Indoor Fingerprint Localization Algorithm Based on Variable Weight Using Channel State Information

Shasha Sun*, Peng Wang*b

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, Shandong, China

*1315580750@qq.com, b1322751326@qq.com

Abstract. In order to solve the problems of indoor fingerprint localization algorithm based on static weight, such as low positioning accuracy and poor environmental adaptability, a variable weight indoor based on channel state information with Euclidean distance as weight reference is proposed. In the preprocessing stage, the collected CSI amplitude values are first subjected to Butterworth low-pass filtering denoising processing, and then the values of each sampling point are averaged, and the reference point location fingerprint database is established by combining the known coordinates. The weight index $\alpha$ is introduced in the online positioning stage, and the nearest neighbor is found by the KNN algorithm with the CSI eigenvalue reference. Then the weight index $\beta$ is introduced, and the Euclidean distance is used as the weight reference, and the nearest neighbor is weighted. Get the coordinates of the target position. The experimental results show that the indoor fingerprint localization algorithm proposed by this scheme has higher positioning accuracy and less fluctuation of positioning error than the traditional KNN-based indoor fingerprint localization algorithm.

1. Introduction

At present, in an outdoor environment, GPS can achieve more accurate positioning, but in a complex indoor environment, the GPS signal is affected by the obstruction, and the positioning accuracy is not high. For complex indoor environments, various indoor positioning technologies such as infrared positioning [1-2], RFID(Radio Frequency Identification) [3], ultrasonic positioning [4], Bluetooth positioning [5], Ultra Wideband (UWB) positioning [6], and WLAN [7] positioning technologies have emerged. WLAN-based indoor localization has the advantages of low cost and low power consumption compared to other positioning technologies. The data forms obtained by WLAN-based indoor localization mainly include RSS and CSI. RSS is greatly affected by environmental changes and multipath effects, and the positioning accuracy is not high. CSI describes the small-scale multipath fading phenomenon, and describes the wireless channel state more finely. It also contains amplitude and phase information, which is more than single-valued. RSS contains richer frequency domain information. Therefore, CSI shows better stability and location sensitivity than RSS, providing new opportunities for more robust and flexible environment-aware applications.
2. Preliminary

2.1. Channel state information

In wireless communication systems, CSI not only reflects how the signal reaches the receiver, but also the effects of the signal propagation process, including scattering, refraction, and fading. The CSI takes into account the number of antennas and the number of subcarriers, including the channel state information of each subcarrier in the frequency domain, and can obtain a plurality of CSI values that change with time and remain relatively stable at one time.

The frequency domain model of channel state [8] is described as:

$$Y = HX + N$$

$Y$, $X$, $H$, $N$ represent the received signal vector, the transmitted signal vector, the channel matrix, and the Gaussian white noise, respectively.

The channel matrix represents a plurality of subcarrier information having a dimension of $N_t$ (number of antennas at the receiving end) $\times$ $N_r$ (number of antennas at the transmitting end) $\times$ $S$ (number of OFDM subcarriers), and the channel matrix is described as:

$$H_{ij} = |H_{ij}| e^{j\angle H_{ij}}$$

$H_{ij}$ represents the amplitude and phase matrix of each subcarrier of each antenna [9].

2.2. Indoor fingerprint location technology based on KNN algorithm

The fingerprint localization algorithm based on KNN model is widely used due to low computational complexity and low hardware cost. It extracts the feature values collected by the CSI on the mobile terminal and then matches all the fingerprint points in the fingerprint library, and finds the K closest fingerprint points. The coordinates of the test points are the centroids of the K fingerprint points, and the KNN algorithm can be effective. Solving the problem of occasionality leading to large positioning errors. The improvement of the weight based on the KNN fingerprint localization algorithm is mainly in two aspects. On the one hand, the weight is introduced by matching the Euclidean distance. In [10], an environmental adaptive fingerprint localization algorithm AWKNN is proposed. The offline process is used to solve the mean value of the acquired CSI amplitude. In the online phase, the AP weighting is used to calculate the Euclidean distance of all reference points of the point to be measured and the fingerprint database. The final target position is calculated by giving each neighbor point a different weight. The environmental adaptability and positioning accuracy of the algorithm are improved. On the other hand, when finding the nearest neighbor to solve the centroid, the weight is introduced as the weighted centroid, and the position of the centroid is taken as the position of the point to be measured. The literature [11], an optimized weighted centroid algorithm is used, which uses the reference node to wait. The reciprocal of the actual distance between the nodes is used as the weight. Due to the influence of the multipath effect caused by the occlusion of the indoor object on the wireless signal and the interference of the plurality of wireless devices on the positioning signal, the actual distance deviation determined based on the signal attenuation model is large. It has a certain influence on the positioning accuracy; the literature [12] combines the clustering algorithm with the KNN algorithm, and uses the clustering algorithm to extract the CSI, which improves the discrimination of fingerprints between different locations. Taking the similarity between the test sample and the training sample as the weight reference, the weight of the centroid algorithm is used to obtain the coordinates of the test point. The algorithm is basically consistent with the KNN accuracy in the WIFI positioning, but the positioning time is shortened accordingly.

Based on the above indoor fingerprint localization algorithm, an indoor fingerprint localization algorithm based on Euclidean distance is proposed to deal with complex indoor environment. The
algorithm first performs Butterworth low-pass filtering on the collected CSI original values to reduce the influence of noise on the positioning result, and then constructs the fingerprint database by taking the mean value. In the nearest neighbor algorithm matching stage, the weight index $\alpha$ is introduced, and the weight index $\beta$ is introduced in the target positioning stage. $\alpha$ and $\beta$ are adjusted according to the change of the environment scene, so that the positioning algorithm can adapt to the positioning environment, thereby further reducing the positioning error. By curve fitting and solving the extreme values of the function, the optimal values of the two weight indices are obtained. Therefore, the Euclidean distance weighting algorithm has better environmental adaptability, and the positioning accuracy and stability of the positioning result are greatly improved.

3. Variable weight fingerprint localization algorithm
Aiming at the problems of traditional indoor fingerprint localization algorithm based on static weights, such as low positioning accuracy, unstable positioning results and poor environmental adaptability, the Euclidean distance weighting algorithm is used to locate this paper, improve positioning accuracy and reduce positioning error. Figure 2 shows a flow chart of the algorithm, which achieves positioning through two stages of the offline sampling phase and the online positioning phase.

![Figure 1. Fingerprint positioning block diagram](image)

3.1. Data preprocessing
In the indoor environment, the original amplitude of the acquired CSI is affected by factors such as co-channel interference, multipath effect, and obstacle occlusion. Therefore, the original CSI amplitude value needs to be filtered. The data processing by Butterworth low-pass filtering model can effectively reduce the influence of some small probability and strong interference noise on the CSI values collected by the training nodes and test nodes, and achieve the purpose of improving the positioning accuracy. The CSI amplitude information is preprocessed for a period of time in order to construct the offline location fingerprint database and the online real-time positioning stage fingerprint matching service. In order to make the data information processing more reasonable and accurate, the positioning scheme adopts a sliding window of size M. The transmission rate is set to 12pkt/s.

The collected CSI raw data contains high frequency noise, which affects the positioning accuracy to some extent. Therefore, high frequency noise filtering is required for the frequency domain data information. In this paper, the Butterworth second-order low-pass filter is used. Among them, the N-order Butterworth low-pass filter is expressed as equation (3):

$$|H(w)|^2 = \frac{1}{1 + c^2(\omega / \omega_c)^2^\nu}$$

(3)
Where $\varepsilon$ is a constant less than 1, and $w$ is the lower corner frequency of the passband. In this scheme, a second-order Butterworth transfer function with $n$ is used to filter out high-frequency noise, $h$ is recorded as a transfer function, the original CSI data is $H$, and the filtered matrix is denoted as $H'$. The formula is as follows:

$$H' = h \cdot H$$  

(4)

3.2. Fingerprint database construction

N training points are selected in the indoor environment, and the physical coordinates $(x_i, y_i)$ of each training point are known, and the physical position information of the N training points constitutes a position space $L = (L_1, L_2, ..., L_N)$. For each training node, the CSI amplitude values from P APs are collected multiple times according to a certain time interval, and the collected data are averaged by Butterworth low-pass filtering and averaged, and these average values are used as training node $L_i$. The original fingerprint information is recorded as $F_i = (p_{k1}, p_{k2}, ..., p_{kP})$ the physical coordinates of the training node and the corresponding fingerprint constitute a fingerprint database.

3.3. Nearest neighbors acquisition

The CSI amplitude values are collected at the P APs constitute a P-dimensional fingerprint vector, and KNNs are matched with the fingerprints in the fingerprint database to find the nearest K points. This paper uses the improved weighted Euclidean distance algorithm to match and define for:

$$D(j) = \sqrt{\sum_{i=1}^{P} \omega_i \cdot (s_i - s_j)^2} \quad \left( i = 1, 2, 3, ..., n \right) \quad \left( j = 1, 2, 3, ..., n \right)$$  

(5)

$$\omega_i = \frac{1}{|s_i|^2}$$  

(6)

$s_i$ represents the CSI amplitude value of the training node, $s_j$ represents the CSI amplitude value of the test node, and $\alpha$ is the weight index. The change of $\alpha$ will cause the change of the average error, so there is a functional relationship between $\alpha$ and the average error. The function expression of $\alpha$ and the average error is determined by curve fitting. The $\alpha$ value of the average error taking the minimum value is the optimal value.

3.4. Target localization

After the WKNN algorithm is used to find the nearest K points, the E-series distance is sorted from small to large, and the first K training points with the smallest Euclidean distance are taken as the nearest neighbor. The traditional method of calculating the coordinates of the node to be tested is the nearest neighbor. The K points are obtained by geometric mean, and the coordinates of the test nodes are obtained. In this paper, the Euclidean distance is used as the weight reference. Considering the influence of environmental factors on the positioning, the weights in different environments should be adjusted, and the index $\beta$ is introduced. An improved weighted centroid algorithm with weights as follows:

$$\omega_j = \frac{(1/d_j)\beta}{\sum_{i=1}^{k} (1/d_i)\beta}$$  

(7)

d represents the Euclidean distance between the sample point to be tested and the $i$ training node with the smallest Euclidean distance. The improved weighted centroid algorithm obtains the positioning coordinates of the test point. The formula is as follows:
\[
\begin{align*}
X &= \omega_1x_1 + \omega_2x_2 + \cdots + \omega_xx_i \\
Y &= \omega_1y_1 + \omega_2y_2 + \cdots + \omega_yy_i
\end{align*}
\] (8)

Same as the way to calculate \( \alpha \), the relationship between \( \beta \) and the mean error is determined by curve fitting, and the extreme value of the function is solved to determine the optimal \( \alpha \) value.

The Euclidean distance weighting algorithm introduces the weight index \( \alpha \) based on the CSI eigenvalue and the weight index \( \beta \) based on the Euclidean distance in the nearest neighbor algorithm matching and the test point coordinate solution respectively, by selecting appropriate values in different environments. The positioning error is minimal and there is a certain environmental adaptability.

4. Experiment and analysis

4.1. Experimental environment

The selected experimental site was a laboratory with a length of 14 m and a width of 12 m at school. There are tables and computers and other debris in the laboratory. The experimental environment is complex and meets the requirements of the actual indoor positioning scene. Figure 2 shows the real scene of the experimental site. The total area of the laboratory is \( 14m \times 12m = 168m^2 \), and \( 7m \times 7m = 49m^2 \) is taken as the experimental area. Two PCs were placed on two diagonal tables in the lab.0.8m from the ground. Both computers are equipped with an Intel 5300 NIC. There are 3 antennas with the network card. The PC in the upper left corner sends a broadcast signal, and the PC in the lower right corner acts as a receiving device. In this experiment, CSI sample information was collected at 25 test nodes and 64 training nodes.

4.2. Experimental steps

\textbf{Step 1}, the experimental area is meshed, and the CSI amplitude value and the actual physical coordinates of each training point at the intersection of the grid are collected, and the CSI data value is subjected to Butterworth low-pass filtering processing to obtain an average value, and then a fingerprint database is established;

\textbf{Step 2}, randomly select 25 test points in the sampling area, and collect the CSI feature values and real coordinates of these test points;

\textbf{Step 3}, determine the weight index \( \alpha \) and apply the formula (5) to find the K training nodes of the nearest neighbor;

\textbf{Step 4}, determining the weight index \( \beta \), applying the formulas (7) and (8) to obtain a weighted average of the K neighbor points found in the state (3), and obtaining the test point coordinates;

\textbf{Step 5}, find the error between the coordinates of the test points and the real coordinates;

\textbf{Step 6}, analyze various performance indicators and compare performance with other algorithms;

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{experimental_environment.png}
\caption{Experimental environment}
\end{figure}
4.3. Determination of the weight index $\alpha$
Figure 3 is a fitting curve of the weight index $\alpha$ and the average error. The fitting curve is used to determine the functional relationship $F(\alpha)$ between $\alpha$ and the mean error, and the minimum value $\min F(\alpha)$ of the fitted curve is obtained to obtain the optimal value of $\alpha$. After calculation, when the experimental environment $\alpha$ takes a value of 4.0, the average error is the smallest, that is, the optimal value is $\alpha = 4.0$.

![Figure 3. Fit curve of $\alpha$ and mean error](image)

4.4. Determination of weight index $\beta$
After $\alpha = 4.0$ has been determined, determine the value of $\beta$. Figure 4 is the fitting curve of the weight index $\beta$ and the average error. As with the calculation of $\alpha$, the function relationship $g(\beta)$ between $\beta$ and the average error is first determined by fitting the curve, and then the minimum value $\min g(\beta)$ of the function is obtained to obtain the optimal $\beta$ Value, after calculation, the optimal value in this experiment is $\beta = 3.8$.

![Figure 4. Fit curve of $\beta$ and mean error](image)

The KNN, B-KNN and Euclidean distance weighting algorithms are run on the same computer, and the above three algorithms are evaluated from four indicators: average error, maximum error, minimum error and error variance. The results are shown in Table 1.
Table 1. Comparison of performance of different algorithms

| Algorithm name     | Average error (m) | Maximum error (m) | Minimum error (m) | Error variance (m) |
|--------------------|-------------------|-------------------|-------------------|-------------------|
| KNN                | 1.38              | 3.54              | 0.25              | 1.13              |
| B-KNN              | 1.16              | 2.49              | 0.24              | 0.89              |
| Algorithm of the article | 0.97              | 2.92              | 0.01              | 0.88              |

It is concluded from Table 1 that the average error of the KNN algorithm is 0.22m larger than that of the BKNN algorithm, and the error variance is 0.24, which indicates that the noise in the indoor environment has a great influence on the data. After the low-pass filtering process by Butterworth, the stability of the localization result and the positioning accuracy has been significantly improved. In addition, this paper compares the performance of Euclidean distance weighting algorithm and other three algorithms. The average error of the Euclidean distance weighting algorithm is 0.41m and 0.19m lower than KNN and BKNN respectively. The positioning accuracy is obviously improved, and the maximum error and minimum error are improved. The difference between the gap and the error variance is smaller than other algorithms, indicating that the positioning performance is relatively stable.

5. Conclusion

In this paper, the KNN fingerprint localization algorithm is improved in consideration of more adapting to the environment and improving positioning accuracy. By introducing weight index $\alpha$ and $\beta$, a variable weight indoor fingerprint localization algorithm is proposed. The collected CSI amplitude values are pre-processed by Butterworth low-pass filtering, and the average value is used to construct a fingerprint database and then weighted KNN algorithm matching. Then, the improved weighted centroid algorithm is used to obtain the test point coordinates. The experiment proves that the Euclidean distance weighting algorithm is superior to the KNN algorithm in localization accuracy and localization stability, which can meet the high precision localization requirements. At the same time, the weight index $\alpha$ and $\beta$ are proposed to make the algorithm have certain environmental adaptability and stronger localization accuracy. In future, the clustering algorithm is combined with the algorithm of this paper to reduce the computational complexity and overcome the shortcomings of fingerprint localization are considered.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61801270, in part by the Shandong Province Higher Educational Science and Technology Program under Grant J17KA075.

References

[1] Stojanović D, Stojanović N. Indoor Localization and Tracking: Methods, Technologies and Research Challenges [J]. Facta Universitatis, Series:Automatic Control and Robotics, 2014, pp: 57-72.
[2] Zhu L, Yang A, Wu D. Survey of Indoor Positioning Technologies and Systems [M]. Life System Modeling and Simulation. Springer Berlin Heidelberg, 2014, pp: 400-409.
[3] Kumar S, Vasisht D, Katabi D. Decimeter-Level Localization With a Single WiFi Access Point [C]. Usenix Conference on Network Systems Design and Implementation, USEnix Association, Santa Clara, CA, USA, 2016.
[4] Xiong J, Sundaresan K, Jamieson K. ToneTrack: Leveraging Frequency-Agile Radios for Time-Based Indoor Wireless Localization [C]. International Conference on Mobile Computing and Networking, ACM, Paris, France, 2015.
[5] Gjengset J, Xiong J, McPhillips G. Phaser: Enabling Phased Array Signal Processing on Commodity WiFi Access Points [c]. International Conference on Mobile Computing and Networking. ACM, Havwai, USA, 2014.
[6] Kotaru M, Joshi K, Bharadia D. SpotFi: Decimeter Level Localization Using WiFi [J]. ACM
Sigcomm Compute Communication Review, 2015, pp: 269-282.

[7] Shi Ke, Chen Hongsheng, Zhang Rentong. An 802.11 Wireless Indoor Positioning Method Based on Support Vector Regression [J]. Journal of Software, 2014, pp: 2636-2651.

[8] Liu Yibo, Xiu Chunzhen, Zhu Yunlong. Indoor Positioning Method Using Channel State Information [J]. Computer Engineering and Applications 2016, pp: 238-241.

[9] Mohammed, Abdulaziz, Al-qaness. Human behavior recognition system based on channel state information [J]. Journal of Wuhan University of Technology, 2016, pp: 76-80.

[10] Tian Lizhen. Research on indoor fingerprint localization algorithm based on CSI [D], 2017.

[11] PENG Wei, ZHAO Yang, XIA Tian-peng. WSN Weighted Centroid Localization Algorithm Based on Optimized RSSI Accuracy [J]. Computer Engineering and Applications, 2015, pp: 88-91.

[12] Tian Guangdong, Yang Pinzhang, Wang Shan. CSI indoor positioning based on Kmeans clustering [J]. Application of Electronic Technique, 2016, pp: 62-64.