

# Indoor Localization Based on Fingerprint Clustering

Shi Chen<sup>1,2</sup>

<sup>1</sup> School of Computer Science and Technology, Nanjing University of Technology, Nanjing, China

<sup>2</sup> School of Information Engineering, Yancheng Teachers University, Yancheng, China

Correspondence: Shi Chen, School of Computer Science and Technology, Nanjing University of Technology, Nanjing, China.

Received: December 20, 2020 Accepted: December 30, 2020 Online Published: December 31, 2020

doi:10.5539/nct.v5n2p40

URL: <https://doi.org/10.5539/nct.v5n2p40>

## Abstract

With the rapid development of the huge promotion of the Internet and artificial intelligence, the demand for location-based services in indoor environments has grown rapidly. At present, for the localization of the indoor environment, researchers from all walks of life have proposed many indoor localization solutions based on different technologies. Fingerprint localization technology, as a commonly used indoor localization technology, has led to continuous research and improvement due to its low accuracy and complex calculations. An indoor localization system based on fingerprint clustering is proposed by this paper. The system includes offline phase and online phase. We collect the RSS signal in the offline phase. We preprocess it with the Gaussian model to build a fingerprint database, and then we use the K-Means++ algorithm to cluster the fingerprints and group the fingerprints with similar signal strengths into a clustering subset. In the online phase, we classify the measured received signal strength (RSS), and then use the weighted K-Nearest neighbor (WKNN) algorithm to calculate the localization error. The experimental results show that we can reduce the localization error and effectively reduce the computational cost of the localization algorithm in the online phase, and effectively improve the efficiency of real-time localization in the online phase.

**Keywords:** Indoor localization, Clustering, Fingerprint technology, WKNN

## 1. Introduction

With the rapid increase of data services and multimedia services, people's demand for localization and navigation is increasing. When people enter a complex indoor environment, they are often used for outdoor real-time localization. Global positioning system (GPS) (Ahn, Song, Sung, Kim, & Lee, 2010) technology cannot be positioned normally indoors, but the emergence of indoor localization and navigation technology just makes up for the shortcomings of GPS and other technologies. For example, in environments such as airport halls, exhibition halls, warehouses, supermarkets, libraries, underground parking lots, mines (Liu, Wang, Zhang, & Han, 2019), etc. It is often necessary to determine the indoor location information of mobile terminals or their holders, facilities and objects.

At this stage, the more common indoor localization technologies include Bluetooth localization (Kuxdorf-Alkirata & Brückmann, 2016), Ultra wide band (UWB) indoor localization technology (Xu, Shi, & Wu, 2007), Radio frequency identification (RFID) localization (Buffi, Nepa, & Cioni, 2017), ZigBee indoor localization technology (Arif, Niaz, Shabbir, Zafar, Hassan, & ur Rehman, 2018), ultrasonic localization (Li, Ji, & Zhang, 2010), infrared localization (De-La-Llana-Calvo, Lázaro-Galilea, Gardel-Vicente, Rodríguez-Navarro, & Bravo-Muñoz, 2018) and Wi-Fi localization (Sun, Xue, Yu, Tang, & Lin, 2018; Zegeye, Amsalu, Astatke, & Moazzami, 2016), etc. Although there are many types of indoor localization technologies, in fact, many algorithms are universal.

The proximity method is a rough localization method. The simplest way is to directly select the location of the AP with the highest signal strength, and the localization result is the location of the currently connected Wi-Fi hotspot stored in the hotspot location database. Geometric feature method is divided into trilateral measurement method, triangulation method and hyperbolic localization method. The trilateral measurement method first uses time of arrival (ToA) (X. Zhu, W. Zhu, & Chen, 2018) to measure the distance from the receiving device to each signal source, and then calculates the location of the receiving device through the location of the signal source. The triangulation method uses angle of arrival (AoA) (Zheng, Sheng, Liu, & Li, 2018) to measure the relative position or angle between the receiving device and the signal source, thereby calculating the position of the receiving

device. The hyperbolic localization method uses time difference of arrival (TDoA) (Cho, Yeo, Choi, Park, & Lee, 2012; Zhao, Li, Hao, Wan, & Wang, 2019) to measure the difference in the distance between multiple signal sources of the receiving device to obtain multiple hyperbolas. The intersection of the hyperbola is the location of the receiving device. The fingerprint location method (Poulose & Han, 2020) is to measure the signal characteristics at each position in advance and store it in the fingerprint database. When localization, the current signal characteristics are matched with those in the fingerprint library to determine the location.

Fingerprint location method is currently a popular indoor location algorithm. The fingerprint algorithm is usually divided into offline sampling data phase and online localization phase. On this basis, to improve the localization accuracy, in the offline stage, the collected RSS signals are preprocessed by Gaussian filtering, and then the fingerprints are clustered using the K-Means algorithm. In the online localization stage, we classify the RSS signals measured at the target localization point first, and then use the WKNN algorithm for localization calculations. Our main contributions are as follows:

- 1) We use the Gaussian model to filter the RSS signal values with high probability and remove the singular values, which effectively improves the stability of the data.
- 2) Our method of clustering fingerprints has effectively improved the localization accuracy, and the localization error after clustering is about 2.24m.
- 3) Our method reduces the calculation time in the online localization phase, and effectively improves the efficiency of localization.

## 2. Method

Our experiment includes two phases: offline phase and online phase. The offline stage is responsible for collecting data, processing data and clustering the fingerprint database, and the online stage realizes regional localization.

### 2.1 Offline Phase

#### (1) Data collection

In different physical locations, the expressive power of signal strength is different, that is to say, the signal strength of access point (AP) received at each point is different. By detecting the signal value of APs arranged in the localization environment in advance at each sampling point, extracting the signal strength as a localization feature attribute, training it into a mapping relationship with the physical location, and constructing a corresponding location fingerprint database. Therefore, we need to deploy APs and plan sampling points in the target site, and then collect RSS signals at each sampling point.

#### (2) Preprocess

Due to the complex indoor environment and numerous obstacles, wireless signals will undergo refraction, reflection, diffraction, etc., resulting in multipath transmission effects, which will cause the RSS value of the signal measured at the same location at the same time to be unstable and remain in a fluctuating state. This phenomenon will This leads to inaccuracy of training data in the offline phase, which affects the accuracy of online localization. Therefore, the data must be filtered to reduce noise, effectively eliminate interference, and ensure data stability.

A common preprocessing method is the mean filtering algorithm, that is, the mean value of the RSS signal measured at each sampling point for a period of time is taken to reduce the random error caused by unstable factors. However, this method does not eliminate the singular value, but averages it to each component of the RSS, which cannot fully represent the signal characteristics of the location. So we use another method: Gaussian filtering. The method of Gaussian filtering is to eliminate the small probability signal through the Gaussian model, retain the high probability signal, and then average the high probability signal as the final RSS signal value. The singular value signal is a small probability signal, so it can achieve a noise reduction effect.

Suppose we deploy N APs and plan M sampling points. We sample for a period of time at each sampling point, and then use the Gaussian model to filter the RSS signal value with high probability. The Gaussian function expression is as follows:

$$f(RSS) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(RSS-\mu)^2}{2\sigma^2}} \quad (1)$$

where  $\mu$  and  $\sigma$  are the mean and variance respectively.

We keep the signals with a probability value greater than 0.6, and then average these signals as the fingerprint of the sampling point and store it in the fingerprint database. The representation of the fingerprint of each sampling

point in the database is  $R_i = [RSS_{i1}, RSS_{i2}, \dots, RSS_{iN}]$ , and its corresponding label in the database is the coordinate of the sampling point  $(x_i, y_i)$ ,  $1 \leq i \leq M$ .

### (3) Fingerprint Clustering

After the fingerprint database is constructed, the structure of the map is the one-to-one mapping relationship between sampling points and their RSS fingerprints. Usually people use KNN algorithm or WKNN algorithm to locate the target directly. Due to the spatial distribution of RSS signals, we know that the Euclidean distance of RSS fingerprints of two adjacent points is small. The KNN algorithm and WKNN algorithm select multiple sampling points with the closest Euclidean distance from the RSS fingerprint of the target location as the target location basis. To ensure the accuracy of the selection of these sampling points, we cluster the fingerprint data of the map sampling points. In this way, the selection range of sampling points is reduced, the accuracy of online localization is higher, the time required for localization is shorter, and the localization efficiency is improved.

Here we use the K-Means++ clustering algorithm. Specifically, we set K clusters, and then start to select K cluster centers. The basic idea of selecting the initial centroid is to keep the distance between the initial cluster centers as far as possible. First, from the M fingerprint data, randomly select one fingerprint data as the centroid  $C_1$  of the first cluster. Then the shortest distance between each sample and the existing cluster center is calculated, which is represented by  $D(x)$ . Then, the probability  $P(x)$  of each sample being selected as the next cluster center is calculated. The formula of  $P(x)$  is as follows:

$$P(x) = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (2)$$

Finally, we use the roulette method to select next cluster center  $C_2$ . Then repeat the selection of cluster centers until K cluster centers are selected. Traverse each sample after selecting K cluster centers, and calculate the Euclidean distance between the sample and the K cluster centers. Then divide the sample into the cluster with the smallest Euclidean distance. Then the traversal is completed, the center of each cluster is recalculated, and the traversal operation is continued until the cluster allocation no longer changes.

### 2.2 Online phase

We measured the RSS value of AP at the test point  $[RSS_1, RSS_2, \dots, RSS_N]$ . First, we classify the measured RSS data into cluster  $C_k$ , where  $1 \leq k \leq K$ . This reduces the localization area. Then use WKNN algorithm for localization. Assuming that there are T fingerprint samples in this category, we calculate the RSS value of the test point and the Euclidean distance of the fingerprint in the classified area, the formula is:

$$d_t = \sqrt{\sum_{i=1}^N (RSS_1 - RSS_{ti})^2} \quad (3)$$

where  $1 \leq t \leq T$ . Let  $(x_1, y_1)$ ,  $(x_2, y_2)$  and  $(x_3, y_3)$  be the three closest points in the cluster  $C_k$  with the closest Euclidean distance to the test point fingerprint, and the distances are  $d_1$ ,  $d_2$  and  $d_3$  respectively. Then our formula for calculating the weight is:

$$\omega_i = \frac{\frac{1}{d_i}}{\frac{1}{d_1} + \frac{1}{d_2} + \frac{1}{d_3}} \quad (4)$$

where  $i = 1, 2, 3$ . So we get to the final localization result  $(x_*, y_*)$ :

$$(x_*, y_*) = \omega_1(x_1, y_1) + \omega_2(x_2, y_2) + \omega_3(x_3, y_3) \quad (5)$$

This method of first clustering fingerprint data and then localization is also applicable to complex indoor environments, especially when an area is divided into several sub-areas by walls. For example, in the environment shown in Figure 1, due to the influence of the wall on the signal, in this case, we cluster the fingerprints into 4 categories, and then the method of localization can effectively reduce the error and improve the localization accuracy.

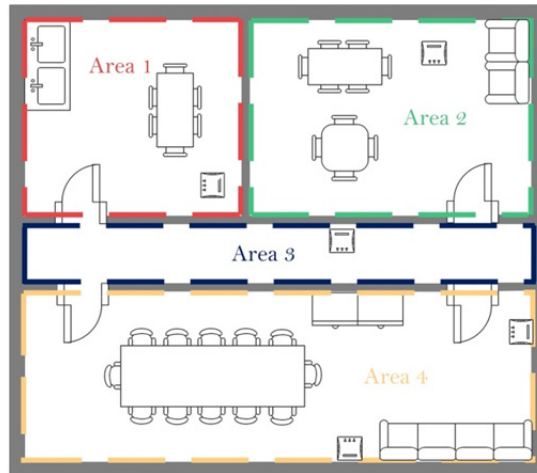


Figure 1.

### 3. Experiment and Analysis

Our experimental scene was in an underground parking lot, and we chose an area of 650 square meters. We deployed 16 APs. We planned a sampling point every 1 meter, planned a total of 702 sampling points, and collected the RSS signal for two minutes at each sampling point. Then for the RSS signal of each sampling point, we perform preprocessing, use the Gaussian model to filter out signal values with a probability greater than 0.6, and average these filtered signal values as the fingerprint of the sampling point. In the measurement, due to the large space, not all AP signal values can be collected at every sampling point. We consider the uncollected signal value to be -100 dBm. In this way, a fingerprint database is constructed.

We cluster these fingerprints using K-Means++ algorithm. For the setting of the K value of the number of clusters, we tried to set K to 2, 3, 4, 5 and 6. We randomly select 200 test points on the map, measure their RSS signals, and perform Gaussian filtering to get the average. Then the RSS signal is classified, and the localization error is calculated by WKNN algorithm after classification. As shown in Figure 2, under our data map, clustering into four categories is the best. In the case of  $K = 4$ , we average the accuracy error of 200 test points, and the error accuracy is about 2.24 m.

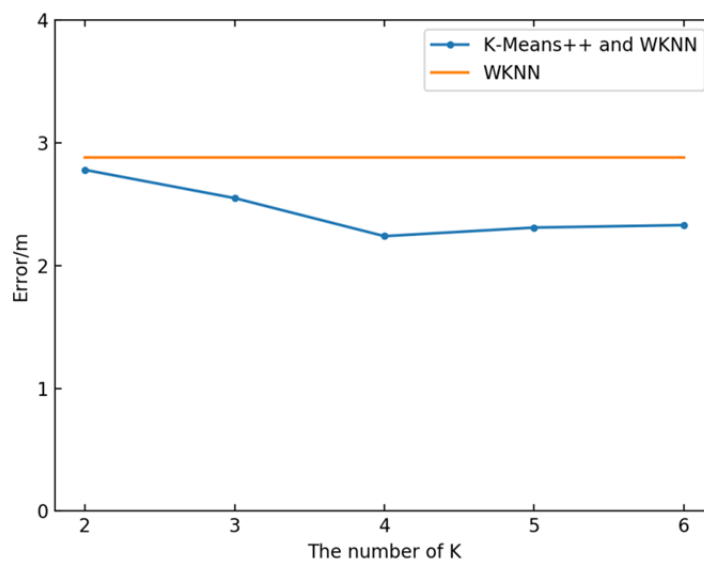


Figure 2.

In addition, we will compare the localization algorithm with clustering and the localization algorithm without clustering. We also directly perform WKNN algorithm on 200 test points to calculate the localization error. From Figure 2, we can see that the localization error without clustering algorithm is significantly higher than our method. Therefore, this illustrates the effectiveness of our method. At the same time, in the online localization stage, we first classify the RSS signals of the test points, which effectively reduces the calculation cost of the WKNN algorithm and improves the efficiency of real-time localization.

#### 4. Conclusion

In the traditional fingerprint localization method, fingerprints are collected in the offline phase and positioned in the online phase. On the basis of traditional fingerprint localization, we propose two goals: reducing errors and improving localization efficiency. First, we preprocess the collected RSS signal. The Gaussian model is used to calculate the probability of the RSS signal value, retain the RSS signal with a probability greater than 0.6, and average these signals to eliminate singular values. Then we use the K-Means++ algorithm to cluster the fingerprints, and group the similar RSS signals into the same cluster. K-Means++ algorithm effectively solves the problem of random initial points, which makes the clustering effect better and the convergence speed faster. Our experimental results show that our method can not only improve the accuracy of localization, but also reduce localization time and improve localization efficiency.

#### References

- Ahn, J., Song, H. Y., Sung, S., Kim, J. B., & Lee, Y. J. (2010, October). An efficient GPS parameter prediction method using GPS ephemeris patterns for self-assisted GPS. In *ICCAS 2010* (pp. 2431-2434). IEEE. <https://doi.org/10.1109/ICCAS.2010.5670251>
- Arif, S. A., Niaz, M. H., Shabbir, N., Zafar, M. H., Hassan, S. R., & ur Rehman, A. (2018, August). RSSI based trilateration for outdoor localization in ZigBee based wireless sensor networks (WSNs). In *2018 10th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 1-5). IEEE. <https://doi.org/10.1109/CICN.2018.8864943>
- Buffi, A., Nepa, P., & Cioni, R. (2017, September). SARFID on drone: Drone-based UHF-RFID tag localization. In *2017 IEEE International Conference on RFID Technology & Application (RFID-TA)* (pp. 40-44). IEEE. <https://doi.org/10.1109/RFID-TA.2017.8098872>
- Cho, S. H., Yeo, S. R., Choi, H. H., Park, C., & Lee, S. J. (2012, October). A design of synchronization method for TDOA-based positioning system. In *2012 12th International Conference on Control, Automation and Systems* (pp. 1373-1375). IEEE.
- De-La-Llana-Calvo, Á., Lázaro-Galilea, J. L., Gardel-Vicente, A., Rodríguez-Navarro, D., & Bravo-Muñoz, I. (2018, September). Characterization of multipath effects in indoor positioning systems based on infrared signals. In *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (pp. 1-8). IEEE. <https://doi.org/10.1109/IPIN.2018.8533816>
- Kuxdorf-Alkirata, N., & Brückmann, D. (2016, September). Reliable and low-cost indoor localization based on bluetooth low energy. In *2016 3rd International Symposium on Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)* (pp. 92-96). IEEE. <https://doi.org/10.1109/IDAACS-SWS.2016.7805794>
- Li, Y., Ji, W., & Zhang, T. (2010, August). Ultrasonic positioning in three-dimensional space. In *2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics* (Vol. 2, pp. 18-21). IEEE. <https://doi.org/10.1109/IHMSC.2010.102>
- Liu, F., Wang, J., Zhang, J., & Han, H. (2019). An indoor localization method for pedestrians base on combined UWB/PDR/Floor map. *Sensors*, *19*(11), 2578. <https://doi.org/10.3390/s19112578>
- Poulose, A., & Han, D. S. (2020, February). Performance Analysis of Fingerprint Matching Algorithms for Indoor Localization. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)* (pp. 661-665). IEEE. <https://doi.org/10.1109/ICAIIIC48513.2020.9065220>
- Sun, W., Xue, M., Yu, H., Tang, H., & Lin, A. (2018). Augmentation of fingerprints for indoor WiFi localization based on Gaussian process regression. *IEEE Transactions on Vehicular Technology*, *67*(11), 10896-10905. <https://doi.org/10.1109/TVT.2018.2870160>
- Xu, Y., Shi, J., & Wu, X. (2007). A UWB-based localization scheme in wireless sensor networks. *2007 IET Conference on Wireless, Mobile and Sensor Networks (CCWMSN07)*, pp. 52-55, Shanghai, 2007.

- Zegeye, W. K., Amsalu, S. B., Astatke, Y., & Moazzami, F. (2016, October). WiFi RSS fingerprinting indoor localization for mobile devices. In *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 1-6). IEEE. <https://doi.org/10.1109/UEMCON.2016.7777834>
- Zhao, Y., Li, Z., Hao, B., Wan, P., & Wang, L. (2019). How to select the best sensors for TDOA and TDOA/AOA localization?. *China Communications*, *16*(2), 134-145.
- Zheng, Y., Sheng, M., Liu, J., & Li, J. (2018). Exploiting AoA estimation accuracy for indoor localization: A weighted AoA-based approach. *IEEE Wireless Communications Letters*, *8*(1), 65-68. <https://doi.org/10.1109/LWC.2018.2853745>
- Zhu, X., Zhu, W., & Chen, Z. (2018, May). Direct localization based on motion analysis of single-station using TOA. In *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)* (pp. 1823-1827). IEEE. <https://doi.org/10.1109/IMCEC.2018.8469278>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).