Location Fingerprint Database Filter Algorithm Based on Multi-Mapping Data Structure

Xun Li[1] & Ze-liang Xie[1]

[1] School of Computer Science and Engineering, Changchun University of Technology, Chang Chun, China

Correspondence: Xun Li, Changchun University of Technology, Changchun Jilin 130012, China. Tel: 86-173-7276-0094. E-mail: yinwuyueye@qq.com

Abstract

The key to improve the precision of location fingerprinting algorithm is to reduce the errors between the position vectors in online locating process and the tag vector of the database in offline process. In view of the large errors caused by the construction of locating tags built from a single vector in traditional offline database, this paper proposed a kind of locating mapping tags for establishing the multi-mapping data structure and the offline database, which adapt better the distribution characteristics of RSSI signals in WIFI. Experiments showed that the offline database constructed by the methods effectively match the current actual position, reduce errors, and significantly improve the matching accuracy.

Keywords: indoor positioning, fingerprint location, offline database, data storage structure

1. Introduction

With the establishment of various large-scale shopping malls, stations and other large buildings, people's desire for the service of indoor positioning accurately is becoming more and more intense. However, GPS positioning has great limitations on indoor environment, which is difficult to do something in the process of indoor positioning (Kong, F., 2017). For instance, Shielding of GPS signals in indoor environment leads to a significant decrease in accuracy of GPS positioning. So in order to improve accuracy, WIFI has been employed gradually in the process of indoor positioning (Figuera, C., 2011).

The mainstream research directions of indoor positioning can be divided into two categories: range-free and range-based (Guo, X., 2018). Firstly, the researchers pay more attention to the algorithm of range-based, which is to quantify the distance between the signal source AP (access point) and the receiving device by the eigenvalues and the attenuation model of the signals. The main algorithms are AOA, TDOA, TOA, AOA, TDOA hybrid positioning, and RSSI (Received Signal Strength Indication). but in the process of indoor positioning research, it is found that compare to other algorithms which need to use additional hardware, such as AOA algorithm needs display antenna, TOA and TDOA need high-precision time synchronization., the location research algorithm of RSSI needn't using additional hardware. Moreover, TOA, TDOA and AOA are not suitable for indoor positioning based on the complex transmitting environment of indoor signals. With the popularity of WIFI technology, WIFI indoor positioning technology has been become a hot research direction of indoor positioning. There are two reasons: the first reason is that WiFi indoor positioning technology does not require additional hardware, and all mobile devices are equipped with WiFi modules, that greatly cost saving. The second reason is that WiFi covers the main scenes of people's life and feasible to use WiFi for indoor positioning.

It is found that the positioning method based on location fingerprint is the most accurate way in the practical test through the research of WiFi indoor positioning technology. Locating fingerprint indoor positioning is divided into two phases: offline phase and online phase. In the offline phase, the offline database can be constructed by using the current position coordinates and obtaining the information about access points (such as phase and signal strength) through accessing network of WLAN wireless as the label values and the eigenvalues of the received signal, respectively. Likewise, in the online phase, the specific coordinates of the current position can be estimated by the most similar position data which matching the eigenvalue of the received signal obtained through accessing network of WLAN wireless with the various data of the offline database.

In order to integrality cover the whole positioning area that will be located, it is necessary to establish a complete

RSSI signal Radio Map, as the target reference node has been selected by planning and measuring in the off-line phase. Each target reference node corresponds to certain quantity of AP signal strength values, and each AP node is independent of each other. The coordinates and signal strength values of each target reference node are stored in the off-line database. However, there was a lot of noise accompanies the process of storing in offline database caused by the time-varying of WIFI signal and the complexity of indoor environment, which affected the process of match by mismatching between the offline database and the instant signal distribution in the online process.

2. Database Filtering Algorithm for Multi-Mapping Data Structure

It is necessary to analyze and collect data separately for each AP node when building an offline database, because of each AP node is independent of each other. As shown in Figure 1, we found the RSSI signal values will jump with sampling time at the same distance and AP (MAC address), which jump is difficult to predict. Later, to solve this situation, offline database was generally constructed by filtering the values of RSSI signal. In Eq. (1), in order to reduce the noise, the results of filtering by the one-dimensional vector $\{rssi_1, rssi_2, rssi_3 \cdots rssi_j\}$ of RSSI are stored in the off-line database and the whole database is filtered by neighbor filter.

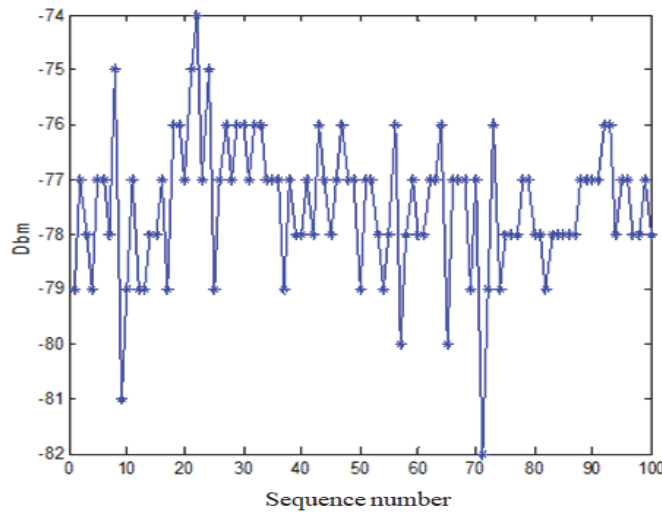$$\overline{rssi} = \frac{\Sigma_1^j rssi_j}{j} \tag{1}$$

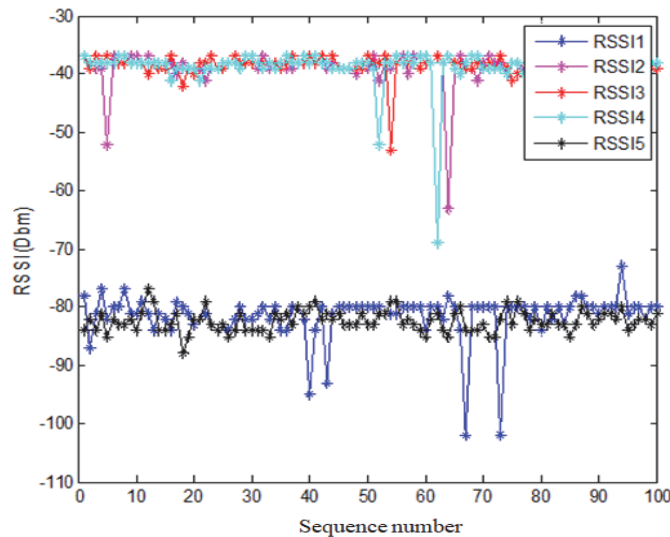Figure 1. Values of RSSI in the same AP and distance

Figure 2. Sequence of RSSI in the different AP

However, neighborhood mean filter is to decrease the noise of the nodes in a small range. It could only filtering the distribution graphs of signal strength that have been constructed. So it doesn't have well efficiency in the process of online. The noises caused by the passage of the sampling signal, when the sampling signal values is matched with the off-line database in the online process, in order to reduce the errors, by matching sampling signal and offline database after filtering the sampling signal. Then, it is necessary to analyze each AP node independently on account of the independence of AP. By sampling the RSSI values of five groups of MAC address AP nodes, as shown in Figure 2. The regulations have found after statistical analysis of the collected results, as shown in Figure 3. The fluctuation of RSSI signal often presents a regional regression state. So we found the fluctuation of the RSSI value is jumped in the fixed value that has the upper and lower bounds. It is barely that the RSSI values will exceed the bounds between upper and lower, it is explained that the fluctuation of the RSSI value is more accurate to meet the normal distribution. In order to adapt the quest of high accuracy, in this paper, a multi-mapping data structure database filtering algorithm presented.
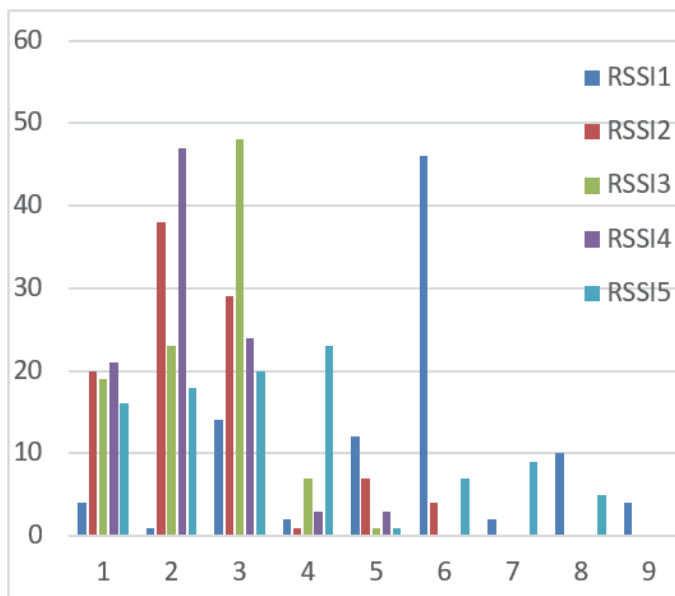

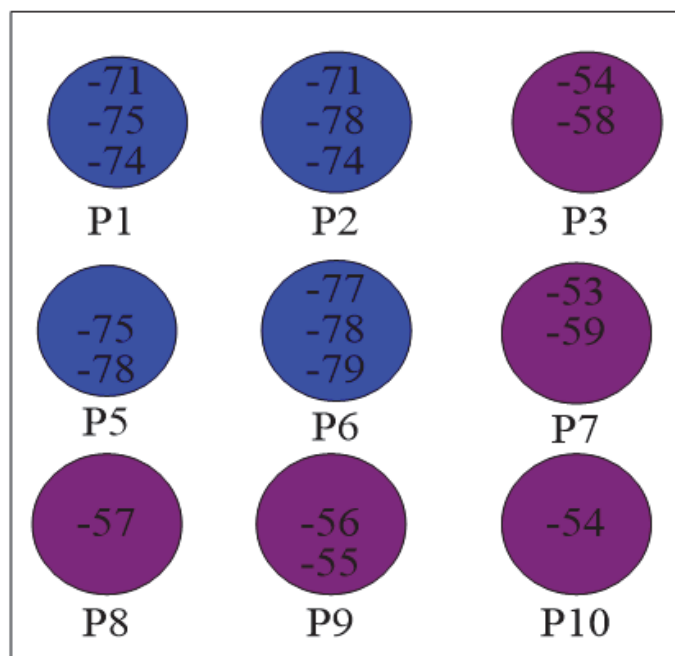
Figure 3. Statistical chart of RSSI

Figure 4. Offline database for the same AP and different coordinates

Since normal distribution showed by signal of RSSI, we used targeted Gaussian filter to filter out the unconventional jump of RSSI signal in the phase that offline database constructed base on fingerprint localization algorithm. For the received values of AP RSSI in each coordinate calculated by Eq. (2), Eq. (3), Eq. (4), where μ and σ denote the average of all RSSI values and the standard deviation of all RSSI values in the test groups respectively. Extracting the signal values with high probability, which values of probability density distribution are in ranges of [0.3,1].

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3}$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{4}$$

Where $RSSI_{gussian}^{x_i}$ and $RSSI_{gussian}^{p(x_i)}$ in Eq. (5) denote the values and sample sets of RSSI signal after Gaussian filtering, respectively. It is worth mentioning that Gaussian filtering mainly filters out the abnormal signal which is free at the upper and lower limits (i.e. the sample groups after Gaussian filtering). According to the probability of sampling signal, the $RSSI_{gussian}^{Sort_{p(x_i)}}$ sequence acquired by sorting from Eq. (6), where $rssi_{x_i}^{n_i}$ denote the frequency of each RSSI signal values. In order to reduce a huge amount of computations in the online phase from a number of RSSI after Gaussian filtering, screening $RSSI_{gussian}^{Sort_{p(x_i)}}$ for gradient descent and the small probability sample $RSSI_{gussian}^{p(x_i)filter}$ after filtering have achieved.

$$RSSI_{gussian}^{p(x_i)} = \{rssi_1, rssi_2, rssi_3 \ldots rssi_i\} \tag{5}$$

$$RSSI_{gussian}^{Sort_{p(x_i)}} = Sort_{p(x_i)}\{rssi_{x_1}^{n_1}, rssi_{x_2}^{n_2}, rssi_{x_3}^{n_3}, rssi_{x_4}^{n_4} \ldots rssi_{x_i}^{n_i}\} \tag{6}$$

$$RSSI_{gussian}^{p(x_i)filter} = \{\frac{RSSI_{gussian}^{Sort_{p(x_i)}}\{n_{p(x_i)}\}}{\sum_1^i n_{p(x_i)}} \geq P_{threshold}\} \tag{7}$$

Where $p_{threshold}$ and $n_{p(x_i)}$ denote the probability threshold of screening and the times of RSSI appear in current probability, respectively. Again, we sort the probability of signal and divide the data for the highest occurrence of probability by the total data in current. The ratio retained after screening in current when it is greater than the probability threshold, if the ratio is smaller than the probability threshold, we divide the current RSSI values added to the next set of RSSI values by total data, then compare the ratio with the threshold until it greater than the threshold. So, we obtain the offline RSSI data sets (i.e. Figure 4) after filtering and storing in the database. There is not one-to-one correspondence about RSSI of each reference position in the same AP node. The storage structure of offline database as shown in Table 1 obtained after collecting all the data of AP through multiple sets of RSSI values correspond to one coordinate. There are various vector groups corresponding the same coordinates when the wireless device matches the collected signal with the offline database in the online phase. As shown in Table 1, the coordinates (X1, Y1) corresponding the values of RSSI such as (-71, -75, -53, -36), (-73, -75, -53, -36), (-77, -75, -53, -36) 18 kinds of combinations and so on. According to the Gaussian distribution of signal, we sort the RSSI combinations of multiple offline datasets, which corresponding the single coordinate. So as to improve the efficiency of online phase for sample data matching with offline database, we firstly match the highest probability in online process. The offline database structure is shown in Table 1.

Table 1. Structure of offline database

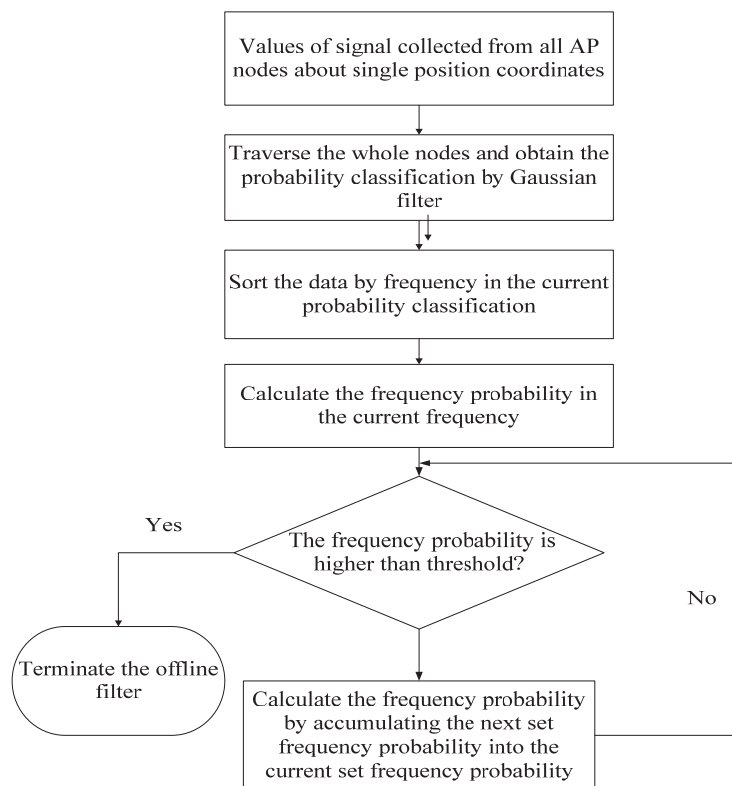| X1,Y1	AP1	AP2	AP3	AP4
MAC	18:64:72:E9:FC:E1	84:1B:5E:6B:BB:D0	8C:AB:8E:01:B2:1C	00:87:36:2C:FD:20
RSSI	-71, -73, -77	-75	-53, -57	-36, -33, -39

Figure 5. Procedure of Fingerprint algorithm based on Multi-mapping database

3. Performance Analysis and Experimental Results

In this section, we theoretically analyzed and simulated the algorithm. We obtained the RSSI signal from WIFI by the NativaWifiApi interface that read all the RSSI information of the AP in the wireless network card of laptop computer. According to the algorithm, in order to receive RSSI signal of all the AP nodes at the same coordinates of position, we firstly traversed all the AP nodes to calculate the classifications of probability after Gaussian filtering, and then sorted the frequency. Secondly, we obtained the frequency probability by calculating the each classifications of probability. Secondly, we compared the frequency probability with the threshold frequency, if the frequency was higher than the threshold frequency, stopped filtering and stored the data in the database, if not, we added the frequency of the next set probabilities into it and compared with the threshold frequency until it was higher than the threshold frequency. The specific process was illustrated in Figure 5. We conducted the RMS error (RMSR) to analyze the performance between our algorithm and Neighbor mean filtering algorithm. It can be seen from Figure 6 that the line graph of Multi mapping filtering algorithm have less RMSR than Neighbor mean filtering algorithm under the same condition, which means our algorithm has the more performance in terms of positioning error. So multi-mapping data structure can effectively reduce the influence of time-varying WiFi signal and reduce the error of positioning. More over RMSE has reduced from 3.3 to 1.4, which maintaining the fitting better between the offline database and real environment.

Luo, J. F., Fu, Y. X., & Wang, Y. G. (2011). WLS location algorithm using RSSI measurement. *Journal of Huazhong University of Science and Technology (Natural Science Edition), 39*(11), 34-38.

Sun, S., Dong, K., Xiu, J. J., & Liu, Y. (2016). A Passive Source Localization Algorithm with Multiple Moving Observers Using TDOA / GROA Measurements Based on CWLS. *Journal of China Academy of Electronics and Information Technology, 11*(5), 540-546.

Zhang, F., Chen, H., & Li, Y. (2017). TDOA-DOA Mapping Using Multi-kernel Least-Squares Support Vector Regression. Journal of Data Acquisition & Processing.

Copyrights