Probabilistic Algorithm based on Fuzzy Clustering for Indoor Location in Fingerprinting Positioning Method

Bo Dong
School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Jian Xing
School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Fei Wu*
School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Yan Zou
School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

***Abstract*—Recently, the location of the fingerprint positioning technology is obviously superior to the signal transmission loss model based on the positioning technology, and is widely concerned by scholars. In the online phase, due to the efficiency of the probabilistic distribution matching computation is low and when clustering the position fingerprint database, hard clustering lead to degrading the positioning accuracy, a probabilistic algorithm based on fuzzy clustering is proposed and applied to the indoor location fingerprinting positioning. Compared with hard clustering fusion algorithm, the proposed method has realized the fuzzy partition of the database, makes online positioning phase can effectively search the desired fingerprint data, and improve the positioning accuracy. Experiments show that the algorithm can effectively deal with the problem of the positioning accuracy of hard clustering.**

***Keywords*—Fuzzy Clustering; Fingerprinting Positioning; Indoor Location; RSSI; Probabilistic Algorithm**

I. INTRODUCTION

With the development of wireless network technology, the demand for location-based services is increasing [1]. In the outdoor environment, GPS can satisfy people's demand for location service, but in the indoor environment, the satellite signal is shielded, so the location accuracy is not high. In daily life, handheld device, mobile terminal, domestic robot and other indoor intelligent equipment require the indoor location, and this demand has been spurred research on indoor positioning technology. Thanks to the development of WIFI technology, WIFI positioning technology has been widely concerned because of its low cost, easy to implement, and small influence [2].

Indoor positioning algorithm consists of Arrival of Angle (AOA) positioning method and Time of Arrival (TOA) positioning method, Time Difference of Arrival (TDOA) positioning method and positioning method based on signal strength (RSSI) [3]. With the location technology based on AOA, TOA, and TDOA, the positioning result is reliable and relatively accurate only in the visibility (Line of sight) signal under the condition of dominant [4-6]. In addition, TOA and TDOA positioning method require very accurate clock synchronization between the receiver and transmitter, and the positioning cost is high. In contrast, localization method based on signal strength has the advantages of low cost, simple implementation method, and so on [7]. With a higher precision, the location fingerprint positioning method based on the signal strength becomes the main research direction in the locating method based on signal strength [8].The main process of the location fingerprinting positioning method can be divided into two phases: the offline phase and the online phase [9].

In the offline phase, the signal intensity of the sample is collected at the reference point, and the location fingerprint database is constructed by using the collected data. In the online phase, the signal is acquired in real time, and the algorithm matches with the fingerprint data constructed in the offline stage to get the target position of the coordinate information.

In the actual situation, due to the efficiency of the probabilistic distribution matching computation is low, and when clustering the position fingerprint database, hard clustering lead to degrading the positioning accuracy, therefore, a probabilistic algorithm based on fuzzy clustering is proposed and applied to the indoor location fingerprinting positioning. Compared with hard clustering fusion algorithm, the proposed method has realized the fuzzy partition of the database, makes online positioning stage can search to the fingerprint data effectively, and improve the precision of position determination.

II. INTRODUCTION OF POSITION FINGERPRINT POSITIONING METHOD

The location of the fingerprinting position method based on the signal strength is divided into two phases: offline and online phases [9].

The mission of the offline phase includes the following aspects:

1) Complete the deployment of Access Point (AP) and the determination of reference point.

2) Ensure that the wireless signal emitted by the AP can reach each reference point position.

3) Signal receiving device placed on the reference point location collect the wireless signal from the AP.

4) The server extracts feature parameters (signal strength, variance, etc.), then along with the location information, store it in the position fingerprint database (Radio Map) as a fingerprint.

In the online positioning phase, the mobile terminal collects the AP signal strength information, and the fingerprint matching algorithm is used to find the best match with the fingerprint data constructed in the offline phase, and finally obtains user location information of the mobile terminal. Its principle is shown in figure 1.

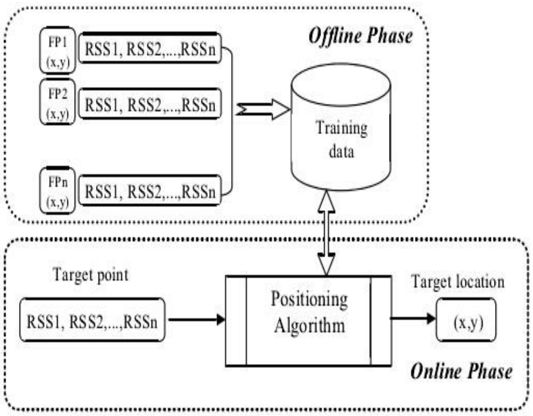

Fig. 1. The principle diagram of the fingerprint location method

III. Theoretical Analysis of the Algorithm

The probabilistic algorithm based on fuzzy clustering is presented in this paper [10]. Firstly, the fuzzy clustering fusion algorithm is used to get the coarse positioning. Then the algorithm selects the appropriate class family based on signal strength acquisition, and uses probabilistic method to obtain the coordinates of the target point, in order to achieve precise positioning location. The basic principle of the algorithm is shown in Figure 2.

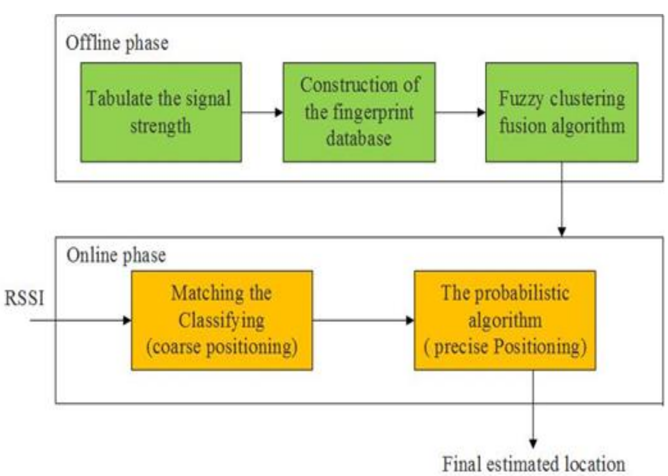

Fig. 2. Basic flow chart of the algorithm

A. Fuzzy clustering fusion algorithm

Fuzzy clustering fusion algorithm is a relatively perfect algorithm in the theory development with the clustering algorithm based on the objective function. Fuzzy clustering algorithm can be used to divide n data $\{x_1, x_2, x_3, \cdots x_n\}$ into C data groups $X_1, X_2, X_3, \cdots X_c (X_1 \cup X_2 \cup X_3 \cup \cdots X_c = X)$, then, get the clustering centers of each group $p_i = (p_{i1}, p_{i2}, \cdots, p_{is}) \in R^c$ to make the minimum sum of squared errors between the samples in each class and the clustering center. Its formula is as follows:

$$J_{FCM}^m(U, P) = \sum_{i=1}^{C}\sum_{j=1}^{n}\mu_{ij}^m d_{ij}^2 \tag{1}$$

There, $\mu_{ij} \in [0,1]$ indicates the degree of x_j is X_i, namely the membership, and must meet the conditions: $\sum_{i=1}^{c}\mu_{ij} = 1$, $\forall j = 1, 2, \cdots, n$; $U = [\mu_{ij}]_{c\times n}$ is the fuzzy classified matrix; *m* is weighted index, if *m* is too large, the clustering effect will be very poor, if *m* is too small, the algorithm will be close to c-means clustering algorithm, so *m* usually assigned to 2 [11]; d_{ij} indicates the distance between the sample x_j and the class center (p_i), also, it can be written as the following form:

$$d_{ij}^2 = \| x_j - p_i \|_A = (x_j - p_i)^T A (x_j - p_i) \tag{2}$$

There, when A is $I_{s\times s}$, d_{ij} is the Euclidean distance.

The minimum of $J_m(U, P)$ is the optimal solution of fuzzy clustering objective function:

$$\begin{aligned}\min\{J_m(U,P)\} &= \min\left\{\sum_{i=1}^{C}\sum_{j=1}^{n}\mu_{ij}^m d_{ij}^2\right\} \\ &= \sum_{j=1}^{n}\min\left\{\sum_{i=1}^{C}\mu_{ij}^m d_{ij}^2\right\}\end{aligned} \tag{3}$$

As the extreme constraint is $\sum_{i=1}^{n}\mu_{ij} = 1$, the Lagrange multiplier method is used to solve:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{4}$$

Make $J_m(U, P)$ is the smallest value, the value of μ_{ij} is:

$$\begin{cases} \mu_{ij} = \dfrac{1}{\sum_{k=1}^{c}\left(\dfrac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} & I_j = \varnothing \\ \mu_{ij} = 0, \forall i \in \bar{I}_j, \sum_{i=1}^{n}\mu_{ij} = 1 & I_j \neq \varnothing \end{cases} \tag{5}$$

Also, p_i can be obtained in a similar way:

$$p_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m x_j}{\sum_{j=1}^{n} (\mu_{ij})^m} \tag{6}$$

If the data set X, the number of cluster categories *m* and the weights *c* are known, optimal fuzzy classification matrix and cluster center can be obtained by the formula.

The pseudo-code of the algorithm is as follows:

Initialization: the number of data samples is n, and the class number is c $(2 \le c \le n)$,the fuzzy weighting exponent m is set to 2, the initial clustering center value $p^{(0)}$,ε indicates the iterative stopping threshold, iteration number $l = 0$;

TABLE I. ALGORITHM PSEUDOCODE

Repeat for l = 1 ， 2……

Step 1：compute the cluster prototypes(means)：

$$p_i^{(l)} = \frac{\sum_{j=1}^{n} (\mu_{ij}^{(l-1)})^m x_j}{\sum_{j=1}^{n} (\mu_{ij}^{(l-1)})^m}, 1 \le i \le c$$

Step 2：compete the distance：

$$d_{ij}^2 = \| x_j - p_i^{(l)} \|_A$$
$$= (x_j - p_i^{(l)})^T A (x_j - p_i^{(l)}),$$
$$1 \le i \le c, 1 \le j \le n$$

Step 3：Update the partition matrix：

For $1 \le j \le n$

If $(d_{ij})^2 > 0$ for all i=1，2，…，c

$$u_{ij}^{(l)} = \frac{1}{\sum_{k=1}^{c} \left(d_{ij}^{(l)} / d_{kj}^{(l)} \right)^{2/(m-1)}}$$

Otherwise

$u_{ik}^{(l)}$ =0 if d_{ik_A} >0 , and $u_{ik}^{(l)} \in [0, 1]$ with $\sum_{i=1}^{c} u_{ik}^{(l)} = 1$

Until $\| U^{(l)} - U^{(l-1)} \| < \varepsilon$

B. Probabilistic algorithm (PM)

The probabilistic distribution location algorithm is used to calculate the matching probabilistic of each point in the wireless signal intensity and location fingerprint, and the maximum of the probabilistic of the reference point is the estimated position of the target [12]. By introducing the probabilistic function of Gauss signal, the signal intensity distribution at any position in the indoor positioning environment is characterized by the expectation and variance of the signal. For the actual collection of wireless signal intensity, it is usually considered to be the general distribution of normal distribution $N(\mu, \sigma^2)$, the statistical parameters are μ and σ^2 [12]. The likelihood function is:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \tag{7}$$

Calculate:

$$\begin{cases} \mu^* = \bar{X} \\ \sigma^{*2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \end{cases} \tag{8}$$

According to the maximum likelihood estimation, μ^* and σ^{*2} are considered as the real value, and the signal intensity of the reference point is considered as the reference point. In the online phase, according to the signal intensity of real-time acquisition, the location of the indoor location fingerprint database is matched in order to achieve precise positioning. By using normal distribution probabilistic formula, the probabilistic of the point (x, y) is obtained:

$$P_i(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(RSSI_i - \mu)^2}{2\sigma^2}} \quad (\sigma > 0) \tag{9}$$

After obtained the wireless signal strength of all APs in the database of probabilistic distribution, take the collected wireless signal strength value into the substitution probabilistic formula, and calculate the probabilistic of each reference point $P(x, y)$ in the class family, as shown in formula (10). There, M is the number of samples for the current class. The maximum probabilistic product is obtained and the coordinate (x, y) is used as the current position estimate coordinates.

$$P(x, y) = \prod_{i=1}^{M} p_i(x, y) = \prod_{i=1}^{M} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(RSSI_i - \mu)^2}{2\sigma^2}} \right) \tag{10}$$

IV. DESIGN OF EXPERIMENT AND RESULT ANALYSIS

In order to verify the effect of this algorithm, the experiment is carried out on the fourth floor of Building 1, local plan is shown in Figure 2. The location area includes the corridor area and the laboratory area. Among them, the laboratory is 6 × 9m, the corridor is 2 × 27m, the location fingerprint interval is 1.5 m; the wall thickness is 33cm, and a total of 138 reference points. The smart router with 5 models named JHR-N845R is used as AP, and 3C Honor is used to collect the RSSI signal of AP, and the sampling frequency is 30samples/min.

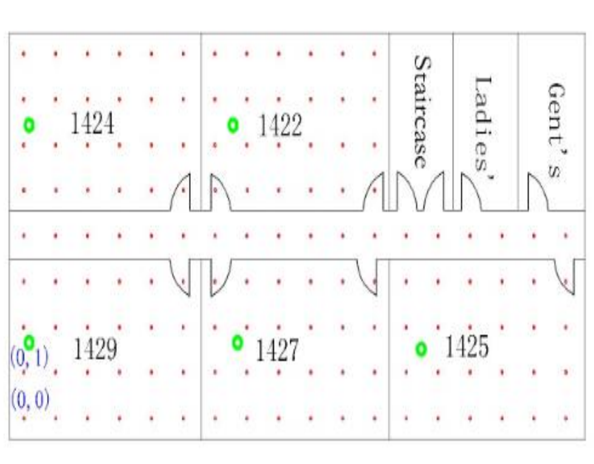

Fig. 3. Experimental environment plan

Figure 3 in the lower left corner is the origin, the vertical direction as y axis and horizontal direction for the X axis to establish coordinate system, and green rings represent the AP location, red dots represent the fingerprint minutiae. In the offline phase, the RSSI values of the reference points are sampled at 50 times, and <the location, the AP name, the RSSI mean, the variance > is recorded in the fingerprint database. The 138 samples are clustered into 6 classes, and the fuzzy weighted index *m* is set to be 2. In the online phase, N times sampling is taken at the target point, and respective calculation are performed in the test position.

In order to evaluate the effectiveness of the algorithm in this paper, the results in the online phase are evaluated, the evaluation parameters are:

$$e = \sqrt{(x_m - x)^2 + (y_m - y)^2} \tag{10}$$

$$e\_aver = \frac{1}{n}\sum_{j=1}^{n} e_j \tag{11}$$

There, (x, y) is actual coordinate, (x_m, y_m) is the calculated coordinates, *e* is the positioning error, e_aver is the multiple positioning error of the mean, *n* is the number of measurement at the current position. The sampling is taken *n* times (*n*=1, 2, 5, 10, 15, 20) at the point (8, 6), the formula (10) and (11) are used to deal with the data obtained, the average error is shown in Figure 4.

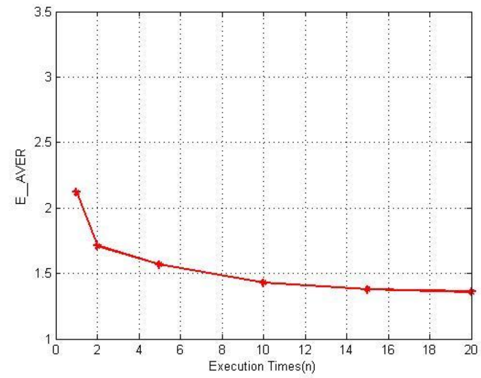

Fig. 4. The average error of N times sampling results

From Figure 4 we can see that in the online positioning phase, when *n*=1, the calculation error of the positioning is large, so the result is not ideal, and the positioning error is reduced after multiple sampling. In the actual situation, we can choose the mean value of the 10 sampling results as the final result.

Again, compared the algorithm proposed in this paper with hard clustering algorithm, the experiment select the hierarchical clustering algorithm (HCA) [13] and K-means clustering (k=6) [14], error results as shown in Figure 5.

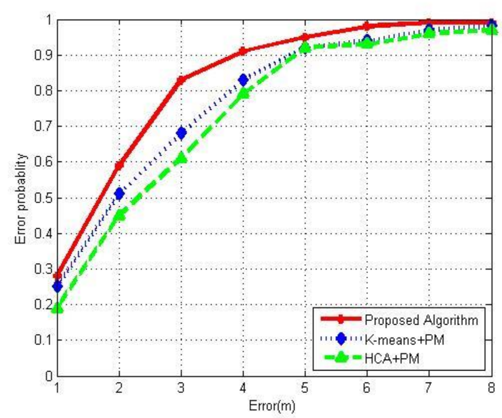

Fig. 5. The cumulative error distribution map

As is shown in figure 5, the results of the experiment show that compared with classic hard clustering algorithm, the proposed algorithm realizes the fuzzy partition of the database, so that the online positioning phase can effectively search to the desired fingerprint data. The probability of the positioning accuracy within 2 meters has reached 60%, and the probability of the positioning accuracy within 3 meters has reached 80%. Compared with the hard clustering algorithm, the positioning accuracy is improved.

V. CONCLUSION

In this paper, a probabilistic algorithm based on fuzzy clustering is proposed, which is used in the indoor location. This algorithm is an effective solution to the unsatisfactory results of the positioning accuracy with the hard clustering algorithm. Experiments show that the algorithm can improve the positioning accuracy by using the method of multiple sampling in the online positioning phase. And, compared with hard clustering fusion algorithm, the probability of the positioning accuracy within 2 meters has reached 60%, and the probability of the positioning accuracy within 3 meters has reached 80%. The proposed algorithm can satisfy the practical application of indoor positioning accuracy.

VI. FUTURE WORK

Aiming at the accuracy of the established database, for the future work, we plan to establish the database by the idea of feedback. Also, a better position of the AP is considered to improve the accuracy of the indoor positioning.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.61272097, 61405045), the Shanghai Municipal Committee of Science and Technology Project (No.13510501400), the Shanghai Municipal Education Commission, A Key Innovation Project (No.12ZZ182).

REFERENCES

[1] Zheng Yang, Chenshu Wu, Yunhao Liu.Location-based Computing: Localization and Localizability of Wireless Networks, Beijing:Tsinghua university press, 2014, pp.111-127.

[2] Dianjun Wang, Hongxing Wei, Fujun Ren. Autonomous mobile robot positioning technology[M]. Beijing:China Machine Press, 62-76, 2013.

[3] Zhang Wenjie, Dong Yuning, Wang Xinheng.Indoor positioning method using RFID and block clustering[J/OL]. Computer Engineering and Applications, 2015-02-16.

[4] Zhou J, Zhang H, Mo L. Two-dimension localization of passive RFID tags using AOA estimation[C]. Instrumentation and Measurement Technology Conference (I2MTC), 2011 IEEE. IEEE, 2011:1-5.

[5] Frank K, Julian L, Christ R. WLAN Mobile Robot Localization with Sensor Fusion[J]. IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. Sep.2009, pp.649-654.

[6] ChengchengJi, Huamei Xin. Indoor fingerprinting localization algorithm based on weighted fuzzy C-means clustering algorithm[J]. Modern Electronics Technique, Vol.36, 2013, 45-47.

[7] Yong Zhang, Jie Huang, Keyu Xu. Indoor positioning algorithm for WLAN based on principal component analysis and least square support vector regression[J]. Chinese Journal of Science Instrument, 2015, 2(36):408-414.

[8] Xiaoxiao Xu, LinboXie, Peng Li. Design of indoor positioning system based on WiFisignal intensity characteristic[J]. Computer Engineering, 2015, 4(41):87-91.

[9] Shengbin Li. Application research on indoor location technology based on WiFi[D]. Nanjing Normal University. 2014.

[10] Li Tang. Indoor WLAN location based on fuzzy cluster KNN algorithm[C]. Harbin Institute of Technology. 2009.

[11] Yilin Du. A fuzzy clustering KNN location fingerprint positioning algorithm [J]. Micro machine and Application, 2012, 31(23):55-58.

[12] Dianjun Wang. Indoor mobile-robot localization system based on WRV[J]. Journal of Tsinghua university. 2011, 12:1849-1854.

[13] Tandong Bai. Classification and Clustering Theory in Data MiningBased Optimization on Fingerprint Indoor Localization[D]. Beijing Jiaotong University. 2014.

[14] Wang Chen, ZhenhongJia, Xizhong Qin, Chuanling Cao. Indoor positioning performance of WLAN based improved K-means algorithm [J]. LASERNAL, 2014, 35(7):11-14.