Research Article

Wireless Localization Based on RSSI Fingerprint Feature Vector

Aiguo Zhang,1 Ying Yuan,1 Qunyong Wu,2 Shunzhi Zhu,1 and Jian Deng1

1College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China
2Spatial Information Research Center of Fujian Province, Fuzhou University, Fuzhou 351002, China

Correspondence should be addressed to Aiguo Zhang; zagcfx@163.com

Received 1 February 2015; Accepted 14 May 2015

Copyright © 2015 Aiguo Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

RSSI wireless signal is a reference information that is widely used in indoor positioning. However, due to the wireless multipath influence, the value of the received RSSI will have large fluctuations and cause large distance error when RSSI is fitted to distance. But experimental data showed that, being affected by the combined factors of the environment, the received RSSI feature vector which is formed by lots of RSSI values from different APs is a certain stability. Therefore, the paper proposed RSSI-based fingerprint feature vector algorithm which divides location area into grids, and mobile devices are localized through the similarity matching between the real-time RSSI feature vector and RSSI fingerprint database feature vectors. Test shows that the algorithm can achieve positioning accuracy up to 2–4 meters in a typical indoor environment.

1. Introduction

Localization technology based on RSSI (Received Signal Strength Indication) makes use of radio frequency signals to estimate the distance between the transmitting and receiving devices. Then the position of mobile device is obtained by those distances with trilateration technique algorithm [1]. Currently, localization algorithm is divided into two categories, ranging algorithm and nonranging algorithm. Among them, ranging algorithm is implemented in the way of trilateration, triangulation, or maximum likelihood estimation by measuring the distance and angle between receptors and transmitters, such as TOA, TDAO, and AOA. Nonranging localization algorithm is implemented by the connectivity between different devices and does not need the information of distances and angles. Furthermore, the distances based on RSSI can be measured by the signals transmitted from wireless communication devices, and this kind of localization will not increase hardware costs, so it is the common method for localization in wireless sensor networks [2–9].

But in the actual applications, the distance measurement based on RSSI is influenced by reflection, multipath, antenna gain, and so forth, and large positioning error is caused. Therefore, the localization method based on fingerprint feature vector is put forward in this paper. First, the location area will be divided into grids, and the received RSSI values of all Wi-Fi sources in each grid will be collected. Second, all the received RSSI values will be stored into fingerprint database which can be formed by training sample vectors. Third, the localization results are obtained through the similarity matching between the real-time RSSI value vector of mobile devices and fingerprint database in actual positioning. In this method of localization, neither the known coordinates of Wi-Fi sources are necessary, nor does the RSSI value need to be transferred into distances, so it can reduce the impact of environmental factors.

2. Wireless Localization Design Based on RSSI Fingerprint Feature Vector

In the course of wireless localization based on RSSI fingerprint feature vector, the RSSI values received from all of the wireless APs make up the fingerprint feature vectors of the location grids, and the fingerprint database is established. Then the real-time RSSI value vector received can be identified for fingerprint positioning. Its positioning process is divided into two stages, which are the establishment of fingerprint database and real-time positioning, and shown in Figure 1.
In the course of wireless localization based on RSSI fingerprint feature vector, first, a lot of Wi-Fi signal devices are arranged, typically wireless APs (Access Point), and the number of APs is in correspondence with the dimensions of vector, here represented by $k$. Then the localization area is divided into a number of grids, here represented by $G_{i,j}$ ($i < m$, $j < n$). Finally, the mobile device (MU) can be located via grid matching between real-time RSSI value vector and fingerprint database.

In the stage of establishment of fingerprint database, the RSSI values and MAC (Media Access Control) address in each center of all the grids are collected, and they are stored into the database. Due to environmental factors, single RSSI value is not stable. In order to weaken the influence of unstable RSSI values, multiple times of measurements are carried out and averaged in each grid. In real-time localization stage, the real-time location of the grid is calculated by comparing real-time RSSI value vector received with grid vectors in the database. To the end, the method includes the RSSI data collection and establishment of fingerprint database, grid matching and localization for real-time RSSI value vector, and accuracy analysis of the positioning results, and they will be introduced separately in the following.

### 3. The Data Collection and Organization of RSSI Fingerprint Feature Vector

#### 3.1. Data Collection Points Arrangement for RSSI Fingerprint Feature Vector

As is shown in Figure 2, location area is divided into closely linked grids, and the grids can be arbitrarily sized and shaped. In order to illustrate the idea of positioning, the grids are set to squares in the paper; at the same time, the size of the grid represents the spatial positioning accuracy. All grids are numbered from southwest to northeast orderly, and there are a lot of wireless signal devices in the targeted area, ranked by AP$_1$, AP$_2$, AP$_3$, ..., AP$_k$. In the course of RSSI collection, all the RSSI values from APs in each grid will be collected and the corresponding MAC address will also be recorded.

Typically, the larger vector dimensions of $k$, the richer expression of vector features, and the higher difference between different vectors will be acquired. In order to improve the wireless Wi-Fi positioning accuracy, enough wireless APs should be laid in the location area, but it also makes the RSSI data collection capacity increase. For example, a location area is $100 \times 80$ m, mesh size is $2 \times 2$ m, and the number of the wireless APs is 20; then the recorded number of RSSI values to be stored is $100 \times 80 \times 20/(2 \times 2) = 40000$. For the purpose of management of all the grid vector data effectively, database is suitable for storing and organization of the received RSSI data.

#### 3.2. Data Organization of RSSI Fingerprint Feature Vectors

Large quantity of RSSI fingerprint feature vector needs to be read for localization each time. Therefore, a more efficient
data organization way needs to be used in terms of matching algorithm characteristics on RSSI feature vectors.

In order to improve data management flexibility for all of the APs, each piece of wireless AP data is stored in a separate table; that is, the number of RSSI data tables is equal to the number of the wireless APs, and all the tables have the same structure. Thus RSSI data in the table is consistent with the array of RSSI data collection grids, as shown in Table 1.

In the table, $m$, $n$ correspond to the values of the row and column of the grids, respectively. “MACid” is the primary key for each record, and C01, C02, ..., Cm correspond to the column of RSSI data table; accordingly, Record 01, ..., and Record $m$ correspond to collection records of RSSI value. Names of RSSI data table are identified by the wireless APs’ MAC address; for example, if MAC address is “00: 24: b2: eb: 21: 21,” corresponding RSSI data table will be named “0024b2eb2121.” To do this, make sure that the data table names are unique, and the element values of data table are in correspondence with RSSI values collected in the grids easily.

In practical applications, the number of wireless APs may appear as changes with equipment damage, updates, and so forth; to this end, the fingerprint database must be adjusted. Therefore, only the changed RSSI data tables need to be updated, and the modification of the database can be accomplished easily; at the same time, the design of the database has a good flexibility. In the course of localization, the RSSI data from all the APs must be queried and read, and its process from data query to build a fingerprint feature vector is shown in Figure 3.

In light of the given row and column of the grid ranked by $G_{i,j}$, the RSSI data of $G_{i,j}$ from all the tables are queried, and the corresponding $k$-dimensional feature vector of $G_{i,j}$ can be built. Thus the $k$-dimensional feature vector can be exported for real-time location matching.

4. Grids Matching and Localization of RSSI Fingerprint Feature Vector

4.1. The Localization Algorithm of RSSI Fingerprint Feature Vector. After a given database, the received wireless AP RSSI value vector will compare with the vectors in the fingerprint database in terms of certain matching algorithm, and the mobile location will be estimated. Among them, the matching algorithm is key to the efficiency and positioning accuracy of localization. The usual matching algorithms are nearest neighbor, $K$-nearest neighbor, and neural networks [10]. The method of vector angle is used in this paper.

Assuming that the observations of received real-time RSSI vector at the mobile device are $V_{rt} = (RtRSSI_1, RtRSSI_2, ..., RtRSSI_k)$ and the fingerprint database has vector $V_{i,j} = (RSSI_1, RSSI_2, ..., RSSI_k)$, where $k$ represents the number of detected different wireless APs on the measuring point, $k \in [1, K_T]$, $K_T$ is the total number of RSSI tables in the fingerprint database, and $V_{i,j}$ is on behalf of $k$-dimensional vector at the row $i$ and column $j$. Therefore, the localization based on RSSI fingerprint feature vector is transformed to determine the similarity between real time received observation vector $V_{rt}$ and the fingerprint feature vector $V_{i,j}$.

4.2. Similarity Matching Based on Vector Cosine. There are two ways on determination of vector similarity, that is, the similarity function and distance measurement [11]. Among them, similarity function is more popular in practical applications, and the common similarity functions are as follows: vector cosine method, correlation coefficient, generalized Dice coefficient, and generalized Jaccard coefficient method. Vector cosine function is adopted in this paper.

Vector cosine is used to calculate the angle between two vectors. Assume that the cosine of the angle between two vectors $x$ and $y$ is shown in

$$\cos(x, y) = \frac{(x, y)}{\|x\| \cdot \|y\|} = \frac{\sum_{i=1}^{N} x_i \cdot y_i}{\left(\sum_{i=1}^{N} x_i^2 \cdot \sum_{i=1}^{N} y_i^2\right)^{1/2}}. \quad (1)$$

The geometric meaning of vector cosine is characterized by the angular dimension between two vectors in $N$-dimensional space. Generally, dimensionless treatment for vector elements is required beforehand, and the vector elements are made to be positive; then the cosine of the angle will be at the range of $[0, 1]$. And the greater the value of the cosine, the lesser the angle, which shows the larger similarity between two vectors. If the value is 1, the two vectors are identical. In addition, the length of the vector is specified in the formula, which means that the role of some important part has not been amplified in similarity calculating [12].

In practical application, assume that the real-time RSSI vector of mobile devices is $V_{rt}$, and RSSI fingerprint database grid vector is $V_{i,j}$, where $i = 0, 1, ..., m$, and $j = 0, 1, ..., n$, shown in Figure 4. Then the vector cosine between $V_{rt}$ with
all $V_{i,j}$ can be calculated separately by using the formula of vector cosine, shown in
\[
\cos(V_{rt}, V_{i,j}) = \frac{\sum_{l=1}^{k} (V_{rt,l} \cdot V_{i,j,l})}{(\sum_{l=1}^{k} (V_{rt,l})^2 \cdot \sum_{l=1}^{k} (V_{i,j,l})^2)^{1/2}}.
\] (2)

Among them, $k$ is the number of APs and also for the dimension of a grid vector. As to each real-time mobile localization, the number of $m \times n$ vector angle cosine values will be calculated and takes the grid with maximum cosine value as the positioning result.

### 5. Wireless Localization Experiment for RSSI Fingerprint Feature Vector

Using Java programming language, the Wi-Fi wireless positioning system based on RSSI fingerprint feature vectors has been designed and implemented. Among them, RSSI data collection for mobile devices is developed in Android platform, and the collected RSSI data can be stored into SQLite database in accordance with Section 3.2 in this paper. Then the data in SQLite is dumped into PostgreSQL database. Finally, the data management and application services are developed by the use of Eclipse + Mybatis programming tools, so the cosine between real-time RSSI vector and fingerprint feature vector can be calculated, and also the positioning result corresponding to the largest vector cosine value can be acquired.

The following scenario is as an example of a gymnasium. A wireless localization experiment based on RSSI fingerprint feature vector is made by using this developed program, and the RSSI collection and distribution of the fingerprint database in the stadium area are in Figure 5.

As shown in Figure 6, the whole localization area is divided into grids by the size of $2m \times 2m$, and it comprises grids by the number of 15 rows $\times$ 28 columns; at the same time, the field RSSI vector data in each grid is collected. Meanwhile, there are 22 kinds of Wi-Fi signals which can be received in any of the grids, including China Telecom and China Mobile, and 20 of them were selected for the experiment. The results of RSSI data distribution and storage of the grid with the MAC address of 00: 24: b2: eb: 21: 21 wireless AP are shown in Figure 6.

In this scenario, RSSI value vector of each grid is selected as real-time reference vector individually and used to perform vector cosine matching localization. Meanwhile, in order to check out the positioning results, all kinds of Wi-Fi signals are divided into three groups of $7 + 7 + 6$ in Table 2, and different RSSI value deviations are given to the elements of positioned vector, whereby the corresponding different real-time vectors are obtained. In different RSSI deviation range and accuracy, the correct rate of localization by a total of 420 grids is shown in Table 2.

As shown in Table 2, (1) when signs of the RSSI value deviations of the three groups are the same, the larger deviation of the RSSI value is set, the less correct rate of positioning accuracy, and the correct rate was significantly reduced by the deviation value at 11 to 16 (or $-11$ to $-16$), which is 95.7%–66.2% (or 95.2%–67.8%). (2) when signs of the RSSI value deviations of the three groups are different, the larger deviation of the RSSI value is set, the less correct rate of positioning accuracy, and correct rate of positioning reduced quickly at deviation $2$–$4$, that is, 86.9%–28.3%. Thus different signs of the deviations influence positioning accuracy obviously. At the same time, the sign of each deviation is randomly selected, which may differ from the actual value change trend of RSSI; (3) in the case of the same RSSI value deviations, the larger the range of accuracy, the higher the correct rate of localization.

Under the conditions of the fingerprint database and the deviations of the real-time RSSI vector elements being within 10, the correct rate of localization is more than 90%, which can meet necessity for the mostly actual application. Meanwhile, with the vector dimension increasing, the correct rate of localization will be improved to some extent.

At the same time, the computational complexity and cost have been tested with Lenovo K23, and the results are shown in Table 3.

As shown in Table 3, with the increase of the times of localization in scenario of this paper, the computational complexity and time costs are also increased, and the growth of time costs is not obvious with the increase of times of localization.

### 6. Conclusion

The errors of fitted distance with RSSI values are large, because they are affected by multipath and other factors in indoor positioning environment. In this paper, without fitting distance, also without position of the known APs, the real-time localization results can be obtained by matching similarity between real-time RSSI feature vector and different fingerprint feature vectors. At the same time, with an actual experiment, good results of localization are achieved.
Table 2: Localization results based on RSSI fingerprint feature.

| Number | 7 APs | 7 APs | 6 APs | Real-time RSSI observations (20 deviations with samples training data) | Accuracy range (unit: grid) | Times of real-time localization | Correct grids of localization | Correct rate |
|--------|-------|-------|-------|-------------------------------------------------|-----------------------------|--------------------------------|-----------------------------|--------------|
| 1      | 1     | 1     | 1     | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 100%         |
| 2      | 2     | 2     | 2     | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 100%         |
| 3      | 4     | 4     | 4     | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 100%         |
| 4      | 7     | 7     | 7     | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 99.5%        |
| 5      | 11    | 11    | 11    | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 95.7%        |
| 6      | 13    | 13    | 13    | 0 deviation                                     | 0 deviation                 | 420                            | 369                         | 87.8%        |
| 7      | 13    | 13    | 13    | 1 deviation                                     | 1 deviation                 | 420                            | 378                         | 90%          |
| 8      | 13    | 13    | 13    | 2 deviation                                     | 2 deviation                 | 420                            | 385                         | 91.7%        |
| 9      | 16    | 16    | 16    | 0 deviation                                     | 0 deviation                 | 420                            | 278                         | 66.2%        |
| 10     | 16    | 16    | 16    | 1 deviation                                     | 1 deviation                 | 420                            | 299                         | 71.2%        |
| 11     | 16    | 16    | 16    | 2 deviation                                     | 2 deviation                 | 420                            | 321                         | 76.4%        |
| 12     | 16    | 16    | 16    | 3 deviation                                     | 3 deviation                 | 420                            | 341                         | 81.2%        |
| 13     | -1    | -1    | -1    | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 100%         |
| 14     | -2    | -2    | -2    | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 100%         |
| 15     | -4    | -4    | -4    | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 100%         |
| 16     | -7    | -7    | -7    | 0 deviation                                     | 0 deviation                 | 420                            | 417                         | 99.3%        |
| 17     | -11   | -11   | -11   | 0 deviation                                     | 0 deviation                 | 420                            | 400                         | 95.2%        |
| 18     | -13   | -13   | -13   | 0 deviation                                     | 0 deviation                 | 420                            | 364                         | 86.7%        |
| 19     | -13   | -13   | -13   | 1 deviation                                     | 1 deviation                 | 420                            | 372                         | 88.6%        |
| 20     | -13   | -13   | -13   | 2 deviation                                     | 2 deviation                 | 420                            | 376                         | 89.5%        |
| 21     | -16   | -16   | -16   | 0 deviation                                     | 0 deviation                 | 420                            | 285                         | 67.8%        |
| 22     | -16   | -16   | -16   | 1 deviation                                     | 1 deviation                 | 420                            | 301                         | 71.7%        |
| 23     | -16   | -16   | -16   | 2 deviation                                     | 2 deviation                 | 420                            | 316                         | 75.2%        |
| 24     | -16   | -16   | -16   | 3 deviation                                     | 3 deviation                 | 420                            | 333                         | 79.3%        |
| 25     | -1    | -1    | -1    | 0 deviation                                     | 0 deviation                 | 420                            | 420                         | 100%         |
| 26     | 2     | 2     | -2    | 0 deviation                                     | 0 deviation                 | 420                            | 365                         | 86.9%        |
| 27     | -4    | 4     | -4    | 0 deviation                                     | 0 deviation                 | 420                            | 119                         | 28.3%        |
| 28     | -4    | 4     | -4    | 2 deviation                                     | 2 deviation                 | 420                            | 186                         | 44.3%        |
| 29     | -4    | 4     | -4    | 4 deviation                                     | 4 deviation                 | 420                            | 265                         | 63.1%        |
| 30     | -7    | 7     | 7     | 0 deviation                                     | 0 deviation                 | 420                            | 28                          | 6.7%          |
| 31     | -7    | 7     | 7     | 2 deviation                                     | 2 deviation                 | 420                            | 99                          | 23.6%        |
| 32     | -7    | 7     | 7     | 4 deviation                                     | 4 deviation                 | 420                            | 190                         | 45.2%        |
| 33     | 11    | -11   | -11   | 0 deviation                                     | 0 deviation                 | 420                            | 10                          | 2.4%          |
| 34     | 11    | -11   | -11   | 2 deviation                                     | 2 deviation                 | 420                            | 58                          | 13.8%        |
| 35     | 11    | -11   | -11   | 4 deviation                                     | 4 deviation                 | 420                            | 126                         | 30.0%        |
| 36     | 13    | -13   | -13   | 0 deviation                                     | 0 deviation                 | 420                            | 7                           | 1.7%          |
| 37     | 13    | -13   | -13   | 2 deviation                                     | 2 deviation                 | 420                            | 44                          | 10.5%        |
| 38     | 13    | -13   | -13   | 4 deviation                                     | 4 deviation                 | 420                            | 99                          | 23.6%        |
| 39     | 16    | 16    | -16   | 0 deviation                                     | 0 deviation                 | 420                            | 18                          | 4.3%          |
| 40     | 16    | 16    | -16   | 2 deviation                                     | 2 deviation                 | 420                            | 75                          | 17.8%        |
| 41     | 16    | 16    | -16   | 4 deviation                                     | 4 deviation                 | 420                            | 143                         | 34.0%        |
Table 3: Computational complexity and cost of wireless localization.

| Times of localization | Computational times of vector cosine values | Time costs (s) |
|-----------------------|-------------------------------------------|----------------|
| 1                     | 420                                      | 1              |
| 420                   | 176400                                   | 2              |
| 12600                 | 52920000                                 | 12             |
| 29400                 | 123480000                                | 33             |
| 63000                 | 264600000                                | 43             |
| 105000                | 441000000                                | 70             |

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The work was supported by the National Natural Science Foundation of China (61373147 and 41204032).
References

[1] S. Cao, “Research progress of indoor location technology and system,” *Computer Systems & Applications*, vol. 22, no. 9, pp. 1–5, 2013.

[2] Z. Li, “Research and implementation of RSSI localization principle,” *Radio Engineering*, vol. 43, no. 7, pp. 8–10, 2013.

[3] M. Yu, X. Chen, and J. He, “Study of the localization algorithm for wireless network based on RSSI,” *Journal of Gansu Sciences*, vol. 25, no. 2, pp. 109–111, 2013.

[4] Y. Che and W. Xu, “RSSI-based positioning of wireless sensor network technology,” *Microcomputer Information*, vol. 26, no. 1–4, pp. 82–84, 2010.

[5] S. Wang, “Study of localization algorithm based on RSSI for WSNs,” *Journal of Yunnan University*, vol. 33, supplement 2, pp. 202–205, 2011.

[6] B. Hu, “RSSI-based location technology research,” *Computer Knowledge & Technology*, vol. 8, no. 32, pp. 7807–7808, 2012.

[7] Z. Tan and H. Zhang, “A modified mobile location algorithm based on RSSI,” *Journal of Beijing University of Posts and Telecommunications*, vol. 36, no. 3, pp. 88–91, 2013.

[8] A. Malekpour, T. C. Ling, and W. C. Lim, “Location determination using radio frequency RSSI and deterministic algorithm,” in *Proceedings of the 6th Annual Communication Networks and Services Research Conference (CNSR ’08)*, pp. 488–495, May 2008.

[9] W. Li, L. Jin, and X. Chen, “Indoor positioning system design and implementation based on android platform,” *Journal of Huazhong University of Science and Technology (Nature Science)*, vol. 41, supplement 1, pp. 88–91, 2013.

[10] H. Lu, X. Liu, and C. Zhang, “Comparison of Wi-Fi localization between triangular and fingerprint algorithm,” *Mobile Communications*, vol. 34, no. 10, pp. 72–76, 2010.

[11] Y. Zhang, Y. Liu, and Z. Ji, “Vector similarity measurement method,” *Technical Acoustics*, vol. 28, no. 4, pp. 532–536, 2009.

[12] R. Tian and P. Xie, “Study on the standardization of similarity evaluation method of chromatographic fingerprints (part I),” *Traditional Chinese Drug Research & Clinical Pharmacology*, vol. 17, no. 1, pp. 40–42, 2006.
