KNN-FCM Hybrid Algorithm for Indoor Location in WLAN

Yongliang Sun, Yubin Xu, Lin Ma, Zhian Deng
Communication Research Center
Harbin Institute of Technology
Harbin, China
syl_peter@163.com

***Abstract*—As a fingerprint match method, *k*-nearest neighbors (KNN) has been widely applied for indoor location in Wireless Local Area Networks (WLAN), but its performance is sensitive to number of neighbors *k* and positions of reference points (RPs). So fuzzy *c*-means (FCM) clustering algorithm is applied to improve KNN, which is the KNN-FCM hybrid algorithm presented in this paper. In the proposed algorithm, through KNN, *k* RPs are firstly chosen as the data samples of FCM based on received signal strength (RSS). Then, the *k* RPs are classified into different clusters through FCM based on RSS and the position coordinates. According to the rules proposed in this paper, some RPs are reselected for indoor location in order to improve the location precision. Simulation results indicate that the proposed KNN-FCM hybrid algorithm generally outperforms KNN when the location error is less than 2m.**

***Index Terms*—WLAN, indoor location, fingerprint match, KNN, FCM**

I. INTRODUCTION

As the development of wireless networks, lots of applications and technologies that are relative to location service have sprung up. As for outdoor location, global positioning system (GPS), as a well developed technology, has been widely used in many fields. But it is difficult to be used in indoor location due to the building barrier. As for another outdoor location system, Cellar system is also not suitable for indoor location, because the location error is hard to bear relatively to the narrow indoor space. Thus, lots of researches have been focused on indoor location [1].

Compared with other location technologies such as angle of arrival (AOA), time of arrival (TOA) and time difference of arrival (TDOA) [2], the indoor location system based on received signal strength (RSS) in Wireless Local Area Networks (WLAN) is widely researched for the simplicity and economy of fingerprint match method, because WLAN has been widely deployed in indoor environment and no extra devices need to be assembled at mobile terminal (MT) besides a wireless network card. Through fingerprint match method, the measured RSS samples that are received by the MT are matched with the radio–map to estimate the position [3]. Generally, the indoor location system based on location fingerprints includes two phases: off-line phase and on-line phase. In off-line phase, enough RSS samples from the available access points (APs) at each of reference points (RPs) are recorded in order to establish the radio-map. In on-line phase, many fingerprint match methods are applied to estimate the position of MT [4]. *K*-nearest neighbors (KNN) is a popular method for its simplicity and performance. But parameter *k* and positions of RPs have great effect on the performance of KNN. In this paper, Fuzzy *c*-means (FCM) clustering algorithm, as a kind of fuzzy clustering algorithm, is combined with KNN to solve the problems.

The rest of the paper is organized as follows: In the second section, the theories of KNN and FCM are firstly introduced and the KNN-FCM hybrid algorithm is presented subsequently. The experimental process is described in detail and experimental results are given and analyzed in the third section. Conclusion and future work are included in the fourth section.

II. KNN-FCM HYBRID ALGORITHM

A. K-nearest neighbors algorithm

In KNN, the distances between measured RSS samples in on-line phase and the location fingerprints in radio-map are computed. Assuming the number of RPs and APs is m and n respectively. The distances are defined as follows

$$\begin{cases} L_{qi} = \left(\sum_{j=1}^{n} \left| s_j - S_{ij} \right|^q \right)^{1/q} \\ i = 1, 2, \cdots, m \end{cases} \tag{1}$$

Where S_{ij} is the RSS sample from jth AP at ith RP in radio-map, s_j is the RSS sample from jth AP measured in on-line phase [5].

There are k RPs are chosen according to the first k minimum distances by (1), then the estimated location is calculated by

$$(\hat{x}, \hat{y}) = \frac{1}{K} \sum_{i=1}^{K} (x_i, y_i) \tag{2}$$

Where (x_i, y_i) is coordinate of ith RP, $(\hat{x}, \hat{y})$ is the estimated coordinate of the test point (TP).

For KNN algorithm, parameter k and q must be considered. Parameter k denotes the number of neighbors, which is the number of RPs for the coordinate estimation. Parameter q determines the distance type. Usually, parameter k and q are

978-1-4244-4543-1/09/$25.00 ©2009 IEEE

selected as positive integer. When parameter q equals 1, it means the Manhattan distance, and parameter q equals 2 denotes Euclidean distance. Considering the algorithm complexity and location precision, parameter q usually equals 1 or 2. Parameter k and q will be assigned with different values according to different circumstances.

B. *Fuzzy c-means clustering algorithm*

Among the objective function-based clustering algorithms, the theory of FCM clustering algorithm is the most mature one. It was derived from the optimization of hard c-means (HCM) clustering algorithm. In order to solve clustering problems by objective function, the general form of objective function is described as follows

$$\begin{cases} J_m(U,P) = \sum_{k=1}^{n}\sum_{i=1}^{c}(\mu_{ik})^m (d_{ik})^2, \quad m \in [1,\infty) \\ (d_{ik})^2 = \left\| x_k - p_i \right\|_A = (x_k - p_i)^{\mathrm{T}} A (x_k - p_i) \end{cases} \tag{3}$$

Where n and c separately denotes the number of data samples and clustering centers, $U = [\mu_{ik}]_{c\times n}$ is the fuzzy classified matrix with membership function $\mu_{ik} \in [0,1]$, $P = [p_i]$ is the set of c clustering centers, m is weighted index, x_k is data sample, and A is symmetric positive definite matrix [6]. When $A = I_{S\times S}$ in this paper, $(d_{ik})^2$ is the Euclidean distance.

Because $J_m(U,P)$ denotes the weighted square sum of distances between data samples and clustering centers, $J_m(U,P)$ must be minimized according to the fuzzy clustering criterion. The steps of the FCM clustering algorithm are described as follows

1) Set up the number of clusters c, iterative stopped threshold ε, iterative counter b=0, and initialize $U^{(0)}$.

2) P is computed by

$$P_i^{(b)} = \sum_{k=1}^{n}\left(\mu_{ik}^{(b)}\right)^m \cdot x_k \Big/ \sum_{k=1}^{n}\left(\mu_{ik}^{(b)}\right)^m, \quad i = 1,2,\cdots,c \tag{4}$$

3) $U^{(b)}$ is updated by

$$\mu_{ik}^{(b+1)} = \left\{ \sum_{j=1}^{c}\left[\left(\frac{d_{ik}^{(b)}}{d_{jk}^{(b)}}\right)^{\frac{2}{m-1}}\right]\right\}^{-1}, \quad \forall i,k \tag{5}$$

4) If $\left\| P^{(b)} - P^{(b+1)} \right\| < \varepsilon$, then the iteration is finished and U and P are exported. Otherwise, let $b = b+1$ and turn to step (2).

C. *Algorithm design*

The entire location process with KNN-FCM hybrid algorithm is described by a flow chart shown in Fig. 1. In off-line phase, the radio-map is established. In on-line phase, the KNN-FCM hybrid algorithm is generally divided into four steps. Firstly, k RPs are chosen through KNN according the collected RSS samples, which provides data samples for the following FCM clustering algorithm.

Secondly, the k RPs are divided into c_{RSS} clusters based on RSS through FCM. The expected cluster is chosen according to sum of RSS distances squared between the components of clustering center vector and the RSS samples from available APs. The cluster with the minimum sum of RSS distances squared is supposed that the RPs in this cluster have the most similar RSS characters with the TP. So it tends to improve location precision by using those RPs.

However, only choosing RPs based on RSS leads to the problem that some RPs may have more similar RSS characters with the TP, but the actual positions of which are very far away from the TP. In this circumstance, choosing this kind of RPs may result in relatively large location error. In order to eliminate the negative effect of those RPs, RPs should be chosen again based on their position coordinates through FCM.

Therefore, in the third step, RPs are chosen based on position coordinates. Specifically speaking, the k RPs are divided into c_{PC} clusters based on their position coordinates through FCM. The cluster that includes the most same RPs with the chosen cluster in the second step is chosen.

In the fourth step, the final set of RPs for indoor location is determined. One rule is to choose the intersection of two clusters that are chosen in the second step and the third step. Another rule is to choose the union of the two clusters. Then, the position of TP is estimated with the coordinates of the RPs, which are finally chosen for indoor location. Through the KNN-FCM hybrid algorithm, the negative effect of some RPs chosen in the first two steps is eliminated and the effect of parameter k of taking different values is also weakened.

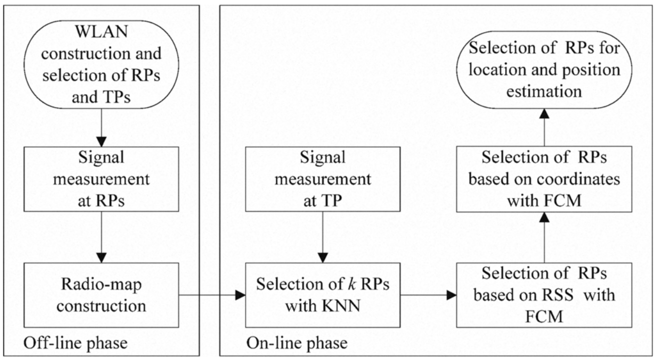


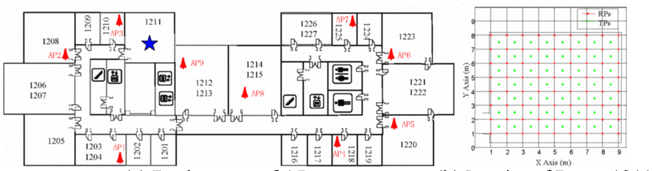
Figure 1. Flow chart of the KNN-FCM hybrid algorithm

III. EXPERIMENT AND ANALYSIS

A. *Setup*

As shown in Fig. 2(a), 9 APs are deployed on the floor with the approximate dimensions of 66m×25m. The performance of the proposed algorithm is evaluated in Room 1211 with the approximate dimensions of 8m×8m, which is marked with a blue star, and the specific environment of Room

1211 is shown in Fig. 2(b). The heights of the office rooms and APs are separately 3m and 0.76m and the width of corridor is 3m.



(a) Deployment of APs (b) Interior of Room 1211

Figure 2. Experimental environment

The test equipments include an ASUS laptop with Windows XP operating system, an integrated Intel PRO/Wireless 3945ABG wireless network card and 9 D-link DWL-2100AP APs that work at 2.4GHz with data rate of 54 Mbps. All the RSS samples are obtained through NetStumbler, which is a kind of public software installed in the laptop.

As shown in Fig. 2(b), there are 72 RPs with 1m gap are depicted as red points and 56 TPs that are depicted as green points. In Room 1211, the laptop can receive the RSS samples from AP1, AP2, AP3, AP8, and AP9. In off-line phase, 360 RSS samples per RP are recorded by the laptop in order to construct the radio-map. In on-line phase, 120 RSS samples per TP are recorded for testing the performance of the proposed algorithm.

B. Experimental design

As analyzed in the second section, the values of parameter k and q have great effect on the performance of KNN. At the same time, the values of parameter k and q determine algorithm complexity based on (1). So it is necessary to analyze the value-taking of parameter k and q. Because the mean of RSS has good location dependence in the experimental environment, the simulation is based on the mean of RSS as well as the most of published papers. The mean error of KNN is shown in Fig. 3 with parameter k and q varies respectively from 1 to 20.

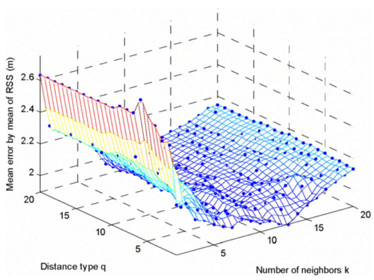
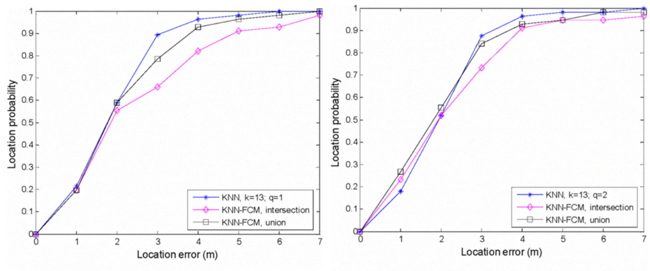


Figure 3. Mean error with different k and q

The result of simulation shows that algorithm complexity of KNN will greatly increase and the location accuracy is not improved when q is greater than 3, so parameter q should be equal to 1 or 2. The mean error tends to be smaller when k ranges from 8 to 13. The minimum mean error is 1.8967m with q equals 1 and k equals 13. When parameter q equals 1 and 2, k equals 13, c_{RSS} equals 2, and c_{PC} equals 3, the performance comparison of KNN algorithm and KNN-FCM hybrid algorithms is shown in Fig. 4.

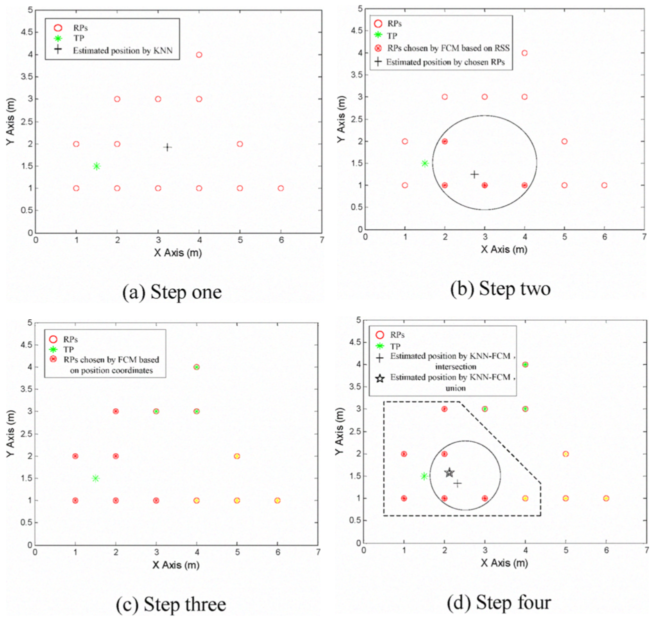


(a) k=13; q=1 (b) k=13; q=2

Figure 4. Performance comparison of the algorithms (k=13; q=1, 2)

In our experimental environment, when KNN is optimal with q equals 1 and k equals 13, although KNN outperforms KNN-FCM when the location error is greater than 2m, KNN and KNN-FCM of choosing union have the same performance when the location error is less than 2m. Furthermore, KNN is difficult to be optimal in practical application, because the workload can be extremely great when the location area is vast and complex. When q equals 2 and k equals 13, KNN-FCM of choosing union is more desirable than the others.

Through computer simulation, a real location process is taken as an example to illustrate the advantages of KNN-FCM hybrid algorithm, which is shown in Fig. 5.



(a) Step one (b) Step two

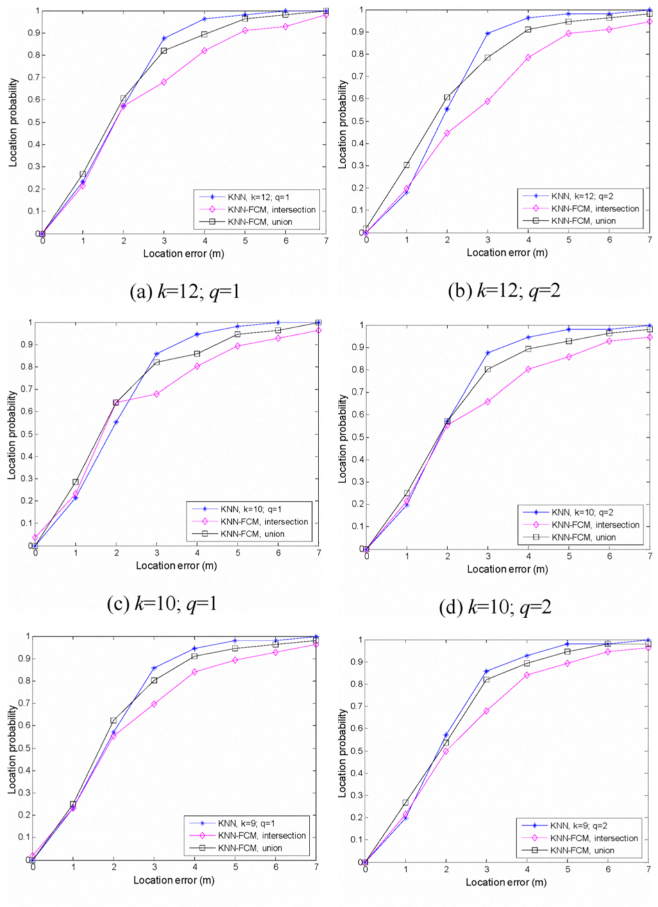
(c) Step three (d) Step four

Figure 5. Location process of the KNN-FCM hybrid algorithm

Let parameter q equals 2, k equals 13, c_{RSS} equals 2, and c_{PC} equals 3. In Fig. 5(a), the coordinate of the TP is (1.5, 1.5) marked with a green "*", RPs that are depicted as red hollow circles are chosen through KNN and the estimated position is the black cross. As shown in Fig. 5(b), RPs marked with red filled circles are chosen through FCM based on RSS, by which the position is estimated and marked with black cross. The result of FCM based on position coordinates is indicated in Fig.

5(c). The RPs are divided into three clusters and the cluster marked with red filled circles is reselected. In Fig. 5(d), RPs included in black circle are the elements of intersection of the two clusters chosen in step two and three, by which the estimated position is marked with black cross in the black circle. At the same time, RPs inside the dashed line are the union of the two clusters and the estimated position is marked with black star. Therefore, under this condition, KNN-FCM hybrid algorithm of choosing union has the best performance and the final results are shown in Fig. 4(b).

In order to test the universality of the algorithm, different values of parameter *k* and *q* are set. Let *k* equals 12, 10 and 9, *q* equals 1 and 2, c_{RSS} equals 2, and c_{PC} equals 3, and the results of simulation with KNN and KNN-FCM hybrid algorithms are shown in Fig. 6.



(a) *k*=12; *q*=1 (b) *k*=12; *q*=2

(c) *k*=10; *q*=1 (d) *k*=10; *q*=2

(e) *k*=9; *q*=1 (f) *k*=9; *q*=2

Figure 6. Performance comparison of the algorithms (*k*=12, 10 and 9; *q*=1, 2)

C. *Analysis of experimental result*

As shown in Fig. 4 and Fig. 6, except the optimal KNN with *k* equals 13 and *q* equals 1, KNN-FCM hybrid algorithm of choosing union outperforms other algorithms when the location error is less than 2m, which is more valuable because the location accuracy within 2m is paid more attention to. So it is worth increasing algorithm complexity to improve the performance, because indoor location system has not a very severe requirement on real time. Furthermore, the example of location process also proves the potential of the KNN-FCM hybrid algorithm and the performance can be also relative to the experimental environment such as building structure, the positions of RPs and TPs. As for the parameters, parameter *q* should be equal to 1 or 2 just as mentioned above. Parameter *k* should not be too small in order to offer enough RPs for following FCM to cluster. At the same time, if parameter *k* is too small, all the RPs tend to be at one side of the TP rather than around it, which will result in large location error. According to parameter *k*, proper values of parameter c_{RSS} and c_{PC} should also be set to make sure that there are enough data samples for clustering. In present experimental environment, the simulation results prove that c_{RSS} should be equals 2 to achieve a better performance and the performance of the algorithm with c_{PC} equals 2 is worse than that when c_{PC} equals 3. According to different environments in practical application, these parameters must be set properly in order to improve the performance.

IV. CONCLUSION

In this paper, a KNN-FCM hybrid algorithm based on mean of RSS in WLAN is presented. Through the algorithm, the negative effect that is generated by KNN can be eliminated and the RPs that are chosen by this algorithm are more suitable for indoor location. In the future, the algorithm will be tested in other experimental environments. The relationships between the parameters and performance of the algorithm, rules of choosing clusters and adaptive parameter selection according to different environments are worth further research.

ACKNOWLEDGMENT

This work is supported by National High Technology Research and Development Program of China (Grant No. 2008AA12Z305).

REFERENCES

[1] J Teruaki Kiasuka, Tsuneo Nakanishi and Akia Fukuda, “Wireless LAN Based Indoor Positioning System WiPS and Its Simulation,” Communications, Computers and signal Processing, 2003. PACRIM. on Vol. 1, 28-30 Aug. 2003. PP: 272-275.

[2] Xie Y Q, Wang Y, Zhu C, et al, “Grid-search-based Hybrid TOA/AOA Location Techniques for NLOS Environments”, IEEE Communications Letters. 2009, 13(4):254-256.

[3] P. Prasithsangaree, P. Krishnamurthy and P. K. chrysanthis, “On Indoor Position Location With Wireless Lans,” The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. 2002: 720-724

[4] Kamol Kaemarungsi, “Efficient Design of Indoor Positioning Systems Based on Location Fingerprinting”, 2005 International Conference on Wireless Networks, Communications and Mobile Computing, 2005: 181-186.

[5] Ahmad Hatami and Kaveh Pahlavan. “Comparative Statistical Analysis of Indoor Positioning Using Empirical Data and Indoor Radio Channel Models,” IEEE CCNC 2006 proceedings. 2006: 1018-1022.

[6] James C. Bezdek, “A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms,” IEEE Transactions on pattern analysis and machine intelligence, vol. pam I-2, No. 1, 1980. January.