Journal of Computers Vol. 28, No. 5, 2017, pp. 162-177
doi:10.3966/199115992017102805015

Research on Location Fingerprint Based WiFi Positioning Algorithm

Maoyuan Huang[1], Shaobo Wu[1] and Yinglei Li[1]

[1] Beijing Information Science & Technology University, Beijing, China.
huangmaoyuan@hotmail.com, wushaobo@bistu.edu.cn, yingleili123@126.com

Received 25 October 2016; Revised 13 June 2017; Accepted 26 June 2017

Abstract. Existing WiFi signals are utilized to carry out indoor positioning, and received signal strength RSSI is usually selected as a positioning feature parameter. An RSSI positioning algorithm is divided into a range-based positioning algorithm and a range-free positioning algorithm. The range-based positioning algorithm calculates a distance by utilizing an indoor transmission loss model and is lager in dependency on an indoor model; and the range-free positioning algorithm adopts a location fingerprint algorithm, a fingerprint database is established by only needing to measure RSSI values, each fingerprint is in only correspondence to one location information, information of an unknown location can be estimated by matching the unknown location fingerprint and the fingerprint database, and realization is simple. This article studies the positioning principle of the location fingerprint algorithm, indicates influence factors which may generate errors in the positioning process, comprehensively and deeply analyzes sources of the errors and also indicates the limitation of the existing location estimation algorithm. This article presents a k-means and WKNN based location fingerprint algorithm. According to the location fingerprint algorithm, the fingerprint database is preliminarily established by measuring collected RSSI values for many times and solving a mean and then is trained by utilizing k-means clustering analysis, and some fingerprints with very small similarity are removed; and actually measured fingerprints are matched with the fingerprint database after being trained, so that the accuracy of the fingerprint database is improved, the search space of matching is reduced, and the influence of the fingerprint database on a positioning result is reduced. In the location estimation stage, a new weight coefficient calculation method is introduced, the location accuracy of WKNN is improved.

Keywords: clustering, fingerprint, k-means, positioning algorithm, RSSI, WiFi, WKNN, WLAN

1 Introduction

The positioning of the indoor environment has always been an unresolved problem. Due to the severe attenuation and multipath effects, outdoor location facilities (such as GPS) do not work effectively in buildings. Positioning accuracy is also a problem, GPS may be able to point out where the mobile device in a building, but people want to get more accurate position information, which requires more sophisticated map information and higher positioning accuracy.

We can build a complete set of positioning infrastructure in the indoor environment, but it costs a lot. For example, the spectrum resources occupied by the signal, additional hardware for sensing the positioning signal embedded in the mobile device, anchor nodes installed in a fixed position to send a positioning signal. So we tend to use those wireless devices which have been widely deployed to achieve indoor positioning.

We consider WiFi as the most suitable basic positioning facility to the positioning method based on wireless signal. Now most mobile devices, including smartphones and laptops, have embedded WiFi modules [1, 17]. In fact, WiFi has been used in outdoor positioning and navigation, through the smart phone and the maintenance of the WiFi hotspot location and its corresponding mac address of the database to locate, many companies maintain such databases, such as Google, Apple, Microsoft, and location service providers such as Skyhook [18]. There are some other technologies, such as Bluetooth

[19], RFID, mobile phone base station signals can also be used to achieve indoor positioning, but they are not so popular as WiFi. Mobile devices (such as laptops and mobile phones) that can connect to WiFi can communicate with each other directly or indirectly (through the AP), so it is possible to consider the positioning function simultaneously with the communication function. But WiFi is not designed for positioning, usually only one single antenna, low bandwidth, complex indoor signal propagation environment makes the traditional method such as time of arrival and time difference of arrival (TOA/TDOA) difficult to achieve [20], the method based on arrival signal angle is also difficult to achieve, and install a directional antenna in a WiFi network will cost a lot. Therefore, the location fingerprint method becomes popular in recent years.

The WLAN (wireless local area network) standard is officially formulated at the end of the 20th century, then application of the WLAN becomes very wide and covers many fields, and the WLAN is considered to be one of most promising technologies in the wireless communication field [2] and widely recognized for positioning the indoor positioning target by utilizing the widely existed WiFi networks. The WLAN is considered to be promising by many researchers as having very good development prospect, and many WiFi indoor positioning systems which can be used as reference are also born, more typically, including a Radar system [8], a Horus system [9], a Nibble system [10] and a Weyes system [11], wherein for the Radar, Horus and Weyes systems, by increasing the amount of sample, carrying out an average value or a median solution on the samples or calculating probability distribution of the samples, the influence of instability of receiving signals on the accuracy of data samples is reduced, and thus the interference problem of the wireless network environment is solved. Facing the complex and various wireless network environments, the Weyes system uses a difference value model to establish a signal space, and thus the influence brought to the openness of the wireless local area network is lowered. The amount of sampled sample needed by the solutions is large, and the establishment time of a database is long, resulting in influence on the aspects of the reliability, the usability and the like of the positioning system; together with the requirement for the special positioning terminal, the positioning cost is increased, and therefore, the systems cannot be applied to production and life of people at present and are all only some prototype testing systems. At present, a really matured commercial WiFi indoor positioning system has not been discovered at home and abroad yet, and some actual technical problems, requiring for deep study and solving, still exist in accurate positioning on a mobile terminal based on the WiFi network.

Location fingerprint positioning algorithm [12] is a common indoor positioning algorithm and is performed in an order of training stage and positioning stage; in the first stage, the work is to establish the fingerprint database; and in the second stage, a corresponding algorithm is performed to find out one or more location fingerprints with highest matching with the parameter features of received signals of the positioning terminal, and then the actual location of a user is estimated according to a certain calculation method by using a location coordinate in the fingerprints. The traditional location fingerprint positioning algorithm mainly includes nearest neighbor in signal space (NN) [13], k-nearest neighbor in signal space (KNN) [13-14], k weighted neighbor in signal space [14] and a naive Bayes algorithm [15]. In several traditional location fingerprint algorithms, final positioning results are all simply obtained by matching or mapping the fingerprints based on the signal strength value basically. But, the strength of the signals received by the same location at different moments is not stable: firstly, the complexity of layout of the indoor environment itself ensures AP transmitting signals to occur the multipath phenomenon in the process of arriving a receiver; and in addition, factors including changes of humidity and temperature of air and activities of personnel can also influence transmission of the wireless network signals, and they all can lead to fluctuation of the RSSI value of the signal strength. Now, the positioning environment [3] of many rooms or many floors are not considered in many algorithms, the fingerprint data volume in the database is certainly increased in such environment, and the positioning time can be prolonged if we match the positioning fingerprints with the fingerprint data one by one in the positioning stage. Based on the negative factors, we must consider to effectively filter the RSSI value and carry out rapid selecting method on the location fingerprint data in the database to improve the positioning algorithm in the algorithms, and thus the purposes of improving the positioning precision and the positioning real time are reached.

Aiming at the problem of larger error of the positioning algorithm of the location fingerprint method, the two aspects of improving the accuracy of the fingerprint database and the existing positioning algorithm can be set about. There are many existing solutions of, such as, reducing the influence of the database on the positioning result by introducing a new technology such as kernel function, support vector machines (SVM) [4] and regression analysis [16] to train the database, improved centroid algorithm, improved algorithm of the improved nearest neighbor in signal space algorithm and the like. The literature [4] uses nonlinear discriminant feature extraction based on RSS kernel directly discriminant analysis (KDDA). The features are extracted by KDDA in a kernel space, and RSSI with inaccurate information is recombined, and meanwhile, redundant features and noises are abandoned. An experiment result shows that the method is higher in precision and also remarkable in lowering the data collection data. The literature [5] analyzes the distribution feature of the signal RSSI and indicates that the signals submit to a multi-Gaussian mixture model, and therefore, multi-Gaussian mixture model modeling is adopted to the signal RSSI of the reference points to train the database, and the parameters of the multi-Gaussian mixture model are estimated by adopting an expectation maximization algorithm, so that the precision of the signal strength and the positioning precision of the positioning system are improved. The enhancement type probability based positioning algorithm [17] uses a Gaussian-polynomial continuous distribution curve to carry out least square fitting on RSSI of original signals, so that the accuracy of the RSSI and the precision of the positioning result are improved. The literature [6] proposes an improved UKF (unscented Kalman filter) algorithm to filter the acquired signal strength RSSI, and the influence of RSSI on the positioning precision of autonomous positioning of a robot is reduced. The literature [7] proposes a least squares support vector machine (LS-SVM) based location fingerprint positioning method. The LS-SVM turns the positioning problem to a classification problem, also adopts one-to-one and one-to-many methods to turn the positioning problem to a plurality of existing classification problems respectively, trains the database, lowers the influence of the database on the positioning result, reduces the positioning error by adopting the support vector machines (SVM) and improves the positioning precision.

On the whole, the above algorithms and studies are independently carried out aiming at training or positioning algorithms of the various databases respectively, are effective in respective ranges of study and mainly investigates the training method for the fingerprint database without considering the errors, which can generate, of the positioning algorithms with fingerprint matching.

This paper introduces the basic principle of location fingerprint positioning algorithm and the nearest neighbor algorithm, combine the k-means clustering algorithm and the k weighted nearest neighbor algorithm, presents a k-means and WKNN based location fingerprint algorithm KWKNN. Based on the algorithm, this paper collects a large number of RSSI data from laboratory, and the simulation shows that KWKNN algorithm positioning result is more accurate, and the effectiveness of the algorithms is also proved.

2 Summary on WiFi Positioning Technology

WiFi is high in transmission rate, wide in coverage and strong in mobility and provides various research points for researchers, for example, routes, protocols, security, networks and positioning. Since the rise of WiFi to now, there are infrastructures of WiFi around the world, users can receive the WiFi signals anytime and anywhere and access a backbone network so as to realize internetworking. By utilizing the existing infrastructures, WiFi can carry out wireless positioning by utilizing the features of the signals. The positioning principle of WiFi is shown in Fig. 1. Reference points are preset in a positioning area, i.e. existing WiFi access points AP, when a positioning target moves into the positioning area, the WiFi signals from the access points are received, a value of RSSI is extracted, and the location of the positioning target can be positioned with combination of a special positioning algorithm.

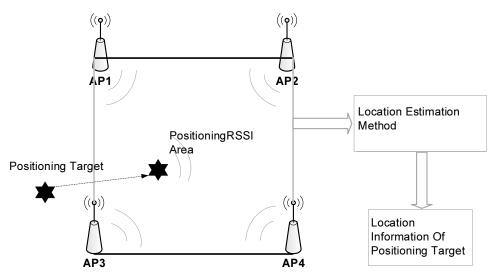

Fig. 1. WiFi indoor positioning principle

2.1 Ranged-based Positioning Algorithm

The wireless positioning technology is usually divided into a range-based positioning method and a range-free positioning method; and the range-based positioning algorithm is usually divided into three stages, i.e. a ranging stage, a location estimation stage and a location correction stage. The ranging stage is carried out by two steps: Step One: starting from an observation point, measuring the distance or angle between the observation point and a reference point; and Step Two: calculating the distance or angle from the observation point to a reference point. In the location estimation stage, a coordinate of an unknown node is calculated by itself by utilizing a trilateral positioning method, a multilateral positioning method or a triangulation measurement method.

2.2 Range-free Positioning Algorithm

The range-free positioning algorithm does not need the ranging stage and lowers the requirements for hardware but increases the positioning error compared with the range-based positioning algorithm. But in some application scenarios without high precision requirement, the range-free positioning algorithm is more applicable. In the indoor environment, the range-free positioning algorithm comprises approximation, the location fingerprint algorithm and the like.

2.3 Location Fingerprint Algorithm

The principle of the location fingerprint based indoor positioning algorithm is shown in Fig. 2, and the location fingerprint based indoor positioning algorithm is divided into a database establishment stage and a location estimation stage. The main part of the location fingerprint algorithm is a database and a positioning algorithm, and the database is established by detecting the strength of WiFi signals by utilizing the existing WiFi infrastructures. A proper positioning area is selected; and in the positioning area, a plurality of sampling points are selected, the location of each sampling point is known, the WiFi signals can be detected at each sampling point, and a signal strength sequence is acquired and stored as a fingerprint in the database. In the WiFi environment, the signal strength RSSI is selected as the feature parameter of the positioning information, and the location fingerprint database is composed of serial RSSI sequences. Each fingerprint is in correspondence to the only location information.

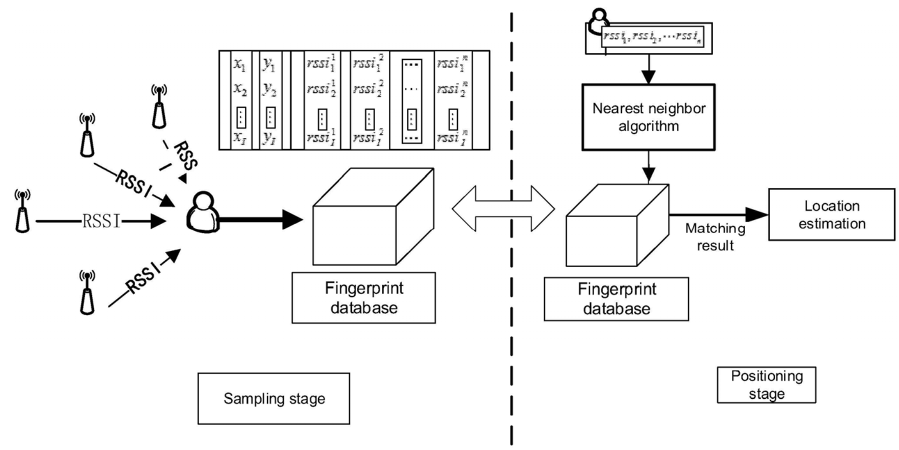

Fig. 2. Principle of location fingerprint positioning algorithm

Location fingerprint database. In the selected positioning area, suppose there are I sampling points, RSSI of n APs can be acquired, so n RSSI values can be collected at each sampling point as a fingerprint, and I fingerprints can be obtained by traversing all the sampling points and stored in the database, shown as a formula (1).

$$FP = \begin{bmatrix} rssi_1^1 & rssi_1^2 & \dots & rssi_1^n \\ rssi_2^1 & rssi_2^2 & \dots & rssi_2^2 \\ \dots & \dots & \dots & \dots \\ rssi_l^1 & rssi_l^2 & \dots & rssi_l^n \end{bmatrix} \quad \textbf{(1)}$$

wherein rssiij expresses the RSSI value, measured at the i th sampling point, of j th AP; and FPi=(rssii1 rssii2…rssiin) means a fingerprint in the fingerprint database. Each fingerprint is in correspondence to the only location, the location is expressed by a binary coordinate (x, y), and then the location information, corresponding to each fingerprint, is shown as a formula (2).

$$Loc = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \dots & \dots \\ x_l & y_l \end{bmatrix} \quad \textbf{(2)}$$

So, the location fingerprint database (LFDB) LFDB=[LOC FP].

Location estimation method for location fingerprint algorithm. According to different location fingerprint expression manners, the location fingerprint algorithm can be divided into a deterministic positioning algorithm and a probability based positioning algorithm. The deterministic positioning algorithm usually adopts the nearest neighbor in signal space (NN) algorithm, K-nearest neighbor in signal space (KNN) algorithm or the like; and the probability based positioning algorithm usually adopts reasoning technology of Bayesian theory to estimate the location information. Concrete classification of the location fingerprint based positioning algorithm is shown in Fig. 3.

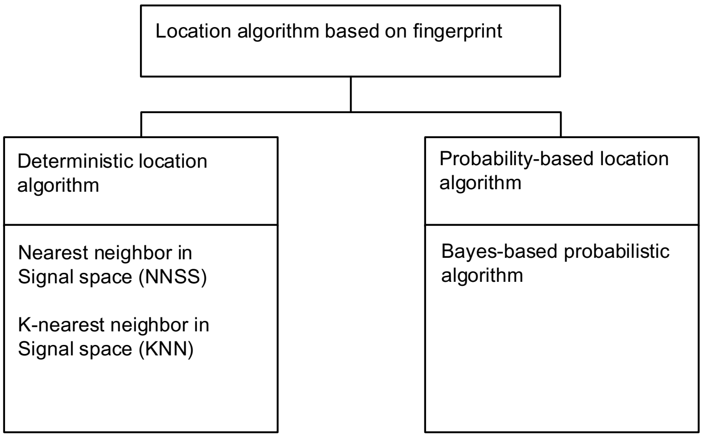

Fig. 3. Classification of location fingerprint based positioning algorithm

(1)Nearest neighbor in signal space algorithm. In the positioning stage, selection of the positioning algorithm is crucial. In the location fingerprint algorithm, the nearest neighbor in signal space algorithm is usually adopted, when a target to be detected enters the positioning area, a finger lf=(rssi1, rssi2, …. rssin) can be obtained, the actually measured finger is matched with finger data in the finger database, and the location information, corresponding to the finger with the maximum similarity, can be used as the location estimation information of the target to be detected.

$$d(lf, FP_i) = \sqrt{\sum_{j=1}^{n} (rssi_j - rssi_i^j)^2} \quad \textbf{(3)}$$

Then the location estimation information of the target to be detected is the location information, corresponding to min(d(lf, FPi)), wherein I = 1, 2,…, l.

The nearest neighbor in signal space algorithm has the advantages of simpleness in deployment and easiness in implementation and has the disadvantages of more singleness in selecting the reference point of the nearest neighbor and easiness in generating positioning errors.

(2)Probability based location estimation method. The probability based positioning algorithm is to use conditional probability to train the fingerprints, establish a probability based fingerprint database and carry out location estimation by adopting the Bayesian reasoning technology in the positioning stage. In this process, the precondition is that the probability distribution of user locations and RSSI probability distribution on each location are known.

Under the initial condition, there is a prior probability P (l) on each location I, and locations in a location collection Loc usually have the same prior probability without more constraint conditions. Hence, the probability based positioning algorithm can acquire the posteriori probability of the locations by adopting the Bayesian principle, namely, the conditional probability of the location l under the condition with known fingerprint lf is as follows:

$$P(l \mid lf) = \frac{P(lf \mid l)P(l)}{P(lf)} = \frac{P(lf \mid l)P(l)}{\sum_{l_k \in Loc} P(lf \mid l_k)P(l_k)} \quad \textbf{(4)}$$

The probability estimation method estimates the location through posteriori probability estimation of the location information, and the location information with the maximum posteriori probability is the estimated location, i.e. the estimated location is as follows:

$$L = \arg\max_{l_k \in Loc} P(l_k \mid lf) = \arg\max_{l_k \in Loc} P(lf \mid l_k)P(l_k) \quad \textbf{(5)}$$

Wherein $\arg\max_{l_k \in Loc} P(l_k \mid lf)$ expresses $l_k \in Loc$ and a value of lk which enables $P(l_k \mid lf)$ to be maximum.

Nibble is a typical indoor positioning system in the WiFi environment, adopts 802.11b protocol and is relatively widely applied. The system is a positioning system with adopting the Bayesian probability reasoning technology early, the quality of measured signals is divided into four levels including a High level, a Medium level, a Low level and a None level, the probabilities on 14 locations in the indoor area are calculated respectively, and then the locations are deduced through a Bayesian network.

The probability based location estimation algorithm has higher performance with the addition of statistical information of the probability distribution, but the probability method needs a large RSSI observation value as the fingerprint to carry out training if high-precision conditional probability distribution needs to be established. As the amount of fingerprint in the fingerprint database is large, although the positioning precision can be improved, implementation is relatively difficult.

3 K-means Clustering and WKNN Based Location Fingerprint Positioning Algorithm

Fig. 4 shows that the location fingerprint algorithm only needs RSSI AS fingerprints in the fingerprint database without needing other feature parameters, and the location can be estimated by matching the fingerprints in the positioning stage. Although compared with the range-based positioning algorithm, the range-free positioning algorithm is lager in positioning error, the location fingerprint algorithm lowers the requirements of the positioning system for the hardware and is low in cost and simple in implementation by adopting RSSI as the feature parameter of the fingerprint database. But, measurement errors can be inevitably introduced in the process of establishing the fingerprint database by collecting RSSI, so that the accuracy of the fingerprint database is influenced, and the subsequent location estimation result can be certainly caused. In the location estimation stage of the location fingerprint algorithm, with the advantage of the nearest neighbor in signal space algorithm of simpleness in implementation, the nearest neighbor in signal space method is selected as the estimation algorithm with the actually measured fingerprint and the fingerprint with the maximum similarity in the fingerprint database as the positioning results and is simpler, but due to singleness of the reference location, the location result of the target to be measured is also inevitably influenced. Therefore, how to improve the accuracy of the fingerprint database and how to reduce the influence of the nearest neighbor in signal space algorithm on the positioning results become one of the urgent problems in the research of the location fingerprint algorithm.

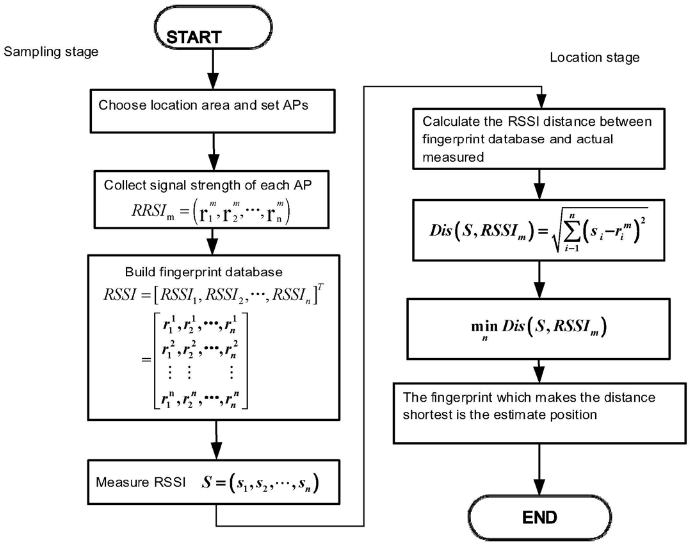

Fig. 4. Algorithmic flow of location fingerprint algorithm

Aiming at the problem that the positioning algorithm of the location fingerprint algorithm has larger error, the two aspects of improving the accuracy of the fingerprint database and the existing positioning algorithm can be set about. There are many existing solutions of, such as, reducing the influence of the database on the positioning result by introducing a new technology such as core function, support vector machines (SVM) and regression analysis to train the database, improved centroid algorithm, improved algorithm of the improved nearest neighbor in signal space algorithm and the like.

On the whole, the algorithms and the researches are independently carried out aiming at training various databases or the positioning algorithm respectively, are effective in respective research range and mainly investigate a training method of the fingerprint database without considering the error which may be caused by the positioning algorithm with fingerprint matching.

Based on the above analysis and researches, this article proposes a location fingerprint algorithm, i.e. the KWKNN algorithm, with combination of k-means clustering and the improved WKNN algorithm and suitability for the WiFi environment. The KWKNN positioning algorithm has the basic idea that the RSSI values which can be received are collected in a data off-line sampling stage to establish the fingerprint database, clustering analysis is carried out on the obtained fingerprints by utilizing a classic k-means clustering method, RSSI information with large deviation is found out to be rejected, and the rest relatively accurate RSSI is utilized to build a new fingerprint database, so that the calculated amount in the fingerprint matching process is lowered; and in the real-time positioning stage, an idea with matching of the actually measured fingerprint and the fingerprint database is adopted to improve the nearest neighbor in signal space algorithm and estimate the unknown location. As shown in Fig. 5.

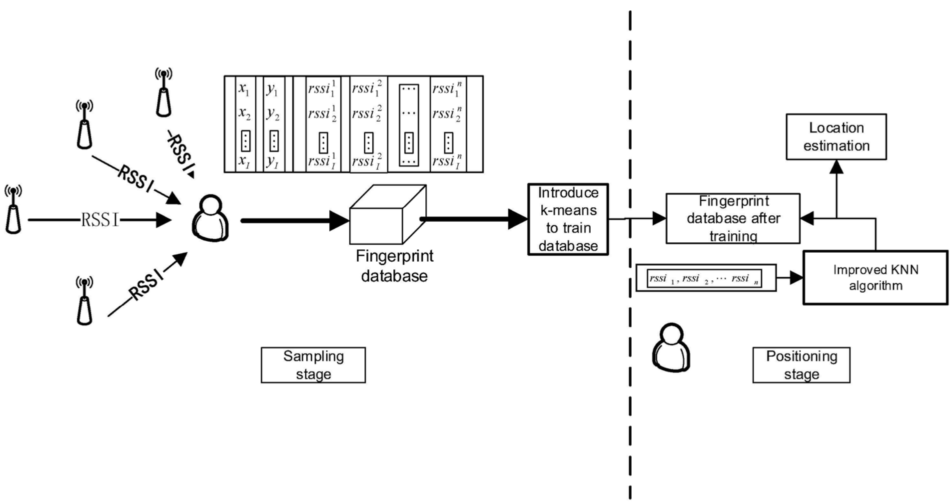

Fig. 5. Working principle of KWKNN algorithm

K-means and WKNN based location fingerprint algorithm (hereafter referred to as a KWKNN algorithm) is an improved location fingerprint positioning algorithm. The location fingerprint algorithm is divided into a fingerprint training stage and a location estimation stage; and in order to improve the positioning precision of the location fingerprint positioning algorithm, this article introduces the k-means clustering algorithm to train the database, lowers the search space of the fingerprint database and improves the existing nearest neighbor in signal space algorithm so as to reduce the positioning error which may be caused by the positioning algorithm.

3.1 K-means Clustering and Weighted KNN Based Location Fingerprint Algorithm

K-means clustering. K-means clustering can also be called k-mean clustering. The k-means clustering method is a clustering algorithm which is widely used and suitable for various data types and is simple in algorithm and rapid in implementation. K-means clustering is a typical distance-based clustering algorithm, takes distance as a measurement of similarity and has the algorithm idea of dividing k subclasses according to the similarity among the existing samples and ensuring the samples with larger similarity to gather together and the samples with smaller similarity to keep away from one another. An experiment shows that, as Fig. 6 shown, k-means clustering can carry out efficient classification to ensure the whole fingerprint database to be divided into different subclasses and reduce the search space of the location fingerprints.

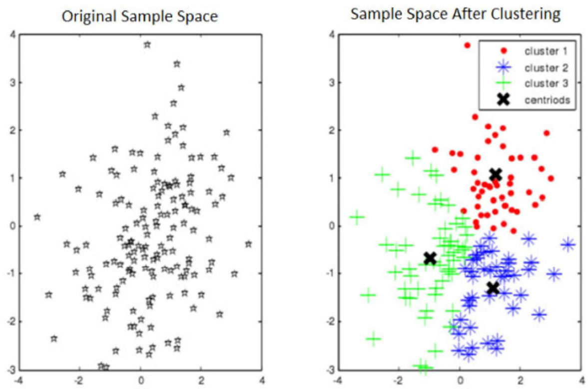

Fig. 6. Comparison diagram of original data in experimental sample space before and after clustering

K-means based location fingerprint database. The first stage of the location fingerprint algorithm is to establish the location fingerprint database. In the WiFi environment, suppose that strength of n WiFi signals, i.e. RSSI, can be detected in some a special indoor area, L sampling points are selected in the positioning area, and the location information of the sampling points is known and is expressed by adopting a two-dimensional space coordinate (x, y). n RSSI values can be observed at each sampling point, and each fingerprint is in one-to-one correspondence to the location of the sampling point of the fingerprint with (rssi1, rssi2, …… rssin) as the fingerprint of the sampling point to form a one-to-one mapping relationship. So, the fingerprint database can be expressed, shown as a formula (6):

$$LFDB = \begin{bmatrix} x_1 & y_1 & rssi_1^1 & rssi_1^2 & \dots & rssi_1^n \\ x_2 & y_2 & rssi_2^1 & rssi_2^2 & \dots & rssi_2^n \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_L & y_L & rssi_L^1 & rssi_L^2 & \dots & rssi_L^n \end{bmatrix}_{L\times(n+2)} \quad \textbf{(6)}$$

wherein LFDB includes an RSSI sequence and the location information. So, the fingerprints and the location information can be separately expressed as a location collection Loc and a fingerprint collection Fp (Fingerprint):

$$Loc=\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \dots & \dots \\ x_L & y_L \end{bmatrix}_{L\times 2} \qquad Fp=\begin{bmatrix} rssi_1^1 & rssi_1^2 & \dots & rssi_1^n \\ rssi_2^1 & rssi_2^2 & \dots & rssi_2^n \\ \dots & \dots & \dots & \dots \\ rssi_L^1 & rssi_L^2 & \dots & rssi_L^n \end{bmatrix}_{L\times n} \quad \textbf{(7)}$$

When the fingerprints in the fingerprint database are trained, fingerprint data needs to be pretreated and is treated by classification by adopting the clustering algorithm, and then the fingerprint database after being trained is KFp. The training process is shown in Fig. 7.

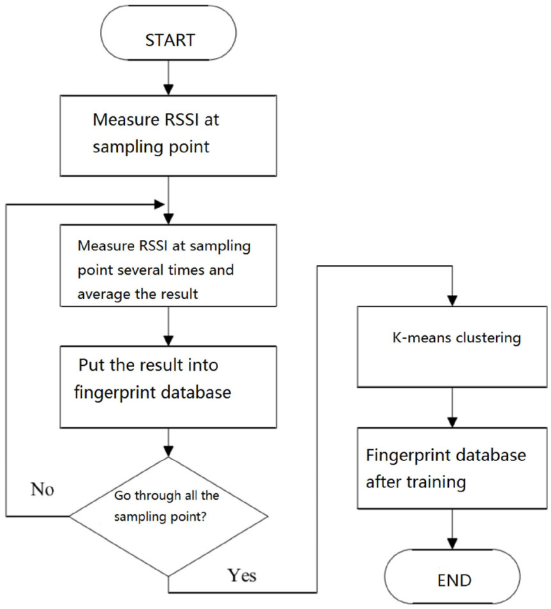

Fig. 7. Training flow of fingerprint database

Improved weighted KNN positioning algorithm. The k-nearest neighbor in signal space (KNN) algorithm is an improved algorithm of the nearest neighbor in signal space (NNSS) algorithm. As being easy in implementation and simple in algorithm, the nearest neighbor in signal space algorithm is selected as the location estimation algorithm in the location estimation stage of the location fingerprint algorithm. The actually measured fingerprint measured in the actual measuring stage is matched with the fingerprint database, the most similar fingerprint is found, and the location of a sampling point, corresponding to the most similar fingerprint, is taken as the estimation location. The reference fingerprints selected by the nearest neighbor in signal space algorithm are relatively single, so that the positioning result is not stable, and a larger error is easy to generate.

Aiming at the problem of singleness of the reference fingerprints, this article applies a weighted k-nearest neighbor in signal space (WKNN) algorithm. In the WKNN algorithm, this article introduces a weight coefficient w, used for expressing the contribution degree of each reference fingerprint. The fingerprint database KFp after being trained is composed of subclasses formed by gathering of a plurality of fingerprints. In the location estimation stage, the actually measured fingerprint lf=(rssi1, rssi2, …, rssin) of the target to be measured is matched with KFp; and the fingerprint database at the moment expresses as KFp=[G1, G2,…., GK], wherein clustering centers in corresponding subclasses are C*=[FP1*, FP2*, …, FPK*] respectively.

In the WKNN algorithm, the usually adopted weighted algorithm directly carries out weighting on the location information, namely,

$$x=\frac{1}{k}\sum_{i=1}^{k}x_i \quad y=\frac{1}{k}\sum_{i=1}^{k}y_i \quad \textbf{(8)}$$

wherein the weight coefficient is wi=1/k, and weights of all the reference locations are the same. While in the WKNN algorithm applied by this articles, selection of the weight coefficient is relevant to the fingerprints in the fingerprint database, and the contribution degree of each fingerprint is calculated through relevant operation of k fingerprints with maximum similarity and is mapped to the corresponding location information to carry out location estimation.

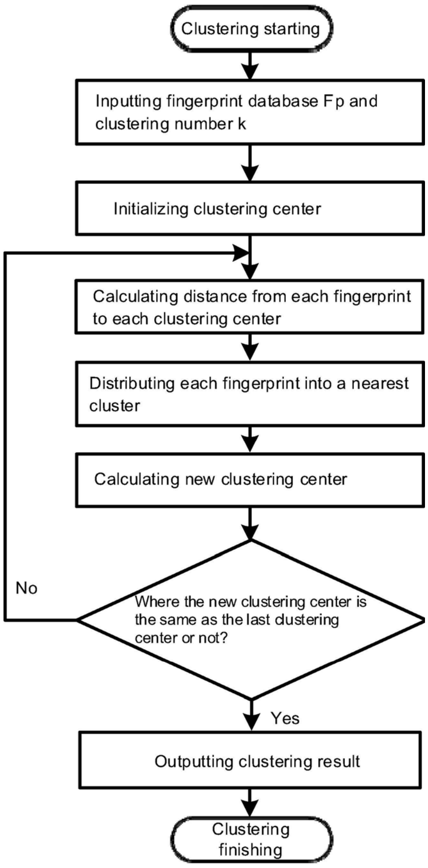

Fig. 8. K-means clustering flow

Suppose that the fingerprint space Fp*, having the maximum similarity with the actually measured fingerprints, have k fingerprints:

$$Fp^* = \begin{bmatrix} rssi_1^1 & rssi_1^2 & \dots & rssi_1^n \\ rssi_2^1 & rssi_2^2 & \dots & rssi_2^n \\ \dots & \dots & \dots & \dots \\ rssi_k^1 & rssi_k^2 & \dots & rssi_k^n \end{bmatrix}_{k\times n} \tag{9}$$

The weight coefficient w_i of each fingerprint in Fp^* is calculated, assume measured fingerprint *lf*= (rssi_1, rssi_2,…, rssi_n), the difference between measured fingerprint and the number I fingerprint in database is $S_{i.}$

$$S_i = \sqrt{\sum_{j=1}^{k} (\mathrm{rsssi}_i^j - \mathrm{rsssi}_j)^2} \tag{10}$$

wherein I = 1, 2, …..k. If $v_i = \frac{1}{S_i}$, the weight coefficient is

$$w_i = \frac{v_i}{\sum_{i=1}^{k} v_i} \tag{11}$$

and then the location of the target to be measured is shown as a formula (12)

$$x_{estimate} = \sum_{i=1}^{k} w_i x_i \qquad y_{estimate} = \sum_{i=1}^{k} w_i y_i \tag{12}$$

The execution process of the WKNN algorithm is shown in Fig. 9 and can be divided into the following six steps:

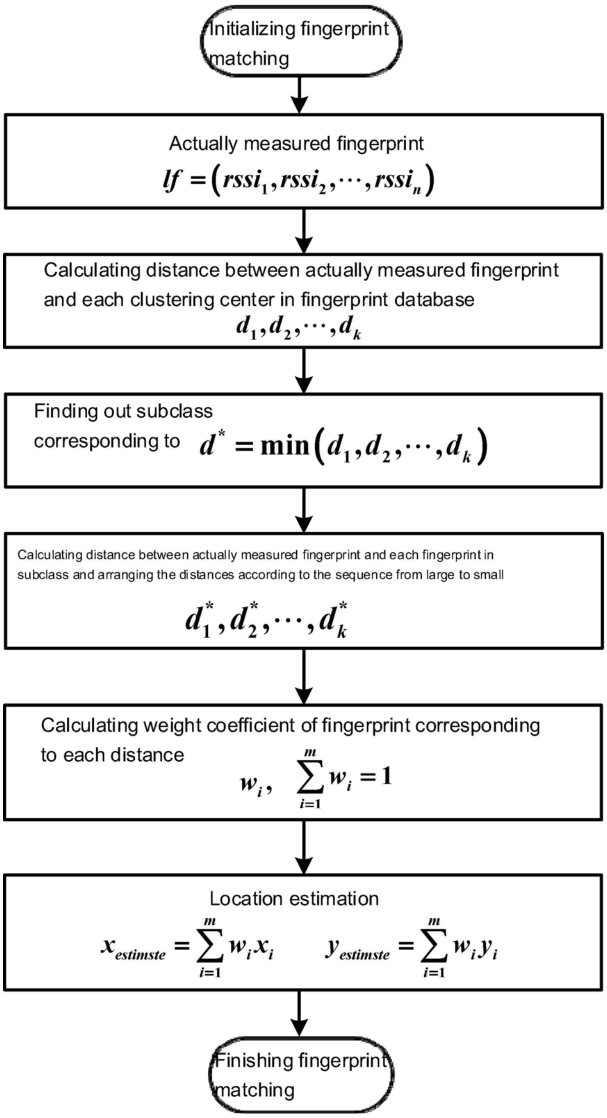

Fig. 9. Location estimation flow of WKNN algorithm

(1)matching the actually measured fingerprint lf=(rssi1, rssi2,…, rssin) with the fingerprint database KFp after being trained, and calculating the distance, denoted as DIS equal to [d1, d2,…, dk], between lf and the center of each subclass;

(2)finding out a subclass, corresponding to min (DIS), and denoting as GSPECIAL;

(3)calculating the distance between the actually measured fingerprint lf and each fingerprint in the GSPECIAL, and denoting as Dis=[d1*, d2*,…, dng*], wherein ng expresses the amount of fingerprint in the GSPECIAL;

(4)arranging Dis according to the sequence from small to large, taking the minimum m distances, and selecting the fingerprints corresponding to the m distances as reference fingerprints with the corresponding location coordinates as the reference coordinates;

(5)calculating the weight coefficient wi of each fingerprint according to formulas (10-11);

(6)calculating the location coordinate (xestimate, yestimate) of the actually measured fingerprint according to the formula (12).

3.2 Algorithm Implementation

On the basis of researching and deeply analyzing the location fingerprint algorithm, this article proposes a k-means and weighted KNN algorithm based location fingerprint algorithm, i.e. KWKNN (k-means Weighted K-Nearest Neighbor in Signal Space) algorithm, k-means clustering is introduced into the fingerprint database training stage for the first time, and the clustering idea is utilized to divide the search space of the fingerprint database; in the location estimation stage, a new weight index calculation method is proposed to ensure the location information estimated by the improved weighted k-nearest neighbor in signal space algorithm to be more accurate. In addition, the actually measured fingerprint is matched with the fingerprint database after being trained by k-means clustering, the nearest subclass is found out, and fingerprints in other subclasses can be removed without considering. The execution process of the KWKNN algorithm is shown in Fig. 10.

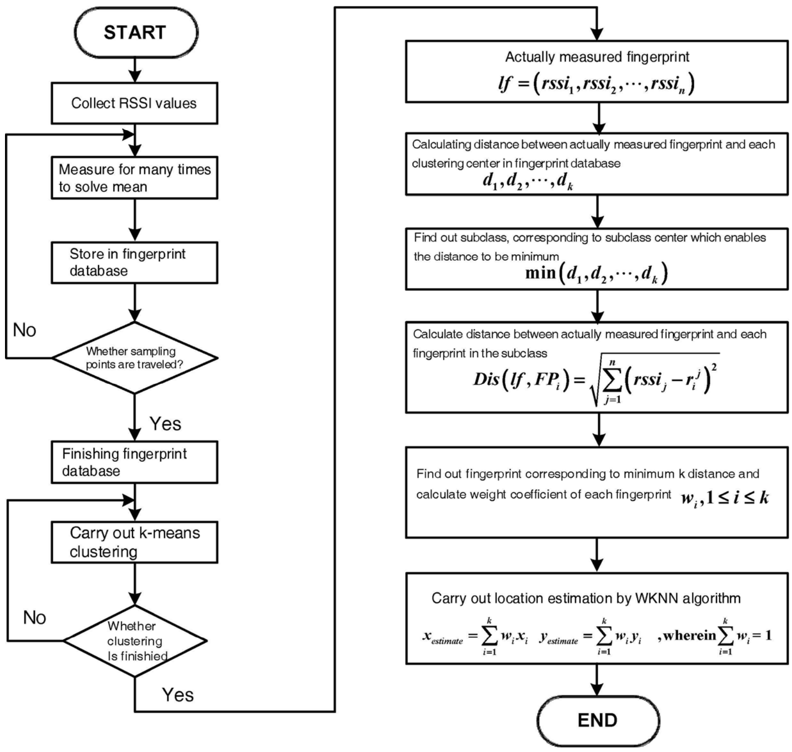

Fig. 10. KWKNN algorithm flow

In the selected indoor area, a plurality of sampling points are selected, RSSI values are measured for many times at each sampling point, a mean is solved to obtain the fingerprint of the sampling point, all the sampling points are traversed to establish the location fingerprint database, and each fingerprint is in correspondence to the location of only one sampling point. An arbitrary location is selected from the positioning area to carry out sampling, and an actually measured fingerprint is obtained with unknown location. The fingerprint with the unknown location information is matched with the fingerprint database, and the location is estimated by adopting the initial nearest neighbor in signal space algorithm and the improved weighted KNN algorithm respectively. The location information obtained by the two algorithms is displayed on a node distribution diagram.

3.3 Algorithm Simulation and Analysis

In order to verify the performance of the KWKNN algorithm, serving as the improved algorithm proposed by this article, this article carries out a simulation experiment on the KWKNN algorithm, adopts a matlab simulation software to simulate the KWKNN algorithm, carries out comparison with the traditional location fingerprint algorithm, and thus verifying the performance of the improved algorithm.
Simulation environment setting. By means of the traditional WiFi infrastructures, No. 614 Laboratory

of Communication Engineering Institute of Beijing Information Science & Technology University is selected as the positioning area. WiFi signals which are distributed at the four corners of the laboratory can be detected in the laboratory, a 8m*8m square area is selected due to the limitation of objective conditions, a fingerprint sampling point is set every 2 meters, and 20 test points are randomly selected. Distribution in the positioning area is shown in Fig. 11.

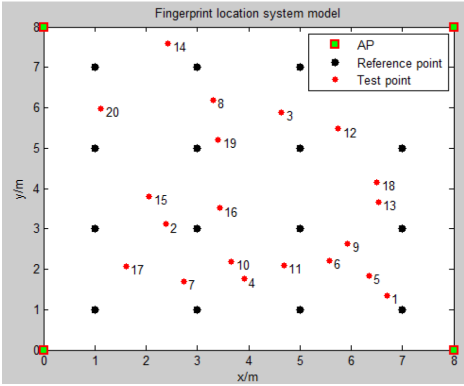

Fig. 11. Sampling point distribution diagram of positioning area

Simulation result analysis. In order to verify the performance of the KWKNN algorithm, serving as the algorithm proposed by this article, the positioning performances of a plurality of actually measured fingerprint verification algorithms are selected in this section. The positioning performances of the algorithms are described through two indexes, one index is positioning error cumulative probability, and the other index is root-mean-square error (RMSE). Suppose that the actual location of the target to be measured is (xactual, yactual), and the location estimated by the positioning algorithm is (xestimate, yestimate), then the calculation formula of RMSE is shown as a formula (13).

$$RMSE = \sqrt{(x_{estimate} - x_{actual})^2 + (y_{estimate} - y_{actual})^2} \tag{13}$$

The Fig. 12 shows comparison of an estimation point and the actually measured point with adoption of the KWKNN algorithm, and a Table 1 shows the value of errors.

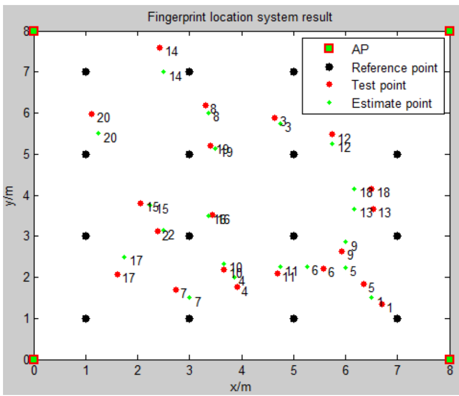

Fig. 12. Comparison diagram of estimate point and test point

Table 1. Error statistics of each point

Test point	Actual x/m	Estimate x/m	Actual y/m	Estimate y/m	Error/m	Average error/m
1	6.700	6.500	1.340	1.500	0.256	
2	2.386	2.491	3.117	3.141	0.108	
3	4.641	4.745	5.879	5.752	0.164	
4	3.915	3.860	1.759	1.996	0.243	
5	6.347	6.001	1.833	2.241	0.536	
6	5.572	5.258	2.216	2.258	0.316	
7	2.738	3.002	1.692	1.500	0.324	
8	3.300	3.355	6.190	6.007	0.191	
9	5.928	6.003	2.633	2.860	0.239	
10	3.668	3.662	2.192	2.331	0.139	0.275
11	4.692	4.748	2.091	2.246	0.165	
12	5.751	5.746	5.480	5.254	0.226	
13	6.530	6.175	3.670	3.657	0.355	
14	2.429	2.500	7.590	7.001	0.594	
15	2.057	2.246	3.796	3.751	0.194	
16	3.434	3.360	3.511	3.502	0.075	
17	1.612	1.748	2.077	2.500	0.444	
18	6.501	6.172	4.150	4.160	0.329	
19	3.405	3.493	5.215	5.144	0.113	
20	1.123	1.246	5.976	5.500	0.492	

Seen from the Table 1, the KWKNN algorithm is adopted to carry out positioning, errors of the 20 points are ranging from 0.10-0.59m, and the average error is only 0.275m. For showing the advantages of the KWKNN algorithm, we compare the KWKNN algorithm with the NN and WKNN algorithms again.

Fig. 13 gives the change conditions of the cumulative probability, obtained by carrying out the original location fingerprint algorithm (nearest neighbor in signal space) and the improved algorithm on the actually measured fingerprint with the actual location to be (4.0, 4.0), of the positioning errors. In the figure, NN represents the original location fingerprint algorithm, the fingerprint database is not trained, and fingerprint matching adopts the nearest neighbor in signal space algorithm; WKNN only represents improvement on the k-nearest neighbor in signal space KNN algorithm, and k-means clustering is not introduced to train the fingerprint database; and the KWKNN algorithm is an improved algorithm proposed by this article, k-means clustering is adopted to train the fingerprint database so as to improve the accuracy of the fingerprint database, and the nearest neighbor in signal space algorithm is improved.

Seen from comparison Fig. 13 under the condition that the actual location is (4.0, 4.0), compared with the original algorithm, the improved algorithm proposed by this article has remarkable increased the convergent speed of a positioning error cumulative probability distribution curve. This illustrates that after the original algorithm is improved, the positioning error is reduced, and the positioning result is more accurate; (2) Shown in the figure, the positioning error cumulative probability distribution curve of the original algorithm starts to converge when the error is about 2.6m, convergence is finished at the position with the positioning error to be 3m, and this illustrates that the positioning error of the original algorithm can reach 3m to the maximum. The positioning error cumulative probability distribution curve of the KWKNN algorithm starts to converge at the position less than 0.8m, convergence is finished at the position about 1m, and this illustrates that the positioning error of the KWKNN algorithm is less than 1m to the maximum. Compared with the original algorithm, the KWKNN algorithm has the maximum positioning error lowered by about 66 percent; and (3) the Fig. 13 shows performance comparison of different improved methods of the original algorithm. By comparing curves of the WKNN and NN algorithms, we can see that after the nearest neighbor in signal space algorithm is improved, and the weight coefficient is introduced, the error of the positioning result is remarkably lowered. Compared with the original algorithm, the weighted k-nearest neighbor in signal space WKNN algorithm has the maximum positioning error lowered by about 50%. Compared with the WKNN algorithm, the KWKNN algorithm increases the training process of k-means clustering to the fingerprint database and has the maximum positioning error lowered by about 25%.

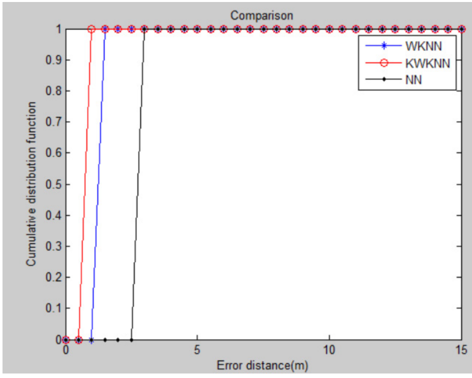

Fig. 13. Performance comparison on NN, WKNN and KWKNN algorithms

4 Conclusion

This article proposes the k-means and WKNN based location fingerprint algorithm, namely the KWKNN algorithm, and the k-means clustering analysis method is introduced into the WiFi positioning technology. Through description of execution process of the location fingerprint algorithm and a corresponding flow diagram, the training stage of the location fingerprint algorithm is elaborated in details. In the stage, the k-means clustering idea is adopted to divide the fingerprints into different subclasses, the actually measured fingerprints are matched with the subclass centers in the positioning process, the fingerprints with smaller similarity are removed, and the rest fingerprints with larger similarity, i.e. more accurate fingerprints, are matched with the actually measured fingerprints. Aiming at the location estimation algorithm, the existing nearest neighbor in signal space algorithm is improved, a new weight coefficient calculation method is introduced, an improved WKNN algorithm is applied, and the positioning error of the WKNN algorithm is about 1.2-1.5m in the same stimulation environment; and compared with the NN algorithm, the maximum positioning error is lowered by 50 percent to the maximum. In addition, compared to the original location fingerprint positioning algorithm with KWKNN, an experiment result shows that the maximum positioning error of the KWKNN algorithm can be lowered by 66 percent, the positioning result is more accurate, and the effectiveness of the algorithms proposed in this article is proved.

Acknowledgements

At the point of finishing this paper, I'd like to express my sincere thanks to all those who have lent me hands in the course of my writing this paper.

This paper was supported by the Beijing Key Laboratory of High Dynamic Navigation Technology (HDN 2015004), the Scientific Research Common Program of Beijing Municipal Commission of Education (KM 201611232011) and the 49th batch of the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References

[1] Y. Wang, Z.H. Ye, Design and implementation of indoor WiFi location system based on android, Electronic Measurement Technology 9(2016) 16-19+25.

[2] I. Galvan-Tejada, E.I. Sandoval, R. Brena, WiFi bluetooth based combined positioning algorithm, Procedia Engineering 35(2012) 101-108.

[3] P. Castro, P. Chiu, T. Kremenek, R. Muntz, A probabilistic room location service for wireless networked environments, in: Proc. the Ubiquitous Computing Conference, 2003.

[4] Z.A. Deng, Y.B. Xu, D. Wu, Indoor positioning in WLAN environment based on support vector regression and space partitioning, Advanced Materials Research 204-210(2011) 1599-1602.

[5] C. Miao, WLAN indoor location system based on gaussian mixture model, Journal of Huazhong University of Science and Technology (Nature Technology) 40(004)(2012) 67-71.

[6] F. Ren, X. Zhang, Y. Jiang, Z. Zun, Application of an improved UKF algorithm in mobile-robot localization system, Journal of Harbin Engineering University 33(10)(2012) 1289-1294.

[7] Y.H. Wei, Y. Zhou, D.L. Wang, Location fingerprint indoor positioning base on LS-VSM, Computer Engineering and Applications Journal (2014) 1-5.

[8] P. Bahl, V.N. Padmanabhan, RADAR: an in-building RF-based user location and tracking system, in: Proc. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, 2000.

[9] M.A. Youssef, A. Agrawala, A.U. Shankar, WLAN location determination via clustering and probability distributions, in: Proc. Pervasive Computing and Communications, 2003.

[10] P. Castro, P. Chiu, T. Kremenek, R.R. Muntz, A probabilistic room location service for wireless networked environments, in: Proc. Ubiquitous Computing. Springer Berlin/Heidelberg, 2001.

[11] L. Victor, K. Xu, G. Cynthia, A difference model for open WLAN environment based location service, in: Proc. IEEE Infocom, 2006.

[12] N. Le Dortz, F. Gain, P. Zetterberg, WiFi fingerprint indoor positioning system using probability distribution comparison, in: Proc. Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012.

[13] Y. Sun, Y. Xu, L. Ma, Z. Deng, KNN-FCM hybrid algorithm for indoor location in WLAN, in: Proc. Power Electronics and Intelligent Transportation System (PEITS), 2009 2nd International Conference on. IEEE, 2009.

[14] D. Wu, Y. Xu, L. Ma, Research on RSS based indoor location method, in: Proc. Pacific-Asia Conference on Knowledge Engineering and Software Engineering, Shenzhen, 2009.

[15] M. Youssef, M. Abdallah, A. Agrawala, Multivariate analysis for probabilistic WLAN location determination systems, in: Proc. Mobile and Ubiquitous Systems: Networking and Services, 2005.

[16] M. Lin, X. Yubin, Z. Mu, Accuracy enhancement for fingerprint-based WLAN indoor probability positioning algorithm, in: Proc. Pervasive Computing Signal Processing and Applications (PCSPA), 2010 First International Conference on. IEEE, 2010.

[17] Y. Wang, Design and implementation of indoor WiFi location system based on android, [dissertation] Nan Jing: Nan Jing University, 2016.

[18] J. Liu, Wireless WiFi location technology and its research and application in wisdom campus, [dissertation] Hunan: Xiang Tan University, 2016.

[19] B. Tian, Research on low-power bluetooth indoor location technology based on filtering fingerprint, [dissertation] Shanxi: Taiyuan University of Technology, 2016.

[20] X. Zhang, Research on indoor location technology based on WLAN location fingerprint and inertial sensor, [dissertation] Shanghai: East China Normal University, 2016.