

# Indoor Localization Based on Optimized KNN

Xuanyu Zhu<sup>1</sup>

<sup>1</sup> Jinling Institute of Technology, Nanjing, China

Correspondence: Xuanyu Zhu, Jinling Institute of Technology, Nanjing, China

Received: November 25, 2020 Accepted: December 9, 2020 Online Published: December 31, 2020

doi:10.5539/nct.v5n2p34

URL: <https://doi.org/10.5539/nct.v5n2p34>

## Abstract

In recent years, with the continuous development of the economic situation, the price of low-end smart phones continues to reduce, the coverage of wireless local area network (WLAN) continues to improve, and individual users pay more and more attention to the real-time information around them, so indoor positioning technology has become a research hotspot. Among them, the indoor positioning based on the location fingerprint method quickly becomes the “Navigator” of indoor positioning direction by virtue of the simplicity of layout, the cost reduction of hardware facilities and the accuracy of positioning effect. However, the traditional indoor positioning methods usually rely on WiFi signal and KNN algorithm. When the KNN algorithm is implemented, there will be a lot of calculation and heavy workload to establish the location fingerprint database offline, and the efficiency and accuracy of online matching positioning points are low. This paper proposes an OKNN algorithm based on the improved KNN algorithm. By improving the efficiency of matching algorithm, the algorithm indirectly improves the positioning accuracy and optimizes the indoor positioning effect.

**Keywords:** localization, WiFi, RSSI, fingerprint

## 1. Introduction

Location based service refers to providing services for mobile users according to their geographic location or known location. At present, LBS can be widely used in many fields, such as mobile location, logistics, and so on. For indoor environment, outdoor positioning technology represented by GPS and cellular wireless positioning can not meet the needs of people’s daily life, and people increasingly need accurate and convenient indoor positioning services.

The indoor positioning system designed for WLAN is a research hotspot in recent years. These indoor positioning systems can be divided into two types: the system of establishing the fingerprint database based on fingerprint and the system of establishing the radio frequency propagation model according to the propagation characteristics. The most commonly used method is based on the signal strength RSS positioning method, which is simple to measure, low cost and easy to build, so it has become the mainstream measurement method.

The scheme of using location fingerprint to establish fingerprint database: due to the popularity of radio transmitter, such as WiFi access point and GSM base station, indoor positioning does not need to deploy additional equipment. The basic method of this kind of location scheme is to receive RSS (received signal strength) received by each receiver for the target space to be located (Strength) is used as a fingerprint vector to measure each location, and a fingerprint vector database of each location is established. Then, the mobile device is located by matching the RSS vector received by the mobile device with the database. Specifically, the positioning process is divided into two stages: offline training stage and real-time positioning stage. In the off-line training stage, several WiFi access points are fixed in the room, and then the mobile devices (such as mobile phones, laptops, etc.) are carried out in different indoor locations for RSS measurement. In the real-time positioning stage, the mobile terminal finds the coordinates closest to the K positions of the fingerprint database by matching the received RSS vector with the established fingerprint database data, and then obtains the user location coordinates by taking the geometric average.

## 2. Related Works

At present, indoor positioning methods can be roughly divided into two categories. The first kind is based on geometry location. The general principle of this method is to obtain the geometric relationship between the user terminal and the known location node through measurement, and obtain the user’s location through the algorithm. There are four commonly used measurement parameters. The first is time of arrival (TOA), which is used in

outdoor GPS positioning. The second is the time difference of arrival (TDOA). This parameter is an upgraded version based on signal delay, which is used for mobile phone base station positioning. The third is RSSI (received signal strength indication), which is mostly used for indoor positioning based on Bluetooth or WiFi. The last one is the angle of arrival (AOA). Since the positioning system using this parameter needs to use the antenna array, the general equipment cost is high. The second method is based on scene analysis. This technology needs preprocessing, abstracting and formalizing the environment, as well as specific quantitative parameters or indicators to describe the environment at specific feature points. Finally, the information is stored in the database, which is usually called “fingerprint database”. Most fingerprint databases take signal strength as the main parameter. In the process of location, the user measures the “fingerprint” of his position, compares it with the fingerprint database, and obtains the location information through a certain algorithm.

Many research institutes at home and abroad are located in the research room. After a lot of work, the following indoor positioning technologies are obtained

#### 1) RFID indoor positioning system

Radio frequency identification (RFID) is a kind of wireless communication technology, which uses the communication of wireless signals to realize automatic identification of targets. The non-contact two-way communication between tag and reader usually uses electrostatic coupling or inductive coupling to realize the data transfer between chip tag card and reader.

Radio signals transmit data from tags attached to items through electromagnetic fields tuned to radio frequencies to automatically identify and track items. The tag contains an electronic chip that can be identified within meters. RFID tags do not need to be placed in the field of vision of the identifier. They can be embedded in items, such as bags.

Most tags, which receive energy from electromagnetic fields, don't need batteries. There are also tags that contain an independent power supply that can actively emit radio waves. RFID positioning system can be divided into passive and active. As the user terminal also needs to send and receive signals, the active positioning leads to higher equipment cost, but the coverage also becomes larger because of the active receiving and sending signals of tags; the passive positioning tag is just a receiver, which can locate by receiving RF signals. At present, the spoton system of University of Washington and MIT's cricket system are typical RFID indoor positioning systems.

Although the operation of RFID is simple, the cost of transmission equipment is high and the transmission distance is limited. It is difficult to complete the complete positioning only by using RFID technology, which usually needs other positioning methods to assist.

#### 2) UWB indoor positioning system

Ultra wide band (UWB) technology is a short-range communication technology for high-speed wireless data transmission by sending and receiving ultra narrow pulse signals. The general process of UWB positioning is to use UWB transmission signal and use TDOA and AOA indoor positioning algorithm to solve the data of multiple sensors to obtain the terminal position.

UWB positioning technology has strong multipath resolution, anti electromagnetic interference ability, and high positioning accuracy (centimeter level). However, due to the short duration of pulse signal, large instantaneous power may affect the normal operation of many systems. Moreover, the technology has high requirements for equipment, the positioning cost is greatly increased, and the cost performance ratio is low.

#### 3) Infrared indoor positioning system

Infrared (IR) positioning is realized by receiving infrared ray. The first generation of indoor positioning system based on infrared is the active badge developed by at & T Laboratory of University of Cambridge. The positioning system is composed of two parts: terminal and base station. Each positioning target needs to be equipped with an infrared transmitting terminal. The modulated infrared signal containing its own identification ID will be sent periodically by the terminal. The system can judge the position of the terminal by the signal received by the base station of the known node. Then, the infrared radiation of the object is measured to locate the object, which is proposed by ambiplex. The positioning accuracy is as high as 20–30cm. The positioning accuracy of IR positioning technology is high, but infrared can not pass through obstacles, so the transmission distance is relatively short, and because the light travels along a straight line, it is disturbed by the larger indoor layout. The practical application is limited.

#### 4) Indoor positioning system based on Bluetooth

Bluetooth wireless communication is widely used, and almost all modern intelligent handheld devices support

Bluetooth function. The indoor positioning system based on Bluetooth uses RSSI to locate. Bluetooth location has the advantages of small footprint, simple deployment and low power consumption. The disadvantages are that the signal is easy to be interfered, the stability is poor and the positioning accuracy is low. The iBeacon device developed by Apple company is specially used for indoor positioning based on Bluetooth.

#### 5) Indoor positioning system based on WLAN

WLAN (wireless local area networks) has developed rapidly in the past decade, and more and more applications in the city make it possible to locate indoor based on WLAN. WLAN positioning system is mainly composed of the user's handheld terminal and the known wireless network.

At present, wireless fidelity (WiFi) technology is the most common technology in WLAN. WiFi, similar to Bluetooth, is also a short-range radio transmission technology that enables users to connect to the network without physical contact.

The indoor positioning technology based on WLAN uses WiFi signal strength to locate. Because WiFi is widely used in home or office, it can avoid deploying new equipment, reduce cost, and has good universality. A typical WLAN location system consists of three parts: user's handheld mobile terminal, location server and known wireless access point (AP). The whole indoor positioning process is divided into two parts: offline acquisition and online matching. In the off-line acquisition stage, the terminal is used to collect the RSSI features in the region at multiple reference points, and the collected data are matched with AP coordinates and stored in the location server database; in the online matching stage, the user holds the terminal terminal, uploads the real-time collected data to the positioning server, and the server matches the data with the positioning database, and obtains the terminal position through certain algorithm And send it to the terminal to complete the positioning.

#### 6) Other indoor positioning systems

In addition to the five indoor positioning systems described above, there are also indoor positioning systems based on optical communication, indoor positioning systems based on geomagnetism, and indoor positioning systems based on pseudo base stations. These systems have the disadvantages of difficult deployment, high equipment cost, small receiver signal range and poor system stability. Some of them are only implemented in the laboratory, so they are not available Extensive research.

### 3. Method

In this paper, an optimized weighted k-nearest neighbor method is proposed, which further improves the positioning accuracy in practical application and optimizes the positioning results in the experiment. In KNN algorithm, the position coordinates of the measurement points are obtained by taking the average value of the Euclidean distance of K reference points, and the difference between the measurement point and the positioning point is given weight, and the collected experimental data are further filtered to optimize the positioning results.

OKNN algorithm is a further improvement on the basis of KNN algorithm. When KNN algorithm calculates the average value of Euclidean distance of K fingerprint positions, the k data are weighted. The calculation results are as follows:

$$(x, y) = \sum_{m=1}^k \frac{d_m}{\sum_{n=1}^k d_n} (x_m, y_m)$$

$(x, y)$  is the coordinate of the measuring point,  $(x_m, y_m)$  is the coordinates of the  $m$ -th reference point,  $d_m$  is the Euclidean distance between the rssi at  $(x, y)$  and  $(x_m, y_m)$ .

When k is fixed, the calculated coordinates around the positioning point in the experiment will be more discrete. This may be because when the positioning points in all data are weighted and averaged, the Euclidean distance with large error between the measurement point and the positioning point is also included in the calculation. Therefore, in view of this situation, some points with large error will be filtered first, and then the next step of calculation will be carried out. DM is calculated according to the setting of reference points in the experimental environment. Through the horizontal unit length of 0.5 meters and the longitudinal unit length of 1.15 meters, through the conversion of the longest distance between the measurement point and the positioning point in the experiment, the threshold value of the Euclidean distance between the measurement point and the fixed point in this experiment is 2.31M. Then the data is removed and weighted to get the final experimental results, which improves the positioning accuracy of OKNN algorithm.

#### 4. Experiments

The experimental step is to take different values for  $K$ , and then calculate the error between the measurement points and the positioning points. According to the positioning results of the positioning points, the errors between the measurement points are constantly corrected. Finally, when the average error tends to be flat, this result is taken as the final positioning result of the measurement point.

$K$ -nearest neighbor algorithm is the experimental step: enter the selected area of the experiment, you can get  $k$  position fingerprint information, and then calculate the distance between the measured point and the reference point, take the minimum value, and take the average value between the  $K$  calculated coordinates, that is to get the position of the measurement point. In the experiment, different values of  $K$  are taken. Table 1 shows the average error between the measurement point and the positioning point when  $k$  takes different values. The following is the experimental analysis of  $k = 1, 2, 3, 4$  and  $5$ . Figure 1 shows the error between the measurement point and the positioning point during positioning.

Table 1. the average error using KNN with different values of  $K$

$K$	1	2	3	4	5
Average error	1.78m	1.97m	1.62m	2.02m	2.18m

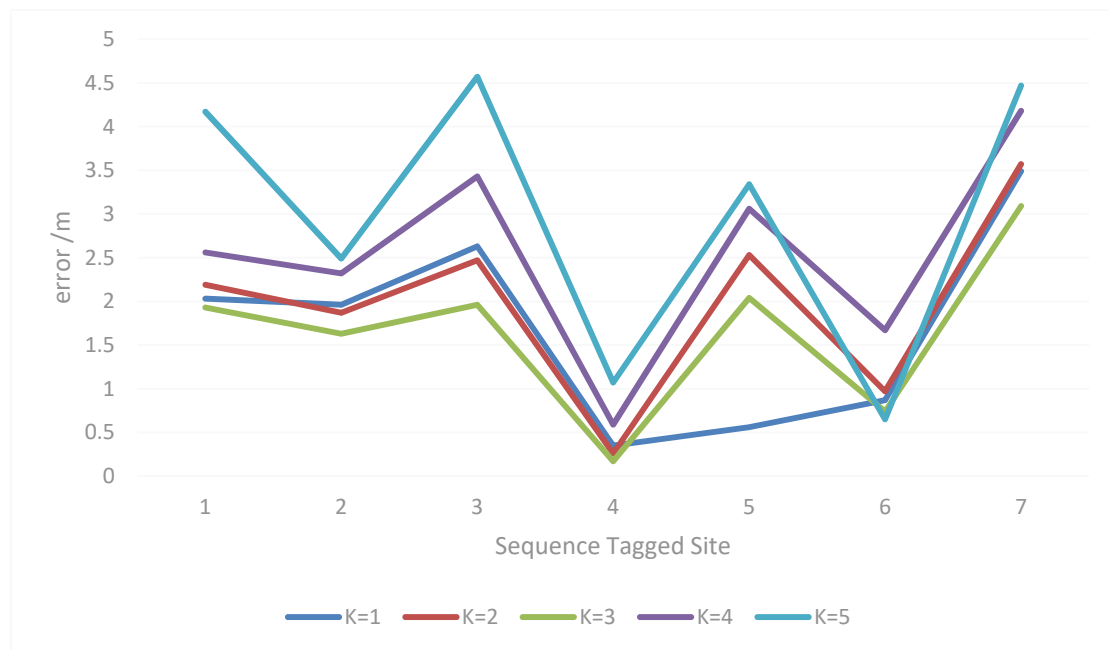


Figure 1. Average error using KNN with different  $K$

When KNN algorithm is selected for experimental positioning, it can be obtained from table 1 that when  $k = 3$ , the average error between measuring points and locating points is the minimum. From the figure, when  $k = 3$ , compared with other  $K$  values, most of the errors between measurement points and positioning points are smaller, and only some of them are larger than others. Through calculation, it can be concluded that when  $k = 3$ , the average error between measuring points and locating points is the smallest. When the value of  $K$  exceeds 3, the error curve begins to fluctuate, and the error between the measuring point and the measuring point begins to increase. When the value of  $K$  is 3, the error curve is relatively gentle compared with other values, and the positioning performance is slightly compared with other  $K$  values. Through the above comparison, it can be concluded that when the value of  $K$  is 3, the KNN algorithm has the best positioning performance, and the next is when the value of  $K$  is 2.

In the experiments using oknn algorithm,  $k = 1$  is removed from the experimental steps of KNN algorithm. Table 2 shows the average error between measurement point and positioning point when using oknn algorithm when  $k$  takes different values. Figure 2 shows the error between the measurement point and the positioning point during positioning.

Table 2. The average error using OKNN with different values of K

K	2	3	4	5
Average error	1.42m	1.18m	1.59m	1.61m



Figure 2. Average error using OKNN with different K

When OKNN algorithm is selected for experimental positioning, it can be obtained from table 2 that when  $k = 3$ , the average error between measurement points and positioning points is the minimum. From Figure 2, when  $k = 3$ , compared with other K values, most of the errors between measurement points and positioning points are relatively small, and only the data obtained by individual measurement points will be larger than other points. It can also be concluded from Fig. 2 that when  $k = 2$ , the error between measurement points and positioning points is small. Some of the measurement points and positioning points have small errors, but there is also a large error between measurement points and positioning points. Therefore, it can be concluded that when using OKNN, the effect of using  $k = 2$  to collect measurement points and positioning points is not as good as that of using  $k = 3$ . Moreover, when  $k$  is taken as 3, the fluctuation of error curve is relatively low. It can be concluded that when  $k = 3$ , the average error between the measuring point and the positioning point is about 1.23m, and the probability of the positioning error between the positioning point and the measuring point is about 71% when the error is less than 1.8m. Therefore, when using OKNN to locate between the positioning point and the measuring point, the efficiency of the positioning effect is about higher than that of the KNN algorithm It's up 28%. When the value of K exceeds 3, the error curve begins to fluctuate, and the error between the measurement point and the positioning point begins to increase. When the value of K is 3, the error curve is relatively gentle compared with other values, and the positioning performance is slightly improved compared with other K values. Through the above comparison, it can be concluded that when the value of K is 3, the positioning performance of OKNN algorithm is the best, followed by When  $k$  is 2.

## 5. Conclusions

With the dual development of wireless communication technology and intelligent mobile devices, location-based services are developing at an amazing speed, and show a broad market prospects. Lbs has gradually penetrated into all aspects of people's daily life and work environment, and brought great convenience for people. With the development of science and technology, the popularity of intelligent terminals, the market demand for indoor positioning is higher and higher. Therefore, a series of indoor positioning schemes based on consumer grade geomagnetism, consumer WiFi, RFID, industrial ZigBee, consumer Bluetooth, computer vision, ultrasonic and so on have appeared in the market. For outdoor positioning, GPS global positioning system has been able to meet most of the needs of life, and then due to multi-path effect and shadow effect and other factors, satellite signal can

not cope with the complex indoor environment, and then the common indoor positioning algorithm accuracy is too low. Therefore, this paper proposes an OKNN algorithm based on the improved KNN algorithm. By improving the efficiency of matching algorithm, the algorithm indirectly improves the positioning accuracy and optimizes the indoor positioning effect.

## References

- Ashraf, I., Hur, S., Shafiq, M., Kumari, S., & Park, Y. (2019). GUIDE: Smartphone sensors-based pedestrian indoor localization with heterogeneous devices. *International Journal of Communication Systems*, 32(15), e4062. <https://doi.org/10.1002/dac.4062>
- Ferreira, D., Souza, R., & Carvalho, C. (2020). QA-kNN: Indoor Localization Based on Quartile Analysis and the kNN Classifier for Wireless Networks. *Sensors*, 20(17), 4714. <https://doi.org/10.3390/s20174714>
- Gan, H., Khir, M. H. B. M., Djaswadi, G. W. B., & Ramli, N. (2018). A hybrid model based on constraint OSELM, adaptive weighted SRC and KNN for large-scale indoor localization. *IEEE Access*, 7, 6971-6989. <https://doi.org/10.1109/ACCESS.2018.2890111>
- Hu, J., Peng, H., Wang, J., & Yu, W. (2020). kNN-P: A kNN classifier optimized by P systems. *Theoretical Computer Science*. <https://doi.org/10.1016/j.tcs.2020.01.001>
- Kalpana, R., & Chandrasekar, P. (2020). An optimized technique for brain tumor classification and detection with radiation dosage calculation in MR image. *Microprocessors and Microsystems*, 72, 102903. <https://doi.org/10.1016/j.micpro.2019.102903>
- Lin, F., Wang, Z., Shen, D., Li, K., Zhao, H., Qiu, S., & Xu, F. (2019, December). Intelligent Flame Detection Based on Principal Component Analysis and Support Vector Machine. In *2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP)* (pp. 339-344). IEEE. <https://doi.org/10.1109/ICICIP47338.2019.9012179>
- Lingampeta, D., & Yalamanchili, B. (2020, February). Human Emotion Recognition using Acoustic Features with Optimized Feature Selection and Fusion Techniques. In *2020 International Conference on Inventive Computation Technologies (ICICT)* (pp. 221-225). IEEE. <https://doi.org/10.1109/ICICT48043.2020.9112452>
- Nie, F., Wang, Z., Wang, R., & Li, X. (2019). Submanifold-preserving discriminant analysis with an auto-optimized graph. *IEEE transactions on cybernetics*. <https://doi.org/10.1109/TCYB.2019.2910751>
- Roy, P., Chowdhury, C., Ghosh, D., & Bandyopadhyay, S. (2019). JUIndoorLoc: A Ubiquitous Framework for Smartphone-Based Indoor Localization Subject to Context and Device Heterogeneity. *Wireless Personal Communications*, 106(2), 739-762. <https://doi.org/10.1007/s11277-019-06188-2>
- Singh, I., Jindal, R., Pandey, K., Agrawal, K., & Kukreja, K. (2020). Revised Grey Wolf Optimized SVM-KNN Ensemble Based Automated Diagnosis of Breast Cancer. *Ingénierie des Systèmes d'Information*. <https://doi.org/10.18280/isi.250216>
- Sun, H., Liu, X., Deng, Q., Jiang, W., Luo, S., & Ha, Y. (2020). Efficient FPGA Implementation of K-Nearest-Neighbor Search Algorithm for 3D LIDAR Localization and Mapping in Smart Vehicles. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 67(9), 1644-1648. <https://doi.org/10.1109/TCSII.2020.3013758>
- Xie, W., Wang, J. S., Xing, C., Guo, S. S., Guo, M. W., & Zhu, L. F. (2020). Adaptive Hybrid Soft-Sensor Model of Grinding Process Based on Regularized Extreme Learning Machine and Least Squares Support Vector Machine Optimized by Golden Sine Harris Hawk Optimization Algorithm. *Complexity*, 2020. <https://doi.org/10.1155/2020/6457517>
- Yang, R., & Kan, J. (2020). Classification of Tree Species at the Leaf Level based on Hyperspectral Imaging Technology. *Journal of Applied Spectroscopy*, 1-10. <https://doi.org/10.1007/s10812-020-00981-9>
- Ying, G., Mingyang, F., Yuxi, S., et al. (2019). Bluetooth indoor localization algorithm based on clustering threshold and EWKNN. *Science of Surveying and Mapping*, 2019.

## 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/>).