Research Article

A Fuzzy Similarity Elimination Algorithm for Indoor Fingerprint Positioning

Yongle Chen,1 Wei Liu,2 Yongping Xiong,2 Jing Duan,3 Zhi Li,4 and Hongsong Zhu4

1 College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China
2 Beijing University of Posts and Telecommunications, Beijing 100876, China
3 Information and Telecommunication Company, Shanxi Electric Power Corporation, Taiyuan 030001, China
4 Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Correspondence should be addressed to Yongle Chen; chenyongle@tyut.edu.cn

Received 25 October 2014; Revised 25 February 2015; Accepted 2 March 2015

Academic Editor: Lin Cai

Copyright © 2015 Yongle Chen 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.

Fingerprint positioning can take advantage of existing WLAN to achieve indoor locations, which has been widely studied. We analyzed the corresponding positions distribution of similar fingerprints, and then found that the fuzzy similarity between fingerprints is the root cause of the larger errors existing. According to clusters distribution feature of corresponding positions of the similar fingerprints, we proposed a $K$-Means+ clustering algorithm to achieve fine-grained fingerprint positioning. Due to the $K$-Means+ algorithm failing to locate the positions of outliers, we also designed a linear sequence matching algorithm to improve the outliers positioning, and reduce the impact of fuzzy similarity. Experimental results illustrate that our algorithm can get a maximum positioning error less than 5 m, which outperforms other algorithms. Meanwhile, all the positioning errors over 4 m in our algorithm are less than 2%. The positioning accuracy has been improved significantly.

1. Introduction

Indoor positioning is the foundation of indoor location services. GPS with indoor limitation and cellular positioning with rough accuracy makes indoor positioning need for fine-grained efficient positioning method. Fingerprint positioning algorithm does not rely on additional hardware overhead, which can utilize the existing infrastructure (e.g., WLAN) to complete the positioning tasks. First, through the acquisition of the received signal strength (RSS) at each sampling point of the object area, the offline fingerprint database will be constructed. Then, we are able to match the signal strength online measured with the fingerprint in the offline database and choose the corresponding position of the optimal matching fingerprint as the positioning result. Therefore, as long as the area can be covered by wireless WiFi networks, where it is easy to realize fingerprint positioning algorithm, the versatility of fingerprint positioning as a technology choice for indoor positioning will be improved.

The current researches on fingerprint algorithm mostly focus on two aspects. One is to reduce the cumbersome workload during offline acquisitions, and another is to improve the fingerprint positioning accuracy. Because the different orientation fingerprints need to be gathered at each sampling point of the target region, the fingerprint acquisition of a large-scale indoor space will be very time-consuming. The current study of this issue is mainly based on the signal model [1, 2] and crowdsourcing method [3, 4], which have achieved many better results. For the positioning accuracy, researchers have designed various online matching algorithms to reduce the positioning error [5, 6] and achieved the median error of around 2 m. But there always are some larger positioning errors with more than 6 m, which is one of the bottlenecks in improving the positioning accuracy. Liu et al. [7] considered the larger positioning errors problem in the practical application of fingerprint positioning. But they ignore the human blocking problem. In fact, the positioning equipment needs to be carried with users, where the human
Finally, we conclude the paper in Section 6. In this paper, we analyze the impacts of different orientations and holding positions and find that body blocking will cause the positioning error over 8 m. This means that the multipath and shadow will further aggravate the larger positioning errors. As far as we know, there are few studies to reduce these large positioning errors caused by human blocking.

For the problem of larger positioning errors, we discuss the position distribution of some offline similar fingerprints and find there is fuzzy similarity between offline fingerprints. The corresponding positions of some similar fingerprints are far apart, called fuzzy similarity, which is the root cause of larger positioning errors. Meanwhile, the multipath and shadow will further aggravate the fuzzy similarity, which leads to more large positioning errors. Through further analyzing positions distribution of similar fingerprints, we also find that most of the corresponding positions had a cluster feature. Therefore, we design a method based on K-Means clustering to achieve fine-grained positioning accuracy, which is called K-Means+ (Algorithm 1). Due to the fact that the K-Means+ positioning algorithm is invalid to the outliers, we further designed a linear sequence matching algorithm to improve the accuracy of outliers positioning, thus eliminating the problem of fuzzy similarity in offline fingerprints. Experimental results demonstrate that our algorithm can improve the positioning accuracy significantly.

The remainder of the paper is organized as follows. The related work is shown in Section 2. We analyze the fingerprint feature in Section 3 while leaving the details of our algorithm design in Section 4. Then we show the experiment and evaluation results of our algorithm in Section 5. Finally, we conclude the paper in Section 6.

### 2. Related Work

Fingerprint positioning with no limit of extra deployment is widely studied. Various methods, such as deterministic kNN [9], Bayesian estimation [10], Sequential Monte Carlo [11], support vector machine [12], and neural network, are used for improving positioning accuracy. But most of fingerprint algorithms rarely reduce the larger errors caused by the body blocking in order to improve the positioning accuracy. Radar system is an earlier attempt of fingerprint positioning using wireless signal strength [8]; although the system finds that body blocking has a serious influence on the positioning accuracy, it does not provide an effective solution. Thereafter, some papers begin to consider body blocking problem during indoor positioning. Papapostolou and Chaouchi [13] build the different direction of signal attenuation model based on a lot of experiments and provide an orientation aware fingerprint positioning algorithm to reduce the influence of body blocking. COMPASS [14] algorithm introduces the device with a digital compass, measuring the personal orientation as a dimension of signal strength fingerprint vector to improve the positioning accuracy with body blocking. LoSF [15] algorithm provides a double node mechanism to avoid the nonline of sight effects from body blocking. These algorithms reduce the influence of body blocking to a certain extent, but they have not considered body blocking impact on fuzzy similarity fingerprint and are unable to avoid larger positioning errors.

Liu et al. [7] consider the fingerprint fuzzy similarity problem in the practical application of fingerprint positioning, which uses the existing mobile phone to provide a peer assisted algorithm. They adopt the sound ranging method to measure the distance between mobile users by the microphone and loudspeaker of mobile phone and use the acquired distance relationship between mobile users to constrain the fingerprint positioning results, which can prevent the emergence of larger errors. This method can avoid larger positioning errors, but it requires additional sound-based ranging method which will increase the energy consumption of positioning service. More importantly, the sound-based ranging method is hardly used in noisy public environments.

We analyze the distribution features of offline similar fingerprints and find similar fingerprints have a cluster position

| 1 | /* Compute the number of clusters of k*/ |
| 2 | k = Precomputed(X) /* X is the subset of fingerprint database*/ |
| 3 | /* X is all the fingerprint database*/ |
| 4 | for all (x ∈ X) do |
| 5 | /* x_k is the fingerprint of cluster center*/ |
| 6 | if (Euclidean(x − x_k) ≤ ε_max) then |
| 7 | x ∈ C_k |
| 8 | end if |
| 9 | end for |
| 10 | if C_k is not convergent then |
| 11 | Center(C_k) /* Compute cluster center*/ |
| 12 | goto (4) |
| 13 | else |
| 14 | return C_k |
| 15 | end if |

**Algorithm 1: K-Means plus.**
distribution feature besides the fuzzy similarity. Therefore, we design an efficient clustering method on offline fingerprints to eliminate fuzzy similarity and avoid the restrictions of sound ranging.

3. Fingerprint Fuzzy Similarity and Positioning Performance

3.1. Body Blocking Influence. To analyze the WiFi signal fingerprints positioning performance, we first conduct a study on the impact of various factors, such as orientation and holding position. Due to the development of smart phones, people use the mobile phone to obtain indoor positioning services increasingly. So we select a GALAXY Note 3 as the WiFi terminal device to acquire signal data. The test mainly studies the multipath and shadow influence on the WiFi signal fingerprint without considering the factor of device diversity.

The testbed is an open lab area of $38 \times 26$ m. Because the desks and chairs cover some parts of indoor office area, we just choose the 76 positions in the passable region (e.g., corridors) to sample signal fingerprint. There are 8 APs deployed for measurement as shown in Figure 1. At each sampling point, the user faces $0^\circ$, $90^\circ$, $180^\circ$, and $270^\circ$ direction and holds the bottom and upper positions of mobile phone, respectively, to measure signal strength. Each measurement acquires 15 groups of signal strengths to calculate an average value. Thus we will generate a total of 608 records in the offline fingerprint database.

Generally, wireless signal strength will change with time leading to some measurement errors [15]. Suppose that these measurement errors follow zero mean normal distribution with $\varepsilon$ variance. Fingerprint matching often uses the Euclidean distance to measure the similarity of fingerprint vectors, and then the maximum measurement error $e_{max}$ between fingerprints could be calculated by the following equation:

$$e_{max} = \sqrt{\sum_{i=1}^{n} \left[ (r_i + \varepsilon) - (r_i - \varepsilon) \right]^2} = 2\varepsilon\sqrt{n},$$

(1)

where $\varepsilon$ is the variance of signal strength distribution, $r_i$ is the received signal strength from $i$th AP, and $n$ is the number of APs. Only when the distance of two fingerprint vectors is greater than $e_{max}$ will there be significant fingerprint dissimilarity, which is called fingerprint granularity in this paper. Since error $\varepsilon$ is the inherent error from signal measurement, we call $e_{max}$ the maximum intrinsic fingerprint granularity error.

To analyze the influence on fingerprint granularity and positioning performance with different orientations and holding positions on mobile phone, we design four group tests using the 608 fingerprint records to evaluate orientation and holding position. We select 30 sampling points of the 76 points in Figure 1 to execute 5 times $k$NN algorithm to compute the average value of the positioning errors. Meanwhile, we compute the Euclidean distances between 608 fingerprints to construct the fingerprint granularity distribution. In Figure 2, different orientation tests include the comparison between $0^\circ$ orientation fingerprints and $90^\circ$ orientation fingerprints and between $0^\circ$ orientation fingerprints while holding the
bottom of mobile device. Different holding positions tests include the comparison between holding bottom and upper position fingerprints and between holding bottom position fingerprints with $180^\circ$ orientation.

Figures 2(a) and 2(b) show the fingerprint granularity cumulative distribution with different orientations and holding positions. Suppose the mean error of WiFi signal strength measurement is 5 db [15]. We can compute that the maximum intrinsic fingerprint granularity error is 28 db by (1). In Figure 2(a), the fingerprint granularity less than 28 db (i.e., similar fingerprints) accounts for 13% and 19%, respectively, where the different orientations have a larger similar proportion. In Figure 2(b), the fingerprint granularity less than 28 db accounts for 18% and 23%, respectively, where the different holding positions have a larger similar proportion. Meanwhile, holding positions have a bigger influence than orientations, which is because the hand is closer to mobile phone than human body. Figures 2(c) and 2(d) show the positioning performance of different orientations and holding positions. We randomly select 30 indoor sampling points and compute the average value of 5 groups of positioning results to compare positioning performance. We find that the higher proportion of similar fingerprints will lead to larger positioning errors. Meanwhile, there are always some larger errors over 6–8 m in the positioning results. This is the main cause of decrease of the positioning performance. To solve this problem, we further analyze the similar fingerprints in the offline database.

3.2. The Root Cause of Larger Errors. In order to analyze the cause of larger errors, we compared the position distribution of similar fingerprints with different orientations. Sampling positions of the similar fingerprints will be displayed in the indoor floor plan. We select the similar fingerprints from 608
fingerprints at the 73rd sampling point with 0° orientation, 40th sampling point with 90° orientation, 51st sampling point with 270° orientation, and 59th sampling point with 180° orientation. The 42nd position distributions are the solid circles in Figure 3, where most of the similar fingerprint positions are close to the 42nd sampling point, and just a few positions are far away. These few outliers are exact cause to produce fingerprint fuzzy similarity; that is, the corresponding positions of similar fingerprints have a large distance, which will lead to larger positioning errors. In addition, we find an obvious cluster feature of sampling points with similar fingerprints. That is, most sampling positions of similar fingerprints are close to each other. The other sampling points in Figure 3 also have a similar distribution feature. Therefore, these outliers with fuzzy similarity in fingerprint database are the root cause of the larger errors. According to the cluster distribution feature of the positions corresponding to similar fingerprints, we try to solve the larger errors problem by clustering method.

4. Removing Fingerprint Fuzzy Similarity

At present, many researches focus on the design of online matching algorithm and filtering optimization algorithm to improve the fingerprint positioning accuracy. The filter can smoothly fit positioning results according to the previous results to avoid the larger deviation. But if there are larger positioning errors, the filtering and fitting will lose efficacy. Based on cluster features of similar fingerprints, we provide a K-Means+ method to cluster the offline similar fingerprints, and we also design a linear sequence matching algorithm to locate the outliers position so as to increase the positioning accuracy.

4.1. Clustering Offline Fingerprints. The traditional K-Means algorithm is a classic machine learning method based on samples similarity measurement. The criterion function is usually the least sum of squared error. Suppose an m-dimensional real vector $X = (x_1, x_2, \ldots, x_m)$ to describe the fingerprint sample; the fingerprint similarity is presented by the Euclidean distance of fingerprint vectors. The traditional K-Means algorithm divides the sample set into $k$ clusters according to the preset parameter $k$. The criterion function can be described as follows:

$$\arg \min_{C} \sum_{i=1}^{k} \sum_{j \in C_i} \|X_j - \mu_i\|^2,$$  \hspace{1cm} (2)

where $C$ is the cluster set, which is described as $C = (C_1, C_2, \ldots, C_k)$. $\mu_i$ is the center of the $i$th cluster. The traditional K-Means algorithm is a dynamic iterative clustering algorithm, but the parameter $k$ must be determined in advance. In the practical application, the $k$ value is difficult to determine and this will directly affect the result of the clustering algorithm. But similar fingerprints positions have significant regional features. So we can obtain the initial $k$ value according to indoor positioning area. Meanwhile, the traditional K-Means algorithm is sensitive to the outliers of clusters. From Figure 3, if two fingerprints have a fuzzy similarity, it will lead to a larger error to compute the center points of clusters. Therefore, we design the K-Means+ algorithm to use the center sampling point instead of the average fingerprint vector as the center of the cluster. Meanwhile, the clustering criterion function uses the maximum inherent fingerprint granularity error $e_{\text{max}}$ instead of the least sum of squared error as clustering decision threshold, which will reduce the iterative times of traditional K-Means. The K-Means+ algorithm is described as follows.

Step 1. Divide the indoor corridors according to a fixed length $l$, and compute the segmentation number $k$. We can obtain $k$ clusters, where the initial cluster center is the center position of each segmentation.

Step 2. Compute the fingerprint granularity between each sampling point and the cluster center, and add the less than $e_{\text{max}}$ sampling point to the corresponding cluster.

Step 3. Recompute the center of all sampling points in each cluster.

Step 4. Set the cluster radius as $l/2$, and repeat Steps 2–3 so the cluster center is in a stable position range or achieves the iteration threshold.

It is important to note that when repeating Step 2, we do not use $e_{\text{max}}$ to cluster sampling point, but decide a sampling point to join the cluster by judging whether the distance upper bound between sampling point and cluster center is more than the given cluster radius $l/2$ in Step 4. This is conducive to reducing the iterative times. Therefore, the algorithm has a low time complexity of $O(\text{nlkt})$, where $n$ is the number of fingerprint samples, $k$ is the number of
clusters, and \( t \) is the iterative times. The K-Means+ algorithm is highly suitable for processing large amount of offline fingerprint data. The essential choice of \( t \) directly determines the clustering number of \( k \) values but is also related to whether the K-Means+ algorithm can guarantee all clusters to cover indoor positioning area. Therefore, the K-Means+ algorithm presents a \( k \) value calculation method. \( l \) value is the physical diameter of similar fingerprint clusters. We need to randomly select \( V_k \) sampling points as the cluster centers and execute once Steps 2 and 3 of K-Means+ method to construct similar fingerprint clusters. The average value of \( l \) can be computed based on the clusters from \( V_k \) sampling points, and then we can compute the clustering number \( k \) based on the average \( l \). According to the precalculated \( k \) values, the K-Means+ algorithm will obtain the offline similar fingerprint clusters.

After clustering the offline fingerprint, we adopt a layered \( k \)NN matching algorithm for positioning. First, we match the sampling real-time fingerprint vector with the fingerprints of the cluster centers. Then we run the exact matching in clusters. The layered \( k \)NN matching algorithm can reduce the matching computational overhead and obtain more accurate positioning result. However, this clustering method cannot solve outliers positioning. When computing the outliers position, it will be false matching to other clusters due to the fact that the outliers do not belong to any cluster. Although the outliers are fewer, they also cause larger positioning error. We further propose a linear sequence matching method to replace the traditional point matching.

4.2. Linear Sequence Matching. Since the K-Means+ algorithm ignores the influence of outlier data, the outlier sampling points are difficult to obtain the accurate positioning result. We propose a linear sequence matching algorithm to replace the traditional point matching. In the process of positioning, we record matching position sequence and set the length of the sequence as \( s \). Since the offline fingerprint cluster covers a larger area, two adjacent positioning intervals usually do not exceed the scope of the cluster. Sequence generating process is described in Algorithm 2.

Algorithm 2: Linear sequence generating.

```
(1) /*Initial position detection*/
(2) InitPos\( (P, S) /* P is the consecutive positioning results set, S is the sequence*/
(3) /*Sequence increase or decrease*/
(4) for all \( (p \in P') \) do
(5) /*P' is the candidate position set*/
(6) if \( (e_t \leq l_o + l/2) \) then
(7) IndirectAdd\( (p, S) \)
(8) end if
(9) if \( (e_t > l_o + l/2) \) and \( (e_t < l_o) \) then
(10) DirectAdd\( (p, S) \)
(11) end if
(12) end for
(13) return \( S \)
```

Step 1 (initial position detection). When there are \( \theta \) consecutive positioning results in the same cluster, we can determine the initial position in the current cluster. We use the \( k \)NN matching method to obtain the precise initial position and take the current position number as the linear sequence header. Otherwise, we continue to execute the positioning and detection. The \( \theta \) is an experimental experience value.

Step 2 (sequence increase or decrease). The initial length of the sequence is 0. The position number will constantly add to the sequence with the fingerprint matching until the length of sequence achieves \( s \), where two consecutive positions can exist in the same cluster. When a new position number adds to the sequence, it must obey the following rules.

(i) Compute the average length \( l_o \) of sequence segments by sliding average method. Set the Euclidean distance between the end position of the sequence and the cluster center as \( e_t \) and between the end position of the sequence and the outlier position as \( e_o \).

(ii) If \( e_t \leq l_o + l/2 \), we select the optimal fingerprint matching position in the cluster to insert in the sequence and delete the sequence header in order to keep the sequence length unchanged.

(iii) If \( e_t > l_o + l/2 \) and \( e_o < l_o \), we add the optimal matching outlier position to the sequence and delete the sequence header. In Figure 4, the dash path represents the numbers that have been deleted, and the solid path represents the current sequence. The same graph stands for the corresponding positions of similar fingerprints. For example, at the \( P \) point, the matching similar fingerprint cluster is represented by circle. The cluster center is away from the sequence end point over \( l_o + l/2 \), and the outlier circle \( P \) can fulfill \( e_o < l_o \). So the outlier point \( P \) adds to the sequence.

(iv) Otherwise, keep the sequence unchanged until next positioning is completed.
Step 3 (matching). The end position of the sequence is the current positioning result.

The linear sequence matching method considers both physical distance and fingerprint distance to avoid the outlier failure problem of $K$-means+ matching algorithm. However, the accuracy of initial position will affect the correctness of the increasing sequence. We will verify the rationality of initial position selection by the experimental method.

5. Experiment Results and Evaluation

5.1. Experiment Design. In order to verify the performance of the linear sequence $K$-Means+ clustering matching algorithm, we still adopt the 608 fingerprint records in Figure 1 with a 38 m * 26 m open office area. We first analyze the impacts of different cluster diameter $l$ on the fingerprint granularity and positioning accuracy and find the optimal $l$ value is mean diameter of clusters generated by maximum intrinsic fingerprint granularity error, which has better clustering performance and reduces the outlier to improve the positioning accuracy. Secondly, we further validate the rationality of selection method of initial position through experiments. Finally, according to the different sequence lengths, we verify the influence of the linear sequence matching algorithm on positioning accuracy.

5.2. Performance Evaluation. Before executing $K$-Means+ method, we need to select $V_k$ sampling points to compute the cluster diameter $l$. In our testbed, 5 sparse sampling points will be selected and 8 APs are deployed where the maximum intrinsic fingerprint granularity error $e_{\text{max}}$ is also 28 db. Based on $e_{\text{max}}$, we can compute the average diameter of 5 sampling clusters as 6.4 m. We take 6.4 m as the optimal cluster diameter. To verify the selection of the optimal cluster diameter, we set $l$ as 3 m, 6.4 m, and 9 m, respectively, to compare the fingerprint granularity and positioning accuracy. Based on the $l$ value, we first compute the $k$ value as 42, 19, and 10. Then we execute the $K$-Means+ algorithm to obtain all similar fingerprint clusters where the outliers will be removed. Based on the fingerprint data holding the bottom position with 0° orientations and holding the upper position with 90° orientations, we recompute the fingerprint granularity distribution according to different $k$ value. As shown in Figure 5, we can find that similar fingerprints will reduce with increasing $k$ value, which is because the outliers with fuzzy similarity are deleted. Meanwhile, the clustered fingerprint granularity between $k = 42$ and $k = 19$ is very close. This is because our $K$-means+ algorithm has no more iterations. We just use the maximum intrinsic fingerprint granularity error $e_{\text{max}}$ to construct similar fingerprint clusters. Although $k$ value is 42, the new cluster diameter computed by $K$-means+ algorithm will be close to 6.4 m, which leads to many overlapping clusters. So the cluster numbers between $k = 42$ and $k = 19$ are almost the same and the fingerprint granularity distribution is similar. When $k = 10$, the cluster numbers will obviously decrease. But the cluster diameter is still 6.4 m around $K$-Means+ algorithm. So the clusters cannot cover the whole positioning area, and many positions will be deleted as outliers lead to fewer similar fingerprints.

We select 30 sampling points of the 76 points in Figure 1 to execute 5 times $K$-means+ algorithm to compare the positioning errors. In Figure 6, when $k = 42$ and $k = 19$, the outliers reduce significantly since the clusters almost cover the whole positioning area. So the larger errors are less than those of $k = 10$. But there are still some position errors nearly 6 m. This is because the deleted outlier positions can not be positioned. When $k = 10$, the clustering results cannot cover all positioning areas. Although similar fingerprints reduce, the positioning performance has not seen any improvement.

When analyzing the influence of different clusters $k$ on fingerprint granularity and positioning accuracy, there are similar performances between $k = 42$ and $k = 19$. The positioning error increases under $k = 10$ although the number of similar fingerprints has declined. To obtain a tradeoff between position performance and clustering overhead, we select $k = 19$ as the experimental value to execute following tests. For the problem of the larger errors after $K$-Means+ clustering, we further verify the performance of the linear sequence matching algorithm.

![Figure 4: A diagram of generating linear sequence.](image)

![Figure 5: Fingerprint granularity distribution under different numbers of clusters.](image)
The initial position of the linear sequence is important to the matching performance. Generally, user cannot move out of the clustering range immediately. According to the above test results, the cluster diameter is about 6.4 m, while the user stride is around 1.5 m. Therefore, successive positioning results are hardly over the range of the same cluster. In the following, we verify whether the adjacent $\theta$ positioning results belong to the same cluster or not, where the $\theta$ is set to 2, 3, and 4. For each trial, we measure 25 groups of the positioning results and analyze the probability and accuracy located in the same cluster. As shown in Figure 7, the initial position of the linear sequence can be determined once with 96 percent and 80 percent, respectively, when $\theta$ is 2 or 3. But the probability will be down to 20 percent when $\theta$ is 4. This is because more successive positions will be beyond the range of the cluster. Although the probability located in the same cluster achieves 100 percent, it needs more positioning tests and is time-consuming. Considering the probability and accuracy synthetically, we set $\theta = 3$ to determine the initial position of the linear sequence in our experiment.

Based on above clustering results, we further verify the influence of linear sequence length on outlier positioning. The sequence length will affect the average sequence segments $l_s$ value. After defining $s$ as 2, 4, and 6, respectively, we analyze the positioning errors of outliers. We select 5 outlier positions with 4 different orientations to execute linear sequence matching according to above 3 groups of sequence lengths. The matching performance is illustrated in Figure 8.

Figure 8(a) describes the matching accuracy of outliers. With the increasing of sequence length, the matching performance has a little improvement. Figure 8(b) represents the average positioning error with wrong matching. The outlier positioning error will reduce while the sequence length increases. But the error falling is not distinct and ranges from around 2 m to 4 m. To achieve the higher positioning accuracy, the sequence length with 6 is a better choice. So we select $s = 6$ to further verify the linear sequence matching performance.

To verify the performance of our fuzzy similarity fingerprint eliminating algorithm, we compare the classical Radar, 2 peers assisted, and $K$-Means+, linear sequence matching methods in Figure 9, which are based on the fingerprint data holding the bottom position with 0° and 90° orientations. We select 30 sampling points of the 76 points to execute 5 times aforementioned algorithms. We find the median error of our $K$-Means+ method is within 2 m and the 90 percent positioning errors are within 3 m, which is better than 4 m in Radar algorithm and 3.6 m in peers assisted algorithm. Particularly, larger errors have reduced significantly. The Radar algorithm can achieve 8 m maximum error, while $K$-Means+ algorithm can lower it to around 6 m. After linear sequence matching, the maximum error will be decreased to 5 m, which is also superior to maximum error of peer assisted algorithm with two phones. This is because the relative ranging in peer assisted algorithm hardly constraints the positioning outliers under human blocking. The proportion of larger errors in positioning results also reduces significantly. The larger errors over 4 m of the Radar method account for 10 percent, while $K$-Means+ algorithm will reduce it to 4 percent, and then the linear sequence matching will further reduce it to 2 percent.

6. Conclusion

Fingerprint positioning is the foundation of indoor location services, which has been widely studied. We analyzed the corresponding position distribution of similar fingerprints and then found the larger errors problem. Through analyzing the corresponding positions distribution of the similar fingerprints, we also found fuzzy similarity is the root cause of larger errors. According to the cluster features of similar fingerprints, we proposed a $K$-Means+ clustering algorithm to achieve fine-grained fingerprint positioning. Due to the $K$-Means+ algorithm failing to locate the positions of outliers, we also designed a linear sequence matching algorithm to
Improving the outliers positioning. Experimental results show that our algorithm can get a maximum positioning error less than 5 m, which is superior to Radar and peer-assisted algorithm. Meanwhile, the proportion of larger errors in our algorithm has significantly declined.

**Conflict of Interests**

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

**Acknowledgments**

This work is supported by National Natural Science Foundation of China (NSFC) under Grant no. 61401300, National High Technology Research and Development Program of China (863) under Grant no. 2012AA013104, the Innovation Program of Institute of Information Engineering Chinese Academy of Sciences (no. Y3Z0071E02), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (STIP) under Grant no. 2014J124, and Youth Foundation of Taiyuan University of Technology (no. 2013Z060 and no. 2014TD054).

**References**

[1] A. Varshavsky, D. Pankratov, J. Krumm, and E. De Lara, “Calibree: calibration-free localization using relative distance estimations,” in *Pervasive Computing*, vol. 5013 of Lecture Notes in Computer Science, pp. 146–161, Springer, Berlin, Germany, 2008.

[2] A. Narzullaev, Y. Park, and H. Jung, “Accurate signal strength prediction based positioning for indoor WLAN systems,” in *Proceedings of the IEEE/ION Position, Location and Navigation Symposium*, pp. 685–688, IEEE, May 2008.

[3] C. Wu, Z. Yang, Y. Liu, and W. Xi, “WILL: wireless indoor localization without site survey,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 839–848, 2013.

[4] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, “Zee: zero-effort crowdsourcing for indoor localization,” in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, pp. 293–304, ACM, 2012.

[5] M. Molina-García, J. Calle-Sánchez, J. I. Alonso, A. Fernández-Durán, and E. B. Barba, “Enhanced in-building fingerprint positioning using femtocell networks,” *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 195–211, 2013.

[6] D. Milioris, G. Tzagkarakis, A. Papakonstantinou, M. Papadopouli, and P. Tsakalides, “Low-dimensional signal-strength fingerprint-based positioning in wireless LANs,” *Ad Hoc Networks*, vol. 12, no. 1, pp. 100–114, 2014.

[7] H. Liu, Y. Gan, J. Yang et al., “Push the limit of WiFi based localization for smartphones,” in *Proceedings of the 18th annual international conference on Mobile computing and networking*, pp. 305–316, ACM, August 2012.

[8] P. Bahl and V. N. Padmanabhan, “Radar: an in-building rf-based user location and tracking system,” in *Proceedings of
the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’00), vol. 2, pp. 775–784, IEEE, Tel Aviv, Israel, 2000.

[9] P. Prasithsangaree, P. Krishnamurthy, and P. K. Chrysanthis, “On indoor position location with wireless lans,” in Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC ’02), vol. 2, pp. 720–724, IEEE, September 2002.

[10] V. Seshadri, G. V. Zaruba, and M. Huber, “A bayesian sampling approach to in-door localization of wireless devices using received signal strength indication,” in Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications (PerCom ’05), pp. 75–84, IEEE, 2005.

[11] C. Morelli, M. Nicoli, V. Rampa, U. Spagnolini, and C. Alippi, “Particle filters for RSS-based localization in wireless sensor networks: an experimental study,” in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’06), vol. 4, p. 4, IEEE, May 2006.

[12] C.-W. Lee, T.-N. Lin, S.-H. Fang, and Y.-C. Chou, “A novel clustering-based approach of indoor location fingerprinting,” in Proceedings of the 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC ’13), pp. 391–396, IEEE, 2013.

[13] A. Papapostolou and H. Chaouchi, “Orientation-based radio map extensions for improving positioning system accuracy,” in Proceedings of the International Wireless Communications and Mobile Computing Conference, pp. 947–951, ACM, June 2009.

[14] T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg, “Compass: a probabilistic indoor positioning system based on 802.11 and digital compasses,” in Proceedings of the 1st International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, pp. 34–40, ACM, Los Angeles, Calif, USA, September 2006.

[15] Y. Chen, H. Zhu, and L. Sun, “Losf: a los-based fingerprint localization algorithm resisting multipath and shadow,” Journal of Computer Research and Development, pp. 524–531, 2013.