# Indoor Fingerprinting Localization Based on Fine-grained CSI using Principal Component Analysis

1st Jingjing Wang*, 2th Xianqing Wang†,
3nd Jishen Peng‡, 4rd Jun Gyu Hwang§, 5th Joon Goo Park¶
*Electronic and Electrical Engineering, Kyungpook National University, Daegu,Republic of Korea*
Email:wjj0219@naver.com*, haongiaong@naver.com†
pjs951121@naver.com‡, cjstk891015@naver.com§, jgpark@knu.ac.kr¶

*Abstract*—With the development of Wi-Fi technology, the IEEE 802.11n series communication protocol and the subsequent wireless LAN protocols use multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) technologies. Channel state information (CSI) fingerprint positioning technology based on fine-grained channel state information is widely used in the field of WIFI indoor positioning. However, the propagation of CSI is still affected by indoor multipath, and we cannot obtain signals in some corner areas. Therefore, CSI needs a suitable calibration method to improve the accuracy of the position estimation system. This paper proposes a fine-grained CSI fingerprint location algorithm based on Principal Component Analysis (PCA). This novel algorithm uses a dimensionality reduction method on the basis of the Discrete Wavelet Transform (DWT) to optimize, eliminate the noise and redundancy of the original data and reduce the positioning error. Experimental results show that the proposed approach achieves significant localization accuracy improvement over using the RSSI fingerprint method and original CSI fingerprint method, while it incurs much less computational complexity. Meanwhile, the algorithm improves the influence of multiple paths in a complex indoor environment on location, and the method can obtain more accurate location results.

*Index Terms*—Fine-grained CSI; Fingerprint positioning algorithm; Principal component analysis

## I. INTRODUCTION

At present, location-based services (LBS) [1] have shown a clear upward trend in people's needs. In indoor environments, we often need to locate objects accurately. Traditional satellite positioning technology uses microwave transmission. It is difficult for signals to pass through the complex outer layers of buildings and cannot meet the requirements of indoor positioning. The current indoor positioning technology is divided into GNSS pseudolite technology, wireless positioning technology, and other positioning technology. Wireless positioning technology is the most widely used, specifically Bluetooth technology, Wi-Fi technology, ultrasonic technology, radio frequency technology, etc.

Wi-Fi positioning technology adopts the 2.4 GHz public frequency band and the widely used IEEE 802.11b standard among wireless positioning technologies. It has the advantages of low price, easy deployment, and wide range and has become the easiest indoor positioning method to be promoted on a large scale. Due to the complexity of the indoor environment, the signal strength indicators (RSSI) [2] will be affected by factors such as wall reflections and interference from obstacles, which will cause the RSSI value received by the node under test to fluctuate impact greatly. It is necessary to propose a stable measurement index that changes with time to improve indoor positioning accuracy. Moreover, the index needs to meet the condition of being less affected by multipath effects.

RSSI is the superposition of signals from multiple paths and is very unstable. The high probability of RSSI overlap in different locations makes it difficult for RSSI to complete high-precision indoor positioning, which is only suitable for rough estimation of the positioning range. CSI is not the superposition of all sub-carrier information, it describes the signal of multiple paths, has more characteristics, and contains the channel state information of multiple sub-carriers. The amplitude and phase of the CSI sub-carriers at different positions are crossed, and a single sub-carrier cannot be distinguished. If the diversity of CSI frequencies is used and all sub-carrier data are used, different positions can be distinguished through algorithms. Generally speaking, CSI based on OFDM system [3], [4] has richer characteristic information than RSSI, and it has the following advantages [5]: (1) It has richer channel characteristics; (2) It is more stable in a static environment and is more sensitive to dynamic environments. The CSI characteristics of the location are different; (3) The anti-interference ability is stronger, and the robustness is better;

Paper [6] investigated the performance of fingerprint positioning in terms of the density of access points (AP) and other aspects based on the actual deployment methods and measurement results. FIFS [7] uses the weighted average CSI amplitude values on multiple antennas to improve the performance of

indoor fingerprint recognition methods. Due to the influence of multiple paths, the weighted average CSI amplitude value obtained does not reflect the location fingerprint information well. DeepFi [8] learns a large amount of CSI data from three antennas for indoor positioning based on deep networks. However, a large amount of CSI data is needed during model training. The CSI-MIMO [9] technology makes full use of the diversity characteristics of space and frequency string of the Multiple-Input Multiple-Output (MIMO) system and further uses the amplitude and phase deviation between the sub-carriers in the CSI. Establish a fingerprint database so that there is a strong correlation between the feature value and the position of the fingerprint point.

The multipath effect in the signal propagation process in the indoor environment will lead to the low positioning accuracy of the existing positioning technology. This paper adopts the more stable and better resolution channel state information CSI in the physical layer to achieve indoor positioning. The obtained CSI undergoes Discrete Wavelet Transform (DWT) [10], [11] to filter out abnormal values. A Filtered CSI fingerprint positioning algorithm based on Principal Component Analysis (PCA) [12] is proposed, which improves the calculation speed and accuracy of positioning.

The rest of this paper is organized as follows. Section II is about the introduction of related work. In section III, we introduce the proposed fingerprint Localization algorithm. The implementation of the enhanced algorithm and experimental evaluations are presented in Section IV. Finally, conclusions and future research are presented in Section V.

## II. RELATED WORK

Compared with the single-value RSSI formed by the fusion of multiple paths, CSI provides more fine-grained channel response information [13]. CSI is transmitted at the same time after being modulated into multiple sub-carriers of different frequencies through orthogonal frequency division multiplexing (OFDM). Each sub-carrier is independent of each other and does not affect each other. It describes the process from signal transmission to reception, and can characterize the multipath propagation phenomenon of the signal. In the process of communication between the transmitting end and the receiving end, the system model after the received signal Y is converted into the frequency domain can be expressed as formula (1):

$$Y = HX + n \tag{1}$$

Among them, $X$ is the set of transmitted signals, H is the channel transfer matrix, and $n$ is the additional channel noise matrix, which is usually regarded as Gaussian white noise, then the channel transfer matrix is estimated as:

$$\hat{H} = \frac{Y}{X} \tag{2}$$

The estimated channel transfer matrix obtained at the receiving end by formula (3) is the channel state information of all sub-carriers. $H$ actually represents data in the frequency domain, including the amplitude and phase of each subcarrier. According to the 802.11a/g/n protocol, the number of sub-carriers is related to the bandwidth. Therefore, H can be expressed as:

$$H = [H_1, H_2, H_3, ..., H_k, ..., H_N] \tag{3}$$

where $H_k$ describes the CSI of the k-th subcarrier. CSI expresses as a matrix of $pxq$ dimension. $H(f_x)$ stands for channel matrix:

$$H(f_x) = \begin{bmatrix} h_{11} & h_{22} & \cdots & h_{1q} \\ h_{21} & h_{22} & \cdots & h_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ h_{p1} & h_{p2} & \cdots & h_{pq} \end{bmatrix} \tag{4}$$

where $H_{pq}$ is a complex number, which contains the amplitude and phase of the subcarrier on each antenna. $H$ appears in the complex form $a + bi$. We can get the modulus $\sqrt{(a^2 + b^2)}$ and argument $\theta = argtan\frac{b}{a}$ of the complex number, that is, the corresponding amplitude and phase.

By modifying the firmware, the common Wi-Fi device can obtain CFR samples on 30 OFDM sub-carriers, so $N = 30$. Each group of CSI represents the amplitude and phase of an OFDM subcarrier:

$$H_k = \|H_k\| e^{j\angle H_k} \tag{5}$$

where $\|H_k\|$ and $\angle H_k$ represent the amplitude and phase of k-th subcarriers respectively. Intel 5300 wireless network card works in the high throughput mode (HT mode) of 20MHz. Table I shows the available information about CSI.

TABLE I
CHANNEL STATE INFORMATION.

| Data Information | Properties |
|---|---|
| Timestamp-low | The lower 32 bits 1MHz clock. |
| Bfee-count | Number of count beamforming |
| Nrx | Number of Nrx receiving antennas |
| Ntx | Number of Ntx transmit antennas |
| rssi-a, rssi-b, rssi-c | RSS of each receiving antenna |
| rate | Rate the transmission rate |
| noise | Noise |
| CSI | 3-dimensional array of Nrx * Ntx * 30 |

## III. PROPOSED ALGORITHM

Most of the existing indoor fingerprint positioning algorithms lack the optimization processing of the original data, which inhibits the running effect of the matching algorithm to a certain extent. Therefore, an indoor fingerprint location algorithm based on PCA principal component analysis.

This section introduces the structure diagram of the fingerprint algorithm, the preprocessing of the original data and the PCA reduction processing of the CSI amplitude after the preprocessing.

323

## A. Structure diagram of proposed algorithm

In the offline phase, the CSI values from different APs are obtained at different reference points, and the obtained CSI values are processed by discrete wavelet. After that, the filtered CSI is subjected to dimensionality reduction processing. The acquired low-dimensional CSI is stored as fingerprint features. In the online phase, the measured value data of any position is obtained in the same way, combined with the WKNN method to match the fingerprint database established in the offline phase. The structure diagram of the indoor fingerprint algorithm is shown in Fig. 1. where $CSI_{i1}, ..., CSI_{im}$ is used to represent
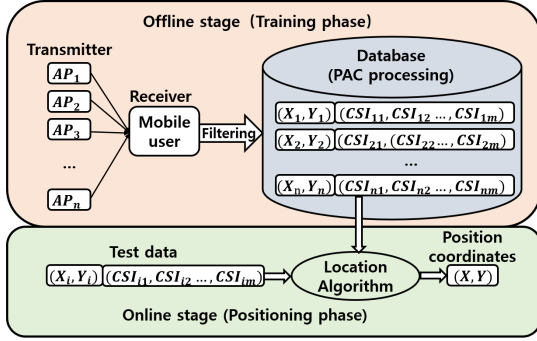


Fig. 1. Structure diagram of fingerprint algorithm.

the dimensionality reduction processing CSI value of the n-th AP collected at an unknown location. Use $CSI_{n1}, ..., CSI_{nm}$ to represent the dimensionality reduction CSI value of the n-th AP recorded at a reference point in the fingerprint database.

## B. Discrete Wavelet Transformation of CSI Data

In this paper, the CSI signal obtained under the 802.11n protocol through the CSI tool is the OFMD orthogonal frequency division multiplexing signal. It is the amplitude and phase information of every other sampled subcarrier among 56 subcarriers. There are 30 sampling subcarriers. The center frequency of these sub-carriers is 5 GHz. The guard frequency interval between the carriers is about 150 kHz. Therefore, a large amount of CSI information can be obtained in a short time. Fig. 2 shows CSI values on different antennas obtained at the reference point. Fig. 4 is the amplitude value of the CSI subcarrier obtained on antenna a.

The wavelet transform denoising can retain the peak value and sudden change part of the valuable signal needed in the original signal. When filtering with Fourier analysis, since the valuable signal is concentrated in the low-frequency part, and the noise is concentrated in the high-frequency part, the high-frequency part of the required signal and the high-frequency interference caused by noise cannot be effectively carried out through low-pass filtering. If the low-pass filter is too narrow, part of the desired
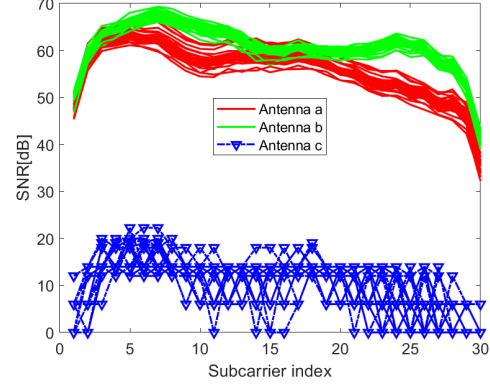


Fig. 2. CSI values on different antennas obtained at the reference point.
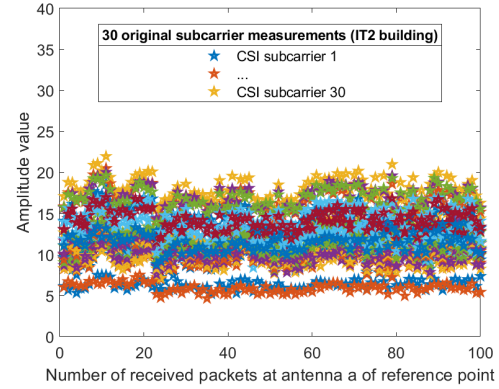


Fig. 3. The amplitude value of the CSI subcarrier obtained on antenna a.

signal is filtered as noise, and its morphological information is erased, resulting in distortion of the original signal. Fourier analysis has poor robustness to signal denoising, and wavelet transform can solve the above shortcomings well. Wavelet transform has good time-frequency localization characteristics, usually linearly expressed as:

$$W_k = W_f + W_c \qquad (6)$$

where $W_k$ is a discrete input signal with a length of N. $W_f$ is a low pass filter, which can filter out the high frequency part of the input signal and output the low frequency part. $W_c$ is a high pass filter, which is the opposite of a low pass filter, which filters out the low frequency part and outputs the high frequency part.

The process of performing a discrete wavelet transform on the acquired raw data can be divided into four stages:

- Step 1: Select a wavelet basis function and sample the signal at equal intervals to obtain the sample point sequence $c(k)$ corresponding to the signal;

- Step 2: Perform N-level discrete wavelet transform (DWT) based on the sample point sequence to obtain N-level wavelet expansion coefficients d(k) of different scales;
- Step 3: Select the corresponding thresholds and threshold rules for the wavelet expansion coefficients at all levels to perform thresholding to obtain the wavelet expansion coefficients at all levels after the threshold processing;
- Step 4: Perform N-level reconstruction according to the wavelet expansion coefficients after threshold processing and the unprocessed wavelet expansion coefficients; obtain the denoising signal CSI.

## C. PCA-based low-dimensional CSI fingerprint features

PCA principal component extraction is a standard algorithm that integrates and compresses the variance in the data. On the premise of ensuring that a large amount of information is not lost, the amount of data is significantly reduced. This data analysis method is a data extraction method. Since the purpose of this article is to achieve the matching of the real-time fingerprint database in the online phase, a large amount of CSI information obtained in a short time should be compressed. The online stage is to perform a fingerprint similarity matching, so we keep the environmental information contained in the CSI as high as possible. The weight of the environmental information reflected by different carriers in different environments is different and changes dynamically. Therefore, we cannot just select specific carrier signals for judgment. At the same time, it is not possible to superimpose all the carrier signals together and then average because the fading of some carriers will weaken other carriers' severe environmental information changes at that moment. Therefore, we need to extract the data that contains much information. The measure of this amount of information in mathematics is variance.

The purpose of data dimensionality reduction in indoor positioning is to be able to represent all the information of the data with new features and eliminate the redundancy caused by multipath loss in a complex environment. On the one hand, it can improve the positioning accuracy, and on the other hand, it can reduce the workload of offline database construction, thereby reducing the amount of online matching operations. Based on the above analysis, PCA can extract the main components of the CSI data, reduce data redundancy, thereby achieving dimensionality reduction and reducing the amount of data.

Principles and steps of PCA principal component extraction is shown in Step 1 to Step 6: The input value is data set CSI amplitude value, which needs to be reduced to k dimensions.

Step 1: Form the original data CSI (with $m$ pieces of n-dimensional data) into a matrix $X$ of $n$ rows and $m$ columns.

$$X = \begin{pmatrix} CSI_{11} & CSI_{12} & \ldots & CSI_{1m} \\ CSI_{21} & CSI_{22} & \ldots & CSI_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ CSI_{n1} & CSI_{n2} & \ldots & CSI_{nm} \end{pmatrix} \quad (7)$$

Step 2: Each feature subtracts its average (ie decentralization). After processing the data in a standardized form, the mean of the sample data is 0 and the variance is 1.

$$X_i' = \frac{X_i - E(X_i)}{\sigma_i} \quad (8)$$

where $X_i'$ represents the new variable after standardization, $X_i$ represents the original variable, $E(X_i)$ represents the mean, and $\sigma_i$ represents the variance. Finally, the standardized matrix is:

$$X - \overline{X} = \begin{pmatrix} y_{11} & y_{12} & \ldots & y_{1p} \\ y_{21} & y_{22} & \ldots & y_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ y_{n1} & y_{n2} & \ldots & y_{np} \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix} \quad (9)$$

Step 3: Calculate the covariance matrix of the sample features (that is, the correlation coefficient matrix. Here, dividing or not dividing the number of samples n or n-1, in fact, has no effect on the calculated eigenvectors.

$$Corr_X = E\left\{(\boldsymbol{X} - \overline{\boldsymbol{X}})(\boldsymbol{X} - \overline{\boldsymbol{X}})^{\mathrm{T}}\right\} = Y * Y^T \quad (10)$$

Step 4: Use the eigenvalue decomposition method to find the eigenvalues and eigenvectors of the covariance matrix. The eigenvalue of the correlation coefficient matrix $Corr_X$ and the corresponding eigenvector $U_i$ are shown below.

$$[Corr_X - \lambda_i] U_i = 0 \quad (11)$$

where $U_i$ represents the unit eigenvector corresponding to the i-th eigenvalue $\lambda_i$.

Step 5: Sort the eigenvalues from large to small, and select the largest $k$ among them. Then the corresponding $k$ eigenvectors are used as row vectors to form the eigenvector matrix. The new feature is a linear combination of the original features. That is, the principal component $S_1, S_2, ..., S_k$ is determined in turn according to the cumulative contribution rate. Combine the eigenvector $U_i$ and the standardized data matrix $X_i'$ to obtain the i-th principal component $S_i$, as shown below.

$$S_i = U_i \times X_i' \quad (12)$$

Step 6: Convert the data to a new space constructed by $k$ feature vectors, which is the data after dimension reduction to $k$ dimensions. Use principal component analysis to extract the first $k$ principal components $S_1, S_2, ..., S_k$ to make a linear combination. Because $k < i$, the compression of the data dimension of $X$ is completed.

## IV. EXPERIMENTS AND PERFORMANCE EVALUATION

### A. Experimental Scenarios

The experiments were carried out in the 1st-floor lobby of Kyungpook National University (KNU) IT2 Building, as shown in Fig. 4. The equipment required
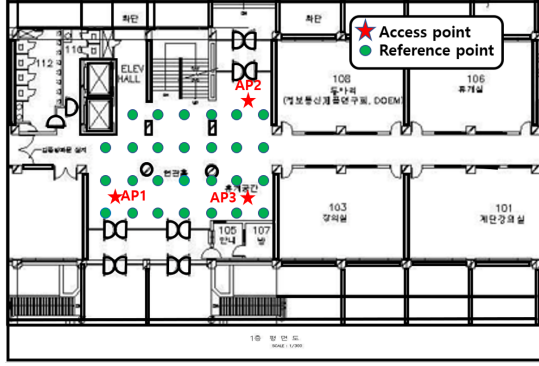


Fig. 4. Representative indoor corridor environment (1st-floor lobby of KNU IT2 Building).

for the ranging solution is: (1) A laptop with an Intel 5300 installed, the operating system is Ubuntu 10.04 LTS, and the kernel and wireless network card drivers are customized; (2) Three 802.11n wireless AP device. In experiments, we use ipTIME N3004 as a node and place it at the height of 0.2m on the ground. An Intel 5300 wireless network card built-in notebook computer DELL Inspiron N4010 is used as a mobile node. We measure sequentially at the reference point with a grid size of 1m*1m. At each reference point, 1600 measurements were performed on three different APs.

### B. Performance Evaluation

The experimentally measured data is processed on Matlab software. In the offline process, a fingerprint database based on the CSI amplitude is established. This paper collects the raw CSI data obtained on 3 APs at various reference points. The acquired raw data is processed by the proposed algorithm and stored in the fingerprint database based on the CSI amplitude. In the process of collecting the CSI value of each reference point, self-programming software is used to obtain the CSI amplitude value.

First, we preprocess the acquired CSI subcarriers. We selected the CSI data measured on antennas a and b for experimental verification. Based on the content of the previous chapter, Fig. 5 shows the CSI subcarriers before and after offline wavelet transform. in Fig. 5.

Use the PCA dimensionality reduction step described in the previous section to perform dimensionality reduction processing on the CSI data after discrete wavelet transform. Each CSI data packet contains 30 subcarriers, that is, 30 dimensions. The online
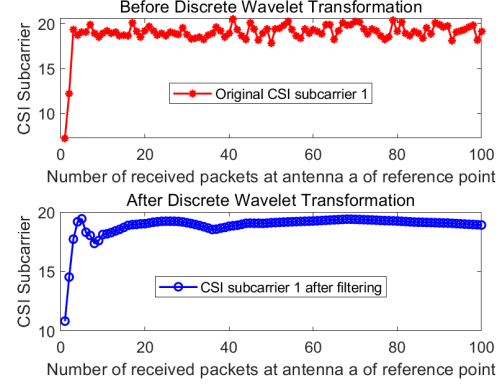


Fig. 5. CSI subcarriers before and after discrete wavelet transform processing.

phase is to perform a fingerprint similarity matching, so the environmental information contained in the CSI is retained as high as possible. We want to extract the components that contain a lot of "information" in the data. The measure of this amount of information in mathematics is variance.
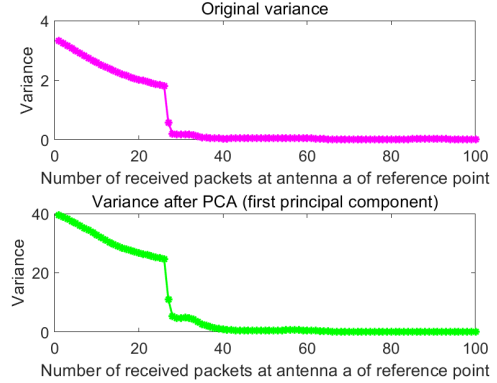


Fig. 6. Variance of CSI data before and after PCA processing.

Fig. 6 uses the data collected at a reference point to perform PCA dimensionality reduction processing. Fig. 6 compares the variance before and after PCA dimensionality reduction. It can be seen from Fig. 6 that the variance before and after dimensionality reduction is not very large, so the CSI after dimensionality reduction better retains the amount of fingerprint data information. Fig. 7 shows the comparison of CSI before and after PCA dimensionality reduction processing.

In the online matching stage, the Weighted K-Nearest Neighbors (WKNN) method is combined to verify the positioning error. The positioning error err and the average positioning error avgerr are respectively:

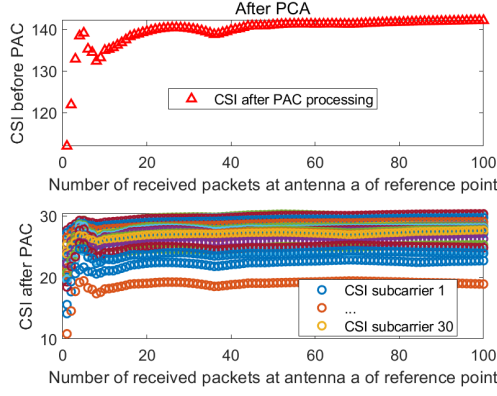$$\text{Error}_i = \sqrt{(x_i - x_i')^2 + (y_i - y')^2} \qquad (13)$$

326

Fig. 7. CSI data before and after PCA dimensionality reduction processing.

$$Average = \frac{1}{n} \sum_{i=1}^{n} Error_i \qquad (14)$$

Under the same experimental conditions, this paper also implements RSSI fingerprint positioning and CSI-based fingerprint positioning algorithms. These existing algorithms are used to compare and verify the advantages of the proposed algorithm in constructing the mapping relationship between positioning features and position coordinates. The positioning results of several algorithms are shown in Table 1.

TABLE II
COMPARISON OF THE POSITIONING ERROR OF THE EXISTING
METHOD AND THE PROPOSED METHOD.

| Fingerprint position-ing methods | Positioning error | Standard deviation |
|---|---|---|
| RSSI-based method | 2.078m | 1.007m |
| CSI-based method | 1.767m | 0.781m |
| The proposed method | 1.372m | 0.652m |

It can be seen from Table 1 that the average positioning accuracy of the PCA-based CSI fingerprint positioning algorithm is higher than the average positioning accuracy based on the RSSI and CSI algorithms. This reflects the advantages of CSI data after PCA dimension reduction in constructing positioning features and position relationships.

## V. CONCLUSIONS AND FUTURE RESEARCH

This paper analyzes the distribution of CSI amplitude, uses discrete wavelet transform to remove the abnormal data, improves the reliability of the data, and reduces the amount of data. At the same time, this paper introduces a clustering algorithm based on PCA principal component analysis based on the CSI fingerprint algorithm. PCA is used to reduce the dimensionality of CSI data and eliminate correlation redundancy, thereby improving the reliability of matching positioning. Experimental results show that the algorithm improves the positioning stability

and positioning accuracy of the original CSI fingerprint library without changing the hardware equipment of the indoor environment and can meet the real-time requirements of indoor positioning. However, the proposed algorithm does not verify the multi-AP and more CSI sub-carrier scenarios, and the matching stage does not use a more refined fingerprint matching algorithm, which will be carried out in further research.

## VI. ACKNOWLEDGMENT

## REFERENCES

[1] Simon Yiu, Marzieh Dashti, Holger Claussen, and Fernando Perez-Cruz, "Wireless RSSI fingerprinting localization," Signal Processing, 2017.

[2] J. J. Wang, J. G. Hwang, and J. G. Park, "A novel indoor ranging method using weighted altofrequent RSSI measurements," in 2017, 11th International Conference on Signal Processing and Communication Systems, ICSPCS 2017 - Proceedings, 2018.

[3] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," IEEE Trans. Signal Process., 2005.

[4] B. Liu, Z. Yan, and C. W. Chen, "Medium Access Control for Wireless Body Area Networks with QoS Provisioning and Energy Efficient Design," IEEE Trans. Mob. Comput., 2017.

[5] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," ACM Comput. Surv., 2013.

[6] Simon Yiu, Marzieh Dashti, Holger Claussen, and Fernando Perez-Cruz, "Wireless RSSI fingerprinting localization," Signal Processing, 2017.

[7] Xiao, J.; Wu, K.; Yi, Y.; Ni, L.M. FIFS: Fine-grained indoor fingerprinting system. In Proceedings of the 2012 21st International Conference on Computer Communications and Networks, ICCCN 2012 - Proceedings; 2012.

[8] Wang, X.; Gao, L.; Mao, S.; Pandey, S. DeepFi: Deep learning for indoor fingerprinting using channel state information. In Proceedings of the 2015 IEEE Wireless Communications and Networking Conference, WCNC 2015; 2015.

[9] Abdelrahman, R.B.M.; Mustafa, A.B.A.; Osman, A.A. A Comparison between IEEE 802.11a, b, g, n and ac Standards. IOSR J. Comput. Eng. 2015, doi:10.9790/0661-17532629.

[10] H. Kanagaraj and V. Muneeswaran, "Image compression using HAAR discrete wavelet transform," in ICDCS 2020 - 2020 5th International Conference on Devices, Circuits and Systems, 2020.

[11] S. Fang, W. Chang, Y. Tsao, H. Shih and C. Wang, "Channel State Reconstruction Using Multilevel Discrete Wavelet Transform for Improved Fingerprinting-Based Indoor Localization," in IEEE Sensors Journal, vol. 16, no. 21, pp. 7784-7791, Nov.1, 2016, doi: 10.1109/JSEN.2016.2602840.

[12] A. M. Martinez and A. C. Kak, "PCA versus LDA," IEEE Trans. Pattern Anal. Mach. Intell., 2001.

[13] J. Wang, J. Peng, J. G. Hwang, J. Park, and J. G. Park, "Indoor positioning method based on optimal channel state information subcarrier," in 2021 International Conference on Electronics, Information, and Communication, ICEIC 2021, 2021.