An Improved WKNN Indoor Fingerprinting Positioning Algorithm Based on Adaptive Hierarchical Clustering

Jian Li[1], Jingqi Fu[1](✉), Ang Li[1], Weihua Bao[2], and Zhengming Gao[2]

[1] School of Mechatronic Engineering and Automation, Shanghai University, NO. 149 Yanchang Road, JingAn District, Shanghai 200072, China
jqfu@staff.shu.edu.cn

[2] Shanghai Automation Instrumentation Co., Ltd., Shanghai, China

Abstract. Aiming at the dependence of the traditional indoor clustering positioning accuracy on the initial center and clustering number selection, an improved WKNN indoor fingerprint localization algorithm based on adaptive H clustering algorithm is proposed in this thesis. Specifically, an adaptive hierarchical clustering combined with positioning environment and fingerprint information without initial clustering center is introduced. At the same time, a RSSI information compensation method based on cosine similarity is proposed aiming at the problem of RSSI information packet loss for test nodes in complicated indoor location environment, with the result of positioning error decrease at test node by using cosine similarity between test nodes and fingerprint points to approximately compensate the missing RSSI information. The experimental results indicate that the proposed adaptive hierarchical clustering algorithm can divide the experimental area adaptively according to fingerprint information, meanwhile the proposed fingerprint information compensation method can decrease the positioning error of the test node with incomplete information, by which the average positioning error in the experimental environment is decreased to 0.78 m compared with other indoor positioning algorithms.

Keywords: Location fingerprint localization · Adaptive hierarchical clustering · Cosine similarity · Received signal strength indication

1 Introduction

With the comprehensive popularization of wireless communication technology, location-based service problems are attracting more and more attention [1]. However, traditional positioning systems such as GPS could not provide a good positioning service when used in such environment. Hence, exploring effective methods of indoor positioning has become a research focus [2].

Fingerprint location algorithm based on received signal strength indicator (RSSI) decreases the effects of the multi-path, non-line of sight and other environmental factors through the idea of mathematical statistics brought by the complexity of the internal structure [3], which is becoming the main method for indoor positioning

M. Fei et al. (Eds.): LSMS/ICSEE 2017, Part I, CCIS 761, pp. 253–262, 2017.
DOI: 10.1007/978-981-10-6370-1_25

technology nowadays. However, the problem of higher computation and lower positioning efficiency arise due to the large amount of fingerprint data when applied in large-scale complicated indoor scenes. Therefore, it is necessary to design corresponding algorithms to achieve the rapid matching for target location.

In [4], the KNN fingerprint localization algorithm is proposed. The node is located by matching its position information with the fingerprints among the whole fingerprint space in the algorithm, but it is difficult to eliminate the interference of singular points. In [5], an enhanced weighted K nearest neighbor algorithm is proposed by changing the number of neighbors to be considered. Literature [6] uses the point sporadic intensity of the nearest neighbor to determine the reference point control network and in order to select the key parameter K of KNN algorithm dynamically. However, the KNN algorithm has high computational complexity and low matching efficiency. In [7], the method of K-means clustering is used to cluster the fingerprint points with matching area determined by matching the node with the sub-category. The computation is reduced, however, it is much easy to trap into the local optimum. In [8], an error beacon filtering algorithm is proposed, but the algorithm does not overcome the problem that the clustering results are sensitive to initial center values. In [9], a KNN algorithm based on fuzzy C-means clustering (FCM) is proposed. It is used to soften the candidate points, meanwhile the membership degree of each candidate point is determined and then clustered. In [10], a WLAN hybrid indoor location method based on FCM and artificial neural network (ANN) is proposed. FCM method is used to select the reference point affected by multipath effect, and ANN is used to approximate the coordinate position. However, the parameter values introduced by the hybrid algorithm are complicated and are difficult to get the optimal combination of parameters.

In order to solve problems mentioned above and the RSSI packet loss of the node to be positioned due to environmental noise in complicated indoor environment, an improved weighted K-nearest neighbor location algorithm based on adaptive hierarchical clustering (AHC-iWKNN) is proposed in Sect. 2. Adaptive hierarchical clustering is used to classify fingerprint space for the sake of avoiding parameter setting and initial clustering center selection in the algorithm. Meanwhile, a RSSI compensation method based on cosine similarity is proposed to reduce the positioning errors of nodes to be positioned with RSSI packet loss. In Sect. 3, the proposed algorithm is verified by the built-in indoor positioning system and the compare results are listed. Conclusions are introduced in Sect. 4.

2 Research on Improved Fingerprint Location Algorithm

The positioning nodes can be located by different methods base on fingerprint location. A novel fingerprints clustering method and an improved Weighted K-Nearest Neighborhood Matching Algorithm are introduced in this section.

2.1 Fingerprint Space Division Based on Adaptive Hierarchical Clustering

Hierarchical clustering algorithm is also called tree clustering algorithm. Unlike the K-means clustering, there is no need to determine the global optimal criterion function and the initial clustering center for hierarchical clustering algorithm. Instead, the threshold criteria is used which clustering according to the given distance threshold or cluster number. A threshold T is set based on human experience to stop the clustering iteration for cohesive hierarchical clustering. In this thesis, an adaptive hierarchical clustering method is proposed, which can generate the distance threshold adaptively with the assurance of difference among clusters by using the maximum distance in the distance similarity matrix as the distance threshold.

The main idea of clustering fingerprints based on adaptive hierarchical clustering is:

(1) At the initial layer, a fingerprint data x_i consisting of the position coordinates and the RSSI information is regarded as a cluster C_i.

$$C_i = \{x_i\}, \forall i. \tag{1}$$

(2) Calculate the Euclidean distance between any two fingerprint data:

$$d_{i,j} = D(C_i, C_j), \forall i, j. \tag{2}$$

Where C_i, C_j are fingerprint data in (2).

Then, generate the distance similarity matrix and find out the maximum distance $(d_{\max})$ as the threshold T.

$$T = \max(d_{i,j}), \forall i, j. \tag{3}$$

In (3), $d_{i,j}$ is the Euclidean distance from fingerprint i to fingerprint j, calculated by formula (2).

As can be seen in Table 1, A, B, C, D, E, F, G are six fingerprint points with known coordinates, d(A → B) represents the Euclidean distance from fingerprint A to fingerprint B, which equals to d(B → A). In addition, $d_{\max}$ is assumed as d(B → D) in Table 1.

Table 1. Distance similarity matrix of 6 clusters.

	A	B	C	D	E	F
A	0	d(A→B)	d(A→C)	d(A→D)	d(A→E)	d(A→F)
B	d(B→A)	0	d(B→C)	d(B→D)	d(B→E)	d(B→F)
C	d(C→A)	d(C→B)	0	d(C→D)	d(C→E)	d(C→F)
D	d(D→A)	d(D→B)	d(D→C)	0	d(D→E)	d(D→F)
E	d(E→A)	d(E→B)	d(E→C)	d(E→D)	0	d(E→F)
F	d(F→A)	d(F→B)	d(F→C)	d(F→D)	d(F→E)	0

(3) Find two clusters (C_a and C_b) with the smallest distance in the range of the whole fingerprint database:

$$a, b = \arg\min_{i,j} d(i,j). \tag{4}$$

Then, merge C_a and C_b into a new cluster C_n and calculate the Euclidean distance between C_n and the other clusters except C_a and C_b. Finally, delete C_a and C_b.

$$C_n = \{C_a, C_b\}. \tag{5}$$

$$Restruture\left(\sum C_i\right) = \sum C_i - C_a - C_b + C_n. \tag{6}$$

For Table 1, find out the minimum distance ($d_{\min}$) in the distance similarity matrix, assume d(D → F). Then merge two clusters with the smallest distance (D, F) into a new cluster G, generate a new distance similarity matrix between cluster G and remaining clusters, as shown in Table 2.

(4) Loop (3), compare T with the minimum distance ($d_{\min}$) in each merging process and stop clustering until $d_{\min}$ is bigger than T.

Table 2. Distance similarity matrix of 5 clusters

	A	B	C	E	G
A	0	d(A→B)	d(A→C)	d(A→E)	d(A→G)
B	d(B→A)	0	d(B→C)	d(B→E)	d(B→G)
C	d(C→A)	d(C→B)	0	d(C→E)	d(C→G)
E	d(E→A)	d(E→B)	d(E→C)	0	d(E→G)
G	d(G→A)	d(G→B)	d(G→C)	d(G→E)	0

2.2 Improved Weighted K-Nearest Neighborhood Matching Algorithm

In the process of locating, beacon nodes are easy to be blocked due to indoor obstacles and frequent flow of personnel, which lead to serious interference in the signal strength or even no RSS from beacon nodes for test nodes. In order not to affect the positioning matching, for conventional methods, remaining RSS is applied to calculate the distance or attach the same value of the fingerprint data from the matching fingerprint point to signal strength values that the node to be located could not receive from the beacon nodes. For former method, locating the node with RSS packet loss easily results in matching misalignment. For the latter, test nodes with forced assignment are easy to lose the signal characteristics at the location of the node to be located, which will reduce the accuracy of matching.

In this thesis, a method of compensating test nodes RSS by cosine similarity is proposed, where a complete RSS vector which has similar characteristics of test nodes is obtained and fingerprint matching based on Euclidean distance is performed. Cosine similarity refers to measuring the cosine of the angle of two vectors to determine the degree of similarity between two vectors. According to the Euclidean inner product formula, the cosine of the angle between two phases which is also called the similarity of two vectors is calculated by:

$$similarity = \cos(\theta) = \frac{A \bullet B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i}\sqrt{\sum_{i=1}^{n} B_i}}. \tag{7}$$

Where A, B in (7) are n-dimensional vector.

The test nodes can receive the RSS signals from all beacon nodes without interference after beacon nodes deployed in the environment and the RSS vector of a test node should be contained by:

$$s^* = \{s_1, s_2, \cdots, s_N\}. \tag{8}$$

Where s^* is N-dimensional vector. N is the number of beacon nodes.

The number of RSS values in s^* reduces to $N - p$ when test nodes couldn't receive signals from p beacon nodes and the algorithm is introduced as follows:

(1) Adaptive hierarchical clustering of fingerprint data.

Define ψ as fingerprint space:

$$\psi = \begin{bmatrix} x_1 & y_1 & RSS_{1,1} & RSS_{1,2} & \cdots & RSS_{1,N} \\ x_2 & y_2 & RSS_{2,1} & RSS_{2,2} & \cdots & RSS_{2,N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_M & y_M & RSS_{M,1} & RSS_{M,2} & \cdots & RSS_{M,N} \end{bmatrix}. \tag{9}$$

In (9), M is the number of fingerprint points, N is the number of RSS values from all beacon nodes, and (x_i, y_i) represents the coordinate of a fingerprint point.

Clustering the fingerprint data through adaptive hierarchical clustering. First, calculate the Euclidean distance between any two fingerprint points:

$$D_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{10}$$

Where i, j in (10) are two different fingerprint point in fingerprint space ψ.

At the beginning of the algorithm, each fingerprint data is regarded as a cluster, the threshold T is obtained from the distance similarity matrix of all fingerprint data. Start clustering by merging two clusters with smallest distance calculated by (10) into a new cluster until the smallest distance is rather bigger than T. Then, calculate the RSS vector of the centroid of each cluster:

$$Centroid_f = \frac{1}{H}\sum_{i=1}^{H}(RSS_{i,1}, RSS_{i,2}, \cdots, RSS_{i,M}), f = 1, \ldots, k. \quad (11)$$

In (11), H is the number of fingerprints in a cluster, k is the number of clusters.

(2) Pretreatment for the vector of the test node (s^*).

All the fingerprint data in ψ is processed as follows: First, remove p corresponding RSS values of all fingerprint points and construct the new fingerprint space v. Calculate similarities between s^* ($N - p$ dimensional vector) and all fingerprint RSS vector in v by formula (7):

$$\cos(\theta) = [\cos(\theta_1), \cdots, \cos(\theta_M)]^T. \quad (12)$$

In (12), M is the number of fingerprint points.

Then, select the maximum p similarities from $\cos(\theta)$, calculate the mean value of p similarities $(\cos(\theta)_{mean})$ and record the corresponding p fingerprint points.

$$\cos(\theta)_{mean} = \frac{1}{p}\sum_{i=1}^{p}\max(\cos(\theta)). \quad (13)$$

In (13), $\cos(\theta)$ is the similarities vector calculated by formula (12).

Finally, regard missing p RSS values in s^* as unknown, meanwhile substitute corresponding p RSS vector that recorded above and $\cos(\theta)_{mean}$ into formula (7) and solve the p-dimensional equations, as can be seen in formula (14). Solutions of the equations approximately equal to p RSS values that s^* lacks so that the new complete RSS vector $s^{*\prime}$ could be constructed.

(3) Matching and location.

Calculate the Euclidean distance between $s^{*\prime}$ and $Centroid_f$:

$$L_j = \sqrt{\sum_{i=1}^{N}(s_i - RSS_i)^2}, j = 1, \ldots, k. \quad (14)$$

Where N is the RSS values of each fingerprint point received from all beacon nodes.

Then, compare the distance between the test node and k centroids, and the cluster that has smallest centroid distance becomes the cluster for test node. Finally, use weight K-nearest algorithm to obtain the estimated coordinate of the test node.

3 Experimental Results and Analysis

In order to verify the effectiveness of the improved algorithm proposed in this thesis, an indoor positioning system based on wireless sensor network is developed independently. The experiment is carried out in Room 305, Western Automation Building, Shanghai University. The area is a 5.7 m × 4 m rectangular laboratory, as is shown in Fig. 1, where there is the office furniture such as tables, chairs and bookcases. The positioning system contains a wireless sensor measurement node A1, 5 beacon nodes, a wireless sensor gateway and detection software in the PC. Measurement node A1 and 5 beacon nodes all contain the TI's CC2530 chip with wireless communication capabilities. In addition, the wireless sensor gateway (including TI's RF module CC2591 and AT91SAM9260 processor) is also integrated in measurement node A1. The experiment area is divided into 35 fingerprint locations and each of which has a 0.5 m × 0.5 m area. A fingerprint point of each location is randomly selected at each area. The positions of 20 test nodes which are randomly selected can receive RSSI values from 3 to 5 beacon nodes. 200 groups of RSS values of the fingerprint point are collected continuously in the offline phase and 20 groups of RSS values of the test nodes are acquired continuously during the online matching phase.

Fig. 1. Experimental area layout

3.1 Results and Analysis of Hierarchical Clustering Experiment

The proposed system is used to build the fingerprint space and MATLAB software is applied in splitting fingerprint points by K-means clustering and adaptive hierarchical clustering. According to reference [7], with the positioning accuracy assured, K-means clustering number is set for 4 on the basis of experimental area size in order to minimize the amount of calculation. The distance threshold T is 3.65 m according to the distance similarity matrix of fingerprints in experimental area.

In order to further verify the clustering effectiveness of adaptive hierarchical clustering on fingerprint points, a positioning experiment for 20 test nodes (p1, p2, …, p20) is conducted by using the above positioning system. Among these test nodes, p9, p10 and p18 lack the RSSI signals from node4 and node5. Three algorithms are KNN, Kmeans+WKNN, AHC+WKNN that can be applied to find the position of test nodes. Table 3 is the maximum errors and average errors of three algorithms for the positioning to test nodes. Figure 2 is the positioning error contrast curve of test nodes among three algorithms.

Table 3. Maximum errors and average errors of three algorithms

Algorithm	Average error (m)	Max error (m)
KNN	1.02	1.81
AHC+WKNN	0.84	1.44
Kmeans+WKNN	1.33	2.83

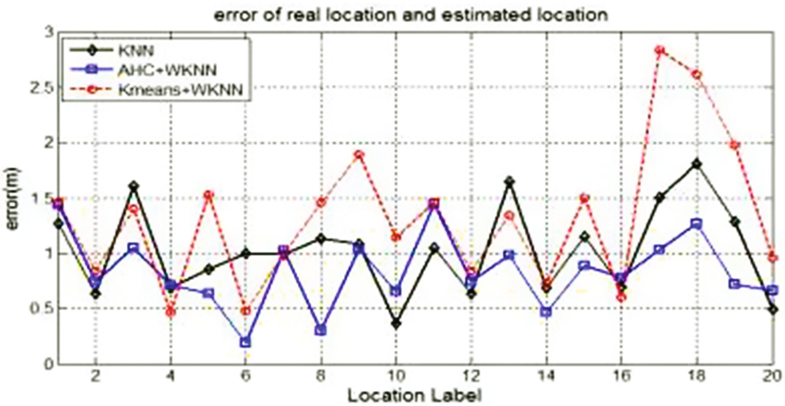

Fig. 2. Positioning error contrast curve of three algorithms

The maximum error and average error of the AHC+WKNN algorithm are 1.44 m and 0.84 m, as is shown in Fig. 2, with 1.39 m and 0.49 m errors decreased respectively when compared with that of the Kmeans+WKNN algorithm from Table 3. The experimental results show that the adaptive hierarchical clustering has a better performance in positioning than K-means clustering.

3.2 Results and Analysis of Improved Weight KNN Location Algorithm

In Fig. 2, for test nodes with incomplete RSSI signals, positioning errors which increase the average positioning error are eliminated more than 1 m except P10 by clustering algorithms. In this thesis, an improved weighted KNN location algorithm for the test nodes with RSS packet loss is proposed, in addition, in order to verify the performance of the proposed algorithm, the standard WKNN location algorithm and the improved WKNN location algorithm are used to estimate the position of test nodes after the hierarchical clustering for the fingerprint space. The positioning errors of the test nodes with RSS packet loss can be seen in Table 4 and Fig. 3 is the contrast curve of positioning errors between the WKNN algorithm and the improved WKNN algorithm after adaptive hierarchical clustering.

Table 4. Positioning errors of test nodes with incomplete RSSI signals

Test node	AHC+WKNN(m)	AHC+iWKNN(m)
P_9	1.04	0.35
P_{10}	0.64	0.45
P_{18}	1.26	0.99
All nodes	0.83	0.78

In Fig. 3, for test nodes with complete RSSI signals, positioning errors are same when using two algorithms respectively. As can be seen in Table 4, positioning errors for all the test nodes with incomplete RSSI signals are reduced, where the decreases are 66.3% at test node P9 and 21.4% at test node P18 respectively. In addition, the average positioning error is decreased from 0.84 m to 0.78 m. Experiments indicate the improved weighted KNN localization algorithm proposed in this thesis can improve the positioning accuracy of the nodes to be positioned with RSS packet loss.

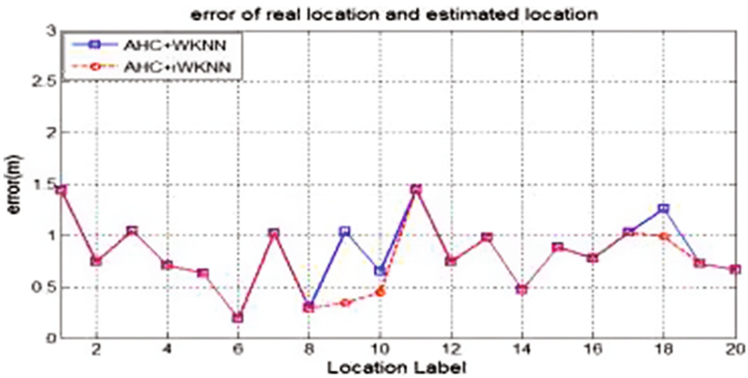

Fig. 3. Positioning error contrast curve of two algorithms

4 Conclusion

Aiming at the dependence of the traditional indoor clustering positioning accuracy on the initial center and clustering number selection, an adaptive hierarchical clustering algorithm is proposed in this thesis. Specifically, an adaptive clustering positioning is realized by the method of combining unnecessary of setting initial clustering center with hierarchical clustering and a method of obtaining the optimal clustering number adaptively according to the fingerprints information proposed in this thesis. Simulation results show that the average positioning error dropped to 0.84 m in the 5.7 m × 4 m experimental environment, achieving a good performance in clustering with the calculation amount increased. At the same time, in order to solve the problem of RSS packet loss for test nodes located in the indoor positioning process in complicated environment, a method to compensate RSSI signals of test nodes by cosine similarity is proposed, which improves anti-noise performance of algorithm in such environment with the result of improving the positioning accuracy further. Experimental results indicate the proposed algorithm can compensate the RSSI of test nodes reasonably in the case of test node packet loss, which optimizes the positioning effects and the average positioning error of all test nodes is reduced from 0.84 m to 0.78 m in the 5.7 m × 4 m experimental environment.

Acknowledgment. This work was financially supported by the Science and Technology Commission of Shanghai Municipality of China under Grant (No. 17511107002).

References

1. Gu, Y.Y., Lo, A., Neimegeers, I.: A survey of indoor positioning system for wireless personal network. IEEE Commun. Surv. Tutor. **11**(1), 13–32 (2009)
2. Xu, J., Luo, H., Zhao, F., Tao, R., Lin, Y.: Dynamic indoor localization techniques based on RSSI in WLAN environment. In: 2011 6th International Conference on Pervasive Computing and Applications (ICPCA), pp. 417–421 (2011)
3. Liao, X.-Y., Ke, H., Min, Y.: Research on improvement to WiFi fingerprint location algorithm. In: 10th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2014), Beijing, pp. 648–652 (2014)
4. Xu, Y., Zhou, M., Meng, W., Ma, L.: Optimal KNN positioning algorithm via theoretical accuracy criterion in WLAN indoor environment. In: Global Telecommunications Conference (GLOBECOM 2010). IEEE (2010)
5. Shin, B., Lee, J.H., Lee, T., Kim, H.S.: Enhanced weighted K-nearest neighbor algorithm for indoor Wi-Fi positioning systems. In: 2012 8th International Conference on Computing Technology and Information Management (NCM and ICNIT), Seoul, Korea (South), pp. 574–577 (2012)
6. Liu, Y., Wang, J.: Normalized KNN model based on geometric clustering fingerprint library. J. Wuhan Univ. (Inf. Sci. Ed.) (11), 1287–1292 (2014)
7. Altintas, B., Serif, T.: Improving RSS-based indoor positioning algorithm via k-means clustering. In: 17th European Wireless 2011 - Sustainable Wireless Technologies, Vienna, Austria, pp. 1–5 (2011)
8. Liu, L., Du, J., Guo, D.: Error beacon filtering algorithm based on K-means clustering for underwater Wireless Sensor Networks. In: 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), Beijing, pp. 435–438 (2016)
9. Sun, Y., Xu, Y., Ma, L., et al.: KNN-FCM hybrid algorithm for indoor location in WLAN. In: 2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS), vol. 2, pp. 251–254. IEEE, Zhangjiajie (2009)
10. Yubin, X., Zhou, M., Lin, M.: Hybrid FCM/ANN indoor location method in WLAN environment. In: 2009 IEEE Youth Conference on Information, Computing and Telecommunication, Beijing, pp. 475–478 (2009)