation

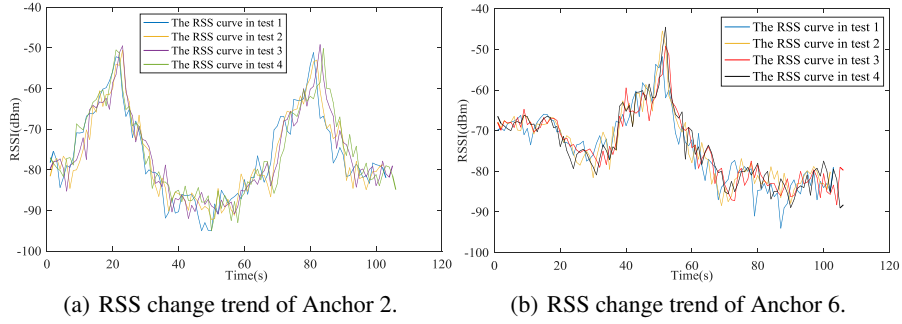
In this paper, we select the first floor of a building in a university to set as the test site, which is a typical office environment, including corridors, halls and rooms. As shown in **Figure 4**, 7 anchors are placed in the environment, and the shadow area is the test area. The total area of about  $1105 \text{ m}^2$ , the hall area is  $140 \text{ m}^2$ , the corridor area is  $116 \text{ m}^2$ .

This system is composed of a tag circuit board, several anchors, a gateway equipment and a computer. The BLE anchor contains a Bluetooth signal transmitting antenna, the tag board is integrated with several sensors. Firstly, the server and gateway are connected, then we supply power to those anchors and tag. On the sever side, the Bluetooth broadcast message can be read, and the MEMS data is hold by the pedestrian terminal.

In the environment, the interval between two blocks of tiles is chosen as sampling distance, and we collect four sampling for each trajectory in sequence. In this paper, each sampling here means walking a round-trip and holding data. Then we select any two anchors and extract the change trend of RSS, which is as shown in **Figure 5**. From **Figure 5**, we can find that in the four tests, the change trend RSS of each anchor is similar, although there is a large error caused by the individual signal jitter. Therefore, we can get stable signal through the dynamic walking way. In this paper, we determine that the sampling interval is 1.6 m, and the sampling number of each track is collected once.

In order to verify the effectiveness of RBF in estimating RSS, we randomly select several samples points for test. The difference between the estimated RSS value and measured value of each point under different cut-off area radius is shown in Table 1. When the cut-off area radius is small, the number of feedback points in the cut-off area is scarce, which leads to large errors. When the cut-off area radius is large, there is complex information and the complexity of the algorithm is increased, which leads to the fall of accuracy. In conclusion, we choose the radius of the cut-off area is 6 m.

In the test area, we collect fingerprints in the direction of  $90^\circ$  and  $270^\circ$ , then we calculate the fingerprints of the other two directions directly according to the mean calculation. The clustering results of the fingerprint database is shown in **Figure 6**, the different colors and shapes are different clusters. It can be seen that each area has similar fingerprint RSS, there is no big noise point and the number of singular point is less than 5% of the number of fingerprint, so the database constructed by the system is good.

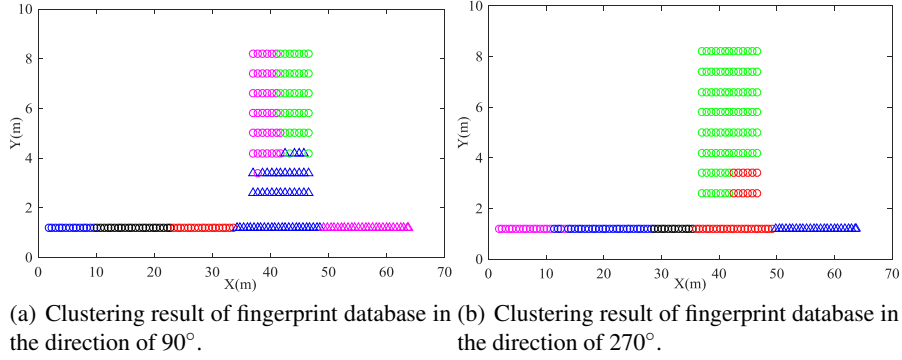


**Fig. 5.** The RSS change trend of different anchor.

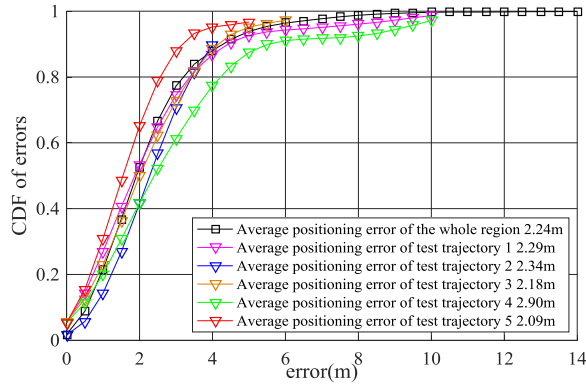
**Table 1.** The difference between the estimated value and measured value of RSS(dBm)

Coordinates	The cut-off area radius (m)				
	r=3	r=4	r=5	r=6	r=7
(37.8,8.2)	4.20	4.53	5.39	5.39	5.43
(40.2,6.6)	0.30	0.83	0.84	0.95	1.01
(40.2,5)	0.27	0.62	0.60	0.60	0.59
(45,3.4)	1.83	1.51	1.26	1.19	1.14
(41.8,8.2)	2.18	2.68	2.77	2.22	2.08
(43.4,6.6)	0.29	0.07	0.93	0.72	0.72
(42.6,5)	2.07	1.42	1.34	1.19	1.20
(39.4,3.4)	1.04	1.21	1.19	1.30	1.29
(39.4,8.2)	0.52	0.67	0.79	0.86	0.88
(44.2,8.2)	0.32	0.20	0.16	0.04	0.04
(46.6,8.2)	1.76	1.04	2.47	1.54	1.77
(45.8,6.6)	1.53	1.60	0.83	0.20	0.17
(38.6,5)	0.67	1.96	1.32	2.26	1.95
(45.8,5)	0.84	6.91	0.79	0.78	0.76
(45.8,6.6)	0.99	0.71	0.88	0.65	0.76
Mean value of the estimated errors(dBm)	1.25	1.73	1.44	1.32	1.32
Bluetooth average positioning error obtained from CDF(m)	2.94	2.85	2.7	2.63	2.65

Subsequently, five sets of random selected test data are collected, including straight line trajectories, closed trajectories and broken line trajectories. The Bluetooth positioning results are as shown in **Figure 7**. From **Figure 7**, we can find that in addition to a few groups of data, these test data can reach the average positioning error of less than 2.4 m, and the cumulative error probability within 3 m is 70%. The black curve represents the location results of the whole area. It can be seen that in complex indoor environment, the database built by our system can achieve the cumulative error probability of 3 m in 78%, and the average error is around 2.2 m, which is a good performance. According to previous studies, in the database from fixed-point collection of fingerprint data method, the cumulative error probability within 2.5 m is can achieve 90%. Although the preci-



**Fig. 6.** Clustering results of fingerprint database.



**Fig. 7.** The Bluetooth positioning accuracy.

sion has fallen in this system, but we have greatly saved the time that takes to collect fingerprint data. In the environment, the total area of the corridor and hall is about  $256 \text{ m}^2$ , and the total time of collecting fingerprint data is  $258 \text{ s}$ . Therefore, the method can reach the efficiency of  $1 \text{ m}^2/\text{s}$ , which is of great effectiveness.

## 5 Conclusion.

To improve the efficiency of fingerprint database construction method, a novel fingerprint database construction method which combines interpolation algorithm and dynamic acquisition of fingerprint data is proposed in this paper. First of all, a number of fingerprints are collected through the dynamical collection approach. Next, a trajectory modification algorithm based on PDR is used to construct an accurate sparse fingerprint database. At last, the RBF is selected to expand the sparse fingerprint database. In the experiment stage, we collect multiple sets of test data. By using the database constructed in this system to carry out experiments, the cumulative error probability within

3 m is 78% and the average positioning error is about 2.2 m. In addition, by clustering the fingerprint database, we find that the fingerprint database can be well clustered into several clusters. At the same time, the method saves more than 80% time compared with the traditional method. So the efficient approach of constructing fingerprint database in this paper can achieve high positioning accuracy, while meeting the existing requirements of people.

## ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China(61771083, 61704015), the Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), Special Fund of Chongqing Key Laboratory (CSTC), Fundamental and Frontier Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), and University Outstanding Achievement Transformation Project of Chongqing (KJZH17117).

## References

1. Y. Zhao, Q. Shen, and L. Zhang. A novel high accuracy indoor positioning system based on wireless lans. *Pier C*. 2011: 25-42.
2. S. Yeh, Y. Chiou, H. Chang, and et al. Performance improvement of offline phase for indoor positioning systems using Asus Xtion and smartphone sensors. *Journal of Communications and Networks*. 2016, 18(5): 837-845.
3. Z. Mu, Y. Tang, Z. Tian, and et al. WLAN indoor localization algorithm based on manifold interpolation database construction. *Journal of Electronics & Information Technology*. 2017, 39(8): 1826-1834.
4. L. Ma, Y. Fan, Y. Xu, and et al. Pedestrian dead reckoning trajectory matching method for radio map crowdsourcing building in WiFi indoor positioning system. *IEEE International Conference on Communications*. 2017: 1-6.
5. J. Liu, Y. Wan, B. Xu, and et al. A novel indoor positioning method based on location fingerprinting. *IEEE International Conference on Communications, Circuits and Systems*. 2013: 239-242.
6. N. Boujnah and P. Korbel. Crowdsourcing based terminal positioning using linear and non-linear interpolation techniques. *IEEE Advances in Wireless and Optical Communications*. 2017: 101-106.
7. M. Zigic and N. Grahovac. Numerical algorithm for rigid body position estimation using the quaternion approach. *Acta Mechanica Sinica*. 2017: 1-9.
8. P. Prasertsung and T. Horanont. A classification of accelerometer data to differentiate pedestrian state. *Computer Science and Engineering Conference*. 2017: 1-5.
9. W. Sun, J. Shen, M. Jin, and et al. A robust magnetic polarity self-sensing method for start-up of PM synchronous machine in fan-like system. *Energy Conversion Congress and Exposition*. 2017, pp(99): 1-8.
10. M. Zhang, W. Shen, Z. Yao, and et al. Multiple information fusion indoor location algorithm based on WIFI and improved PDR. *IEEE Control Conference*. 2016: 5086-5092.