



Research on Positioning Algorithm Based on RPCA in Sparse Fingerprint Environment

Yaqin Xie*, Md Emadur Rahman Ekra, Tianyuan Gu, Xiaoli Wang

School of Electronics and Information Engineering,
Nanjing University of Information Science and Technology, Nanjing, China

*Corresponding Email: xyq@nuist.edu.cn

ABSTRACTS

The Wi-Fi fingerprinting-based indoor positioning technique offers the benefits of being cheap as well as simple to use. However, during signal collection, the presence of significant noise in the environment can cause fluctuations in signal strength measurements due to environmental variations. Additionally, a large number of fingerprints usually need to be collected to achieve high positioning accuracy. To address these issues, In a sparse fingerprint environment, this paper points to a placement technique based on an effective resilient principal component analysis algorithm (RPCA). First, purification is carried out using signal measurement weights in consideration of the outlier noise in the signals that were gathered, and the refined fingerprints are then saved in the fingerprint database. Secondly, given the high cost of collecting fingerprints, this paper generates some virtual fingerprints near reference points based on a transmission loss model, all of which are stored in an offline fingerprint database. Finally, adaptive K-value fingerprint matching is used to obtain the final results. The results show that the proposed algorithm can improve positioning accuracy in a sparse fingerprint environment.

© 2025 Tim Konferensi UNIKOM

ARTICLE INFO

Article History:

Received 28 Dec 2025

Revised 19 Jan 2025

Accepted 25 Jan 2025

Available online 01 Feb 2025

Publication date 01 Dec 2025

Keywords:

RSS,

RPCA,

Fingerprint Matching,

Indoor Positioning

1. INTRODUCTION

As the demand for location-based services (Isaia & Michaelides, 2023; NSC et al., 2022; Zekavat et al., 2021), To deliver dependable and highly accurate location services, several advanced positioning technologies and signal processing techniques have been put forward (Asaad & Maghdid, 2022; Leitch et al., 2023). In outdoor environments, Global Navigation Satellite System (GNSS) was applied to determine the location of users in outdoor environments. However, since the GNSS system is easily affected by obstructions such as buildings and has low local accuracy (Moradbeikie et al., 2021), satellite positioning cannot achieve satisfactory positioning performance indoors (Song et al., 2024). Many positioning systems based on Wi-Fi, Radio Frequency Identification (RFID), ultra-wideband, Bluetooth, and geomagnetism can achieve good positioning results in indoor environments (Abraha & Wang, 2024; Gismalla et al., 2022; Li et al., 2024; Ma et al., 2024; Shu et al., 2022; Xu et al., 2023; Zhuang et al., 2022). Broad availability and ease of deployment are two benefits of the commercial equipment used for location established on the Wi-Fi fingerprint methods. (Tang et al., 2023; Wang et al., 2023).

Utilizing the Wi-Fi signal's received signal strength (RSS) to pinpoint the target location is the basic premise behind Wi-Fi fingerprint placement. Wi-Fi fingerprint placement is based on the fundamental idea of using the received signal strength (RSS) of the Wi-Fi signal to figure out the target. In the offline phase, some reference points (RP) with known coordinates are selected, and the

RSS from different access points (AP) are collected at these RP locations to establish a fingerprint database. In the online phase, the target user obtains the RSS information from each AP, generates an online fingerprint, thereafter so the offline fingerprint database is compared to the online fingerprint. Then, using the best matched RSS result, the user's location is estimated. generally, a few test points (TP) with known coordinates are chosen for the setting of experiment, and the difference between these TPs' predicted and real positions is used to assess the preciseness of positioning. (Zhou et al., 2021).

(Wang et al., 2020) used Euclidean distance to select the nearest RPs to locate the target. But their work did not take random noise into account. (Bazin & Navaie, 2023) proposed a method for increasing indoor positioning accuracy that uses RIS-based RSS optimization to improve signal strength measurement and reduce noise. In order to improve accuracy, some work only needs to focus on fingerprint training. (Lin et al., 2023) improved the traditional fingerprint positioning technology and proposed a distributed positioning system based on clustering algorithm and feature extraction. However, the phase of collecting fingerprint information requires manual measurement and uploading, which consumes a lot of manpower and time. (Ezhumalai et al., 2021) suggested an initial MaxMean approach to generate an offline fingerprint database, choosing a number of the most powerful APs capable of covering the whole positioning region while offline. The offline fingerprint database can be made more resilient using this strategy, but it ignores unforeseen circumstances that can arise

during the online phase. (Duong et al., 2024) redesigned fingerprints using RSS principle component sets after training them using principle Component Analysis (PCA). This method improves positioning accuracy while reducing computational complexity, but does not consider nonlinear components. Applying Kernel Principal Component Analysis (KPCA) to train fingerprints, a new online fingerprint algorithm is proposed that can automatically determine the number of RPs in the online phase (Sun et al., 2024).

In recent years, some research has focused on updating strategies for offline fingerprint databases. (Liu et al., 2023) proposed a kriging interpolation method based on the mayfly optimization algorithm, which utilizes a small amount of fingerprint information to establish a variogram function and uses an improved mayfly optimization algorithm to fit the variogram function model, thereby generating an offline fingerprint database to improve interpolation accuracy and efficiency. (Lan et al., 2022) proposed an approach for updating fingerprint databases based on super-resolution network migration, which fully utilizes the spatial distribution characteristics of RSS at historical moments through the migration of neural networks to achieve fingerprint updates. In order to maintain the real-time performance of offline fingerprints, (Wei & Zheng, 2021) proposed an efficient Wi-Fi fingerprint crowdsourcing method to improve indoor positioning performance, with a focus on the scalability and participation of fingerprint database updates. (Bonthu & Mohan, 2023) used Wi-Fi fingerprint recognition and pedestrian dead reckoning (PDR) to improve indoor positioning accuracy and handle

environmental changes. (Ji et al., 2022) proposed a crowdsourcing based method for generating and updating indoor Wi-Fi fingerprint maps, which improves localization accuracy and reduces unnecessary data. (Lee et al., 2022) updated the database using the average number of occurrence of each AP gathered from user input data. The update approach has performed well in enhancing accuracy. Nonetheless, if some offline APs are uncertain or undetected during the online phase, or if the offline fingerprint database is not updated in a timely manner, it may affect the positioning effect. Also, there are initiatives centered on sensor fusion to enhance Wi-Fi-based indoor accuracy. (Abdellatif et al., 2023) proposes a multi-sensor fusion-based particle filter indoor positioning system that enhances positioning reliability and accuracy by fusing indoor positioning with intelligent mobile terminals. (Tinh et al., 2023) utilized users' motion patterns for target localization in Wi-Fi fingerprint-based algorithms. The Peer phone assisted positioning method overcomes the limitations of Wi-Fi methods by reducing significant errors through improved hardware (Liu et al., 2013). These algorithms work adequately indoors, but they rely on additional hardware support, not just Wi-Fi infrastructure.

Thus, in sparse fingerprint environments, this paper suggests a virtual fingerprint location approach based on the Robust Principal Component Analysis (RPCA) methodology. (Hou et al., 2022). The offline and online phases make up the majority of the positioning system. In the offline phase, RSS data of reference points are first obtained from different APs and denoised using the RPCA algorithm. Then, the data is optimized using the

dual-metric weight purification algorithm to generate a regenerated fingerprint database of reference points. During the online phase, the system obtains RSS data from the AP for testing points, interpolates it, and matches it with the fingerprint database. In order to improve the positioning accuracy, the system adaptively selects the K values of different APs and uses the WKNN algorithm for position estimation. At the same time, the RPCA algorithm and the dual-metric weight enhancing method are used again to optimize the results, and finally high-precision indoor positioning is achieved.

The following are this paper's primary study contributions:

1) This paper proposes a dual-metric purification algorithm based on RPCA to address the issue of outlier noise in collecting RSS data in practical environments. This algorithm can extract the more stable parts of the filtered data and use these data to generate fingerprints, thereby improving the accuracy of the positioning system.

2) In order to deal with the situation where there are fewer RPs or larger RP intervals in the positioning system, this paper proposes an RP regeneration fingerprint algorithm. To enhance the system positioning capabilities, this technique creates fingerprints at short intervals and saves them in an offline fingerprint database.

3) The algorithm implemented in this paper can significantly improve the positioning performance in the case of fewer RPs or larger RP intervals, which has obvious advantages over the existing fingerprint-based positioning algorithms. In addition, this algorithm's scalability is

good, and it can be used in paired with various other methods to increase positioning precision even further.

This paper's following aspects are structured as follows: The second section provides an overview of the positioning system framework proposed in this paper and provides some explanations of the process. The third section provides a detailed introduction to the algorithms used in the positioning system of this paper in the algorithm design section. Then, in the fourth section of result analysis and discussion, this paper conducted some simulations and actual measurements, and provided the results. Finally, summarize the entire text in the fifth section.

2. SYSTEM FRAMEWORK

In order to solve the problem of high time and labor costs in the fingerprint collection process and low positioning accuracy caused by a small number of reference points, this paper proposes a regenerated fingerprint positioning method based on RPCA which is suitable for sparse fingerprint environments, where traditional approaches struggle to provide reliable results. By using RPCA for denoising and purification, and incorporating a dual metric purification algorithm to further enhance the stability and reliability of the data, the system ensures cleaner and more accurate fingerprint information, alongside the generation of virtual fingerprints, this approach significantly improves the accuracy of indoor positioning while reducing the dependency on large manual data collection. The overall architecture of the proposed indoor positioning system is showed in Figure 1, presenting both the offline and online phases of the methodology.

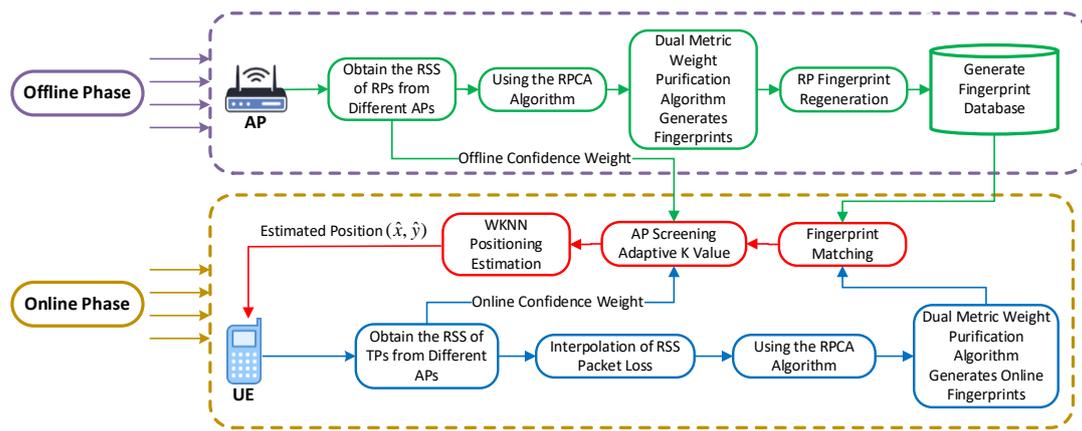


Fig. 1. Architecture of Indoor Positioning System.

Meanwhile, the majority of the key symbols used in this study have been explained in detail in Table 1 to help clarify the system. This paper assumes that there are N RPs and M APs, the distance between nearby RPs is d . The position of the i^{th} RP can be represented as $P_i = (x_i, y_i)$, And the vector formed by the RSS values of M APs received at this reference point is expressed as:

$$r_{SS}_i = [r_{SS_{i,AP_1}} \quad r_{SS_{i,AP_2}} \quad \dots \quad r_{SS_{i,AP_M}}]^T \quad (1)$$

Where, $r_{SS_{i,AP_m}}$ represents the collection of RSS values from the i^{th} RP to the m^{th}

AP, and the RSS vector obtained from all RPs can be expressed as:

$$RSS = [r_{SS_1} \quad r_{SS_2} \quad \dots \quad r_{SS_N}]^T \quad (2)$$

The position information of all RPs can form a position vector, which can be expressed as:

$$p = [p_1 \quad p_2 \quad \dots \quad p_N]^T \quad (3)$$

The offline fingerprint database is constructed by combining the RSS vector from equation(2) and the position information vector from equation(3), as shown in Table 1.

Table 1. Construction of offline fingerprint database.

The position of the i^{th} RP	Received a vector of RSS values from M APs at this reference point
$p_1: RP_1(x_1, y_1)$	$r_{SS_{1,AP_1}} \quad r_{SS_{1,AP_2}} \quad \dots \quad r_{SS_{1,AP_M}}$
$p_2: RP_2(x_2, y_2)$	$r_{SS_{2,AP_1}} \quad r_{SS_{2,AP_2}} \quad \dots \quad r_{SS_{2,AP_M}}$
$p_3: RP_3(x_3, y_3)$	$r_{SS_{3,AP_1}} \quad r_{SS_{3,AP_2}} \quad \dots \quad r_{SS_{3,AP_M}}$
\vdots	\vdots
$p_N: RP_N(x_N, y_N)$	$r_{SS_{N,AP_1}} \quad r_{SS_{N,AP_2}} \quad \dots \quad r_{SS_{N,AP_M}}$

2.1. Offline Phase

In the offline phase, the indoor location system proposed in this paper creates a Wi-Fi fingerprint database. All APs' RSS sample data is gathered at each RP. First, due to the fluctuation of RSS data, the RPCA algorithm is used to denoise the original data. Then, the dual-metric weight purification algorithm proposed in this paper is used to extract stable bands and calculate the mean to build an offline fingerprint database. In addition, in view of the time and labour costs in the collection process, or when there are fewer RPs, this paper proposes a simulated generation of area virtual fingerprints algorithm to generate virtual fingerprints to further improve the offline fingerprint database.

2.2. Online Phase

In the online phase, the user location estimation is achieved through Wi-Fi data obtained from various APs. First, based on the initial data packet loss scenario, the target user measures and records RSS data from multiple APs. Then, they use mean interpolation to reconstruct the RSS data. Next, the same RPCA denoising process and dual-metric purification algorithm as in the offline phase are used to generate online fingerprints. Then, the online fingerprint is matched with the offline fingerprint, and the difference of the offline and online data is extracted to generate confidence weights for screening the appropriate number of RP and AP. Lastly, the target position is estimated and the positioning results are returned using the weighted K-Nearest Neighbor (WKNN) methods.

3. METHODOLOGY

This section will cover RPCA and dual metric purification algorithms in the offline phase, RP regeneration fingerprint algorithm, online data processing, and WKNN positioning estimation.

3.1. RPCA and Dual Metric Purification Algorithm

3.1.1. RPCA Algorithm

In a complex indoor environment, the RSS signal is easily affected by environmental fluctuations, resulting in sharp noise in the sampled data. Figure 2. Displays an AP's RSS detection graph in real time, with the original RSS data displayed as the blue line. The graphic makes it evident that there is a significant amount of outlier noise in the original data. The original data must be denoised in order to increase positioning accuracy since these outlier noises will have a substantial impact on the positioning system's performance.

As shown on offline fingerprint database formulated by equation(1), rss_{i,AP_m} represents all the sampled data collected from the m^{th} AP at the i^{th} RP. Assuming the total number of sampling times is S, the sampled data at the i^{th} RP is:

$$rss_{i,AP_m} = [rss_{i,AP_m}(1) \quad rss_{i,AP_m}(2) \quad \dots \quad rss_{i,AP_m}(S)]^T \quad (4)$$

The $rss_{i,AP_m}(s)$ in equation(4) represents the s^{th} sampled value from the m^{th} AP. Actual testing shows that if the sampling value is below -95dBm, the data is difficult to detect. Therefore, the lost data is filled with -95dBm.

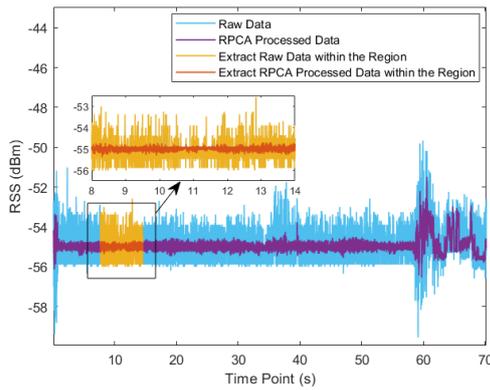


Fig. 2. Data Extraction Based on RPCA.

As mentioned above, sampling information is often affected by noise, so this paper uses the RPCA algorithm for denoising preprocessing. The data have to be a square matrix in order for the RPCA solution to use singular value decomposition, so it is necessary to reset the number of sampled values to s^2 . Specifically, s^2 should meet the dimensional requirements of the data, making the data matrix a square matrix for effective matrix decomposition and denoising processing, and

$$s = \lfloor \sqrt{S} \rfloor \tag{5}$$

Where, $\lfloor g \rfloor$ is the floor function. Equation(2) can be rewritten as follows:

$$\mathbf{R} = \begin{bmatrix} r_{SS_{i,AP_m}}(1) & r_{SS_{i,AP_m}}(s+1) & L & r_{SS_{i,AP_m}}((s-1) \times s + 1) \\ r_{SS_{i,AP_m}}(2) & r_{SS_{i,AP_m}}(s+2) & L & r_{SS_{i,AP_m}}((s-1) \times s + 2) \\ M & M & L & M \\ r_{SS_{i,AP_m}}(s) & r_{SS_{i,AP_m}}(s+s) & L & r_{SS_{i,AP_m}}((s-1) \times s + s) \end{bmatrix}_{s \times s} \tag{6}$$

The main idea of RPCA is to model the original matrix \mathbf{R} the sum of the low-rank matrix $\hat{\mathbf{R}}$ and the noise matrix \mathbf{E} . The RPCA model can be expressed as:

$$\begin{aligned} \min_{\hat{\mathbf{R}}, \mathbf{E}} \quad & rank(\hat{\mathbf{R}}) + \lambda \|\mathbf{E}\|_0 \\ \text{s.t.} \quad & \mathbf{R} = \hat{\mathbf{R}} + \mathbf{E}. \end{aligned} \tag{7}$$

Therefore, the principal component and outlier noise of the RSS raw data are consistent with the description of the RPCA model, so it is more reasonable to use RPCA for noise reduction. $rank(\hat{\mathbf{R}})$ represents the rank of the matrix, $\|\mathbf{E}\|_0$ represents the number of non-zero elements in the noise matrix, where $\lambda > 0$ controls the weight between $\hat{\mathbf{R}}$ and \mathbf{E} . Since the rank and non-zero norm of the matrix are non-convex, the optimization problem described by equation(7) is difficult for non-deterministic polynomials. Therefore, the above model can be transformed into a convex optimization problem:

$$\begin{aligned} \min_{\hat{\mathbf{R}}, \mathbf{E}} \quad & rank(\hat{\mathbf{R}}) + \lambda \|\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{R} = \hat{\mathbf{R}} + \mathbf{E}. \end{aligned} \tag{8}$$

When the convex optimization problem has a minimum value, the Inexact Augmented Lagrange Multiplier (IALM) can solve it. The construction of the enhanced Lagrangian function is as follows:

$$\begin{aligned} L(\hat{\mathbf{R}}, \mathbf{E}, \mathbf{Y}, \mu) = & \|\hat{\mathbf{R}}\|_* + \lambda \|\mathbf{E}\|_1 \\ & + \langle \mathbf{Y}, \mathbf{R} - \hat{\mathbf{R}} - \mathbf{E} \rangle \\ & + \frac{\mu}{2} \|\mathbf{R} - \hat{\mathbf{R}} - \mathbf{E}\|_F^2 \end{aligned} \tag{9}$$

\mathbf{Y} is the Lagrange multiplier, and μ is a positive scalar. The IALM method is used to alternately iterate the matrices $\hat{\mathbf{R}}$ and \mathbf{E} to converge. The iterative update equation of the matrices $\hat{\mathbf{R}}$, \mathbf{E} and \mathbf{Y} is:

$$\begin{aligned} \hat{\mathbf{R}}_{k+1} = & \arg \min_{\hat{\mathbf{R}}} L(\hat{\mathbf{R}}, \mathbf{E}_{k+1}, \mathbf{Y}_k, \mu_k) \\ = & \arg \min_{\hat{\mathbf{R}}} \|\hat{\mathbf{R}}\|_* \\ & + \frac{\mu_k}{2} \left\| \hat{\mathbf{R}} - (\mathbf{R} - \mathbf{E}_{k+1} + \frac{\mathbf{Y}_k}{\mu_k}) \right\|_F^2 \end{aligned} \tag{10}$$

$$E_{K+1} = \arg \min_E L(\hat{\mathbf{R}}_{k+1}, \mathbf{E}, \mathbf{Y}_k, \mu_k)$$

$$= \arg \min_E \|\mathbf{E}\|_1 \tag{11}$$

$$+ \frac{\mu_k}{2} \left\| \hat{\mathbf{R}} - \left(\mathbf{R} - \mathbf{E}_{k+1} + \frac{\mathbf{Y}_k}{\mu_k} \right) \right\|_F^2$$

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k (\mathbf{R} - \hat{\mathbf{R}}_{k+1} - \mathbf{E}_{k+1}) \tag{12}$$

Among them, in the process of solving the RPCA algorithm using the IALM method, the initialization parameter \mathbf{Y}_0 can be expressed as:

$$\mathbf{Y}_0 = \mathbf{R} / \mathbf{J}(\mathbf{R}) \tag{13}$$

The \mathbf{R} in the above equation can be obtained in equation(6) and $\mathbf{J}(\mathbf{R})$ can be solved by the following equation:

$$\mathbf{J}(\mathbf{R}) = \max(\|\mathbf{R}\|_2, \lambda^{-1} \|\mathbf{R}\|_\infty) \tag{14}$$

The denoised matrix $\hat{\mathbf{R}}$ can be expressed as:

$$\hat{\mathbf{R}} = \begin{bmatrix} r\hat{s}_{i,AP_m}(1) & r\hat{s}_{i,AP_m}(s+1) & L & r\hat{s}_{i,AP_m}((s-1) \times s + 1) \\ r\hat{s}_{i,AP_m}(2) & r\hat{s}_{i,AP_m}(s+2) & L & r\hat{s}_{i,AP_m}((s-1) \times s + 2) \\ M & M & L & M \\ r\hat{s}_{i,AP_m}(s) & r\hat{s}_{i,AP_m}(s+s) & L & r\hat{s}_{i,AP_m}((s-1) \times s + s) \end{bmatrix}_{s \times s} \tag{15}$$

The $\hat{\mathbf{R}}$ obtained from equation(15) can be restored to a vector and expressed as:

$$r\hat{s}_{i,AP_m} = [r\hat{s}_{i,AP_m}(1) \ r\hat{s}_{i,AP_m}(2) \ \dots \ r\hat{s}_{i,AP_m}(s^2)]^T \tag{16}$$

3.1.2. Dual Metric Purification Algorithm

In addition, as shown in Figure 2, the purple line represents the result of RPCA processing on the original data. In general, directly taking the mean to build a fingerprint database can achieve better positioning performance. In order to extract more accurate data to build fingerprints, this paper proposes a dual metric purification algorithm, which

aims to extract a more stable data part from a stable sequence to improve the accuracy of the fingerprint.

Firstly, taking the average of $r\hat{s}_i^{AP_m}$ obtained from equation(16) is taken as the first metric, which can be expressed as:

$$r\bar{s}_{i,AP_m} = \frac{1}{s^2} \sum_{n=1}^{s^2} r\hat{s}_{i,AP_m}(n) \tag{17}$$

Secondly, the RSS information values that can be obtained from the measured data can be retained to four decimal places or even more accurately. Based on this phenomenon, three decimal places, two decimal places, and one decimal place of all elements in the $r\hat{s}_i^{AP_m}$ vector are selected to be retained separately. On this basis, we choose the mode as the second metric, which can be expressed as:

$$r\bar{s}_{mode_i,AP_m}^c = Mode(round(r\hat{s}_{i,AP_m}, c)) \tag{18}$$

Based on equations (17) and (18), in order to obtain a more accurate reference metric, the RPCA-processed data obtained by equation(16) is used to calculate the weights of these two metrics to obtain a newly generated metric, which can be expressed as:

$$r\bar{s}_{cal} = k_{ave_i} r\bar{s}_{i,AP_m} + \sum_{c=1}^3 k_{mode_i,c} r\bar{s}_{mode_i,AP_m}^c \tag{19}$$

In equation(19), k_{ave} and $k_{mode,i}$ are the normalized weight coefficients of these two metrics, which can be expressed as:

$$k_{ave_i} = \frac{w_{ave_i}}{w_{ave_i} + \sum_{c=1}^3 w_{mode_{i,c}}} \tag{20}$$

$$k_{mode_{i,c}} = \frac{w_{mode_{i,c}}}{w_{ave_i} + \sum_{c=1}^3 w_{mode_{i,c}}}$$

Where,

$$w_{ave_i} = \frac{1}{\sum_{n=1}^{s^2} (r\hat{s}s_{i,AP_m}(n) - r\bar{s}s_{i,AP_m})} \quad (21)$$

$$w_{mod_{e_{i,c}}} = \frac{1}{\sum_{n=1}^{s^2} (r\hat{s}s_{i,AP_m}(n) - r\bar{s}s_{mod_{e_{i,c}},AP_m}^c)}$$

At this point, the average difference between the new metric rss_{cal} obtained from equation(19) and the data processed by RPCA is selected as the threshold, which can be expressed as:

$$\alpha_{i,AP_m} = \frac{1}{s^2} \left| \sum_{n=1}^{s^2} r\hat{s}s_{i,AP_m}(n) \right| \quad (22)$$

As shown in Figure 2, the yellow area is the range of the above dual-metric extraction RPCA. In this paper, the RPCA-processed data of a segment that satisfies the threshold of equation(22) and has the longest continuous length is extracted, and the mean is taken as the output of the dual-metric purification algorithm, denoted as η_{i,AP_m} .

In summary, after processing the offline fingerprint database, it can be expressed as:

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \dots \\ \boldsymbol{\eta}_M \\ \dots \\ \boldsymbol{\eta}_N \end{bmatrix} = \begin{bmatrix} \eta_{1,AP_1} & \eta_{1,AP_2} & \dots & \eta_{1,AP_M} \\ \eta_{2,AP_1} & \eta_{2,AP_2} & \dots & \eta_{2,AP_M} \\ \dots & \dots & \dots & \dots \\ \eta_{N,AP_1} & \eta_{N,AP_2} & \dots & \eta_{N,AP_M} \end{bmatrix}_{N \times M} \quad (23)$$

3.2. RP Regenerated Fingerprint

To address the second problem, that is, achieving high-precision positioning with fewer RPs, this paper proposes a regenerative fingerprint algorithm. The interpolated RSS data used by this algorithm is generated based on the log-normal path loss model (Li et al., 2020), and its equation is:

$$RSS = P_t - P_{d_0} - 10n \log\left(\frac{d}{d_0}\right) - X_\sigma \quad (24)$$

In equation(24) above, the transmission power P_t is set to 25dBm, P_{d_0} is the path loss at the reference distance d_0 , usually $d_0 = 1 m$. P_{d_0} is set to 37.7dBm, and the path loss index n is 4. X_σ is Gaussian noise with a mean of 0 and a standard deviation of 5dBm.

As shown in Figure 3, the area covered by a reference point is selected in this paper, the distance between two nearby reference points is set to Δ_{RP} , the number of virtual RPs is n_{vir}^2 , and the interval Δ_{vir} for generating new virtual RPs is:

$$\Delta_{vir} = \frac{\Delta_{RP}}{n_{vir}} \quad (25)$$

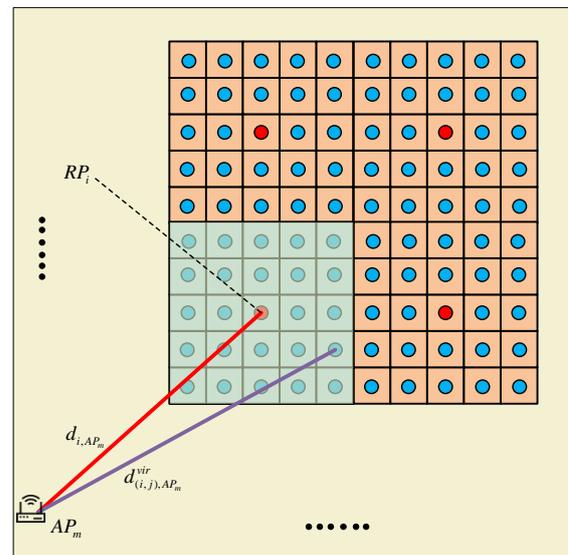


Fig. 3. Schematic Diagram of RP Regenerated Fingerprint.

RSS fingerprints were collected from the m^{th} AP at the position of the i^{th} RP, and the distance between the AP and this RP is d_i . In this paper, s^2 in equation(5) is selected as the number of simulated

sampling times, so the vector of the fingerprint point is:

$$rss_{i,AP_m}^{vir} = [rss_{i,AP_m}^{vir}(1) \quad rss_{i,AP_m}^{vir}(2) \quad \dots \quad rss_{i,AP_m}^{vir}(s^2)]^T \quad (26)$$

In equation(26) $rss_{vir,i}^m$ $i = 1, 2, \dots, s^2$, is the RSS value simulated based on the distance from AP to RP as d_i . By applying the RPCA algorithm and the dual metric refinement algorithm to the vector of equation(26), a new fingerprint point $\phi_{vir,i}^{AP_m}$ can be obtained. In this paper, we set the noise to be fixed within the coverage area of a real RP point, which can be expressed as:

$$\Delta_{noise} = \eta_{i,AP_m} - \eta_{i,AP_m}^{vir} \quad (27)$$

Due to the known location information of the real RP and AP, and the fact that the virtual fingerprint interval Δ_{vir} has been obtained at equation(25), it is easy to obtain the location information of the virtual reference point and the distance $d_{(i,j),AP_m}^{vir}$ from this AP_m to the j^{th} virtual reference point. Substitute this distance into (24), and simultaneously execute the RPCA algorithm and the dual metric purification algorithm to generate the virtual fingerprint point $\phi_{i,j}^{AP_m}$. Combined with the noise obtained from equation (27), this virtual fingerprint can be represented as:

$$\eta_{(i,j),AP_m}^{vir} = \eta_{(i,j),AP_m} + \Delta_{noise} \quad (28)$$

Therefore, the matrix constructed by the virtual fingerprint generated at the i^{th} RP is:

$$\hat{\eta}_{i,AP_m}^{vir} = \begin{bmatrix} \eta_{(i,1),AP_m}^{vir} & \eta_{(i,n_{vir}+1),AP_m}^{vir} & L & \eta_{(i,j_{vir}^2-n_{vir}+1),AP_m}^{vir} \\ \eta_{(i,2),AP_m}^{vir} & \eta_{(i,n_{vir}+2),AP_m}^{vir} & L & \eta_{(i,j_{vir}^2-n_{vir}+2),AP_m}^{vir} \\ M & M & L & M \\ \eta_{(i,n_{vir}),AP_m}^{vir} & \eta_{(i,2n_{vir}),AP_m}^{vir} & L & \eta_{(i,j_{vir}^2),AP_m}^{vir} \end{bmatrix}_{n_{vir} \times n_{vir}} \quad (29)$$

Considering the geographical environment and other factors in the real environment, the RSS data in η_{i,AP_m}^{vir} in equation(29) is partially deleted according to the test environment. The number of RSS fingerprints after the processing of the virtual fingerprint matrix is planned to be r^2 , Therefore, the result of the matrix constructed by the virtual fingerprint generated by the i^{th} RP after processing is:

$$\hat{\eta}_{i,AP_m}^{vir} = \begin{bmatrix} \eta_{(i,1),AP_m}^{vir} & \eta_{(i,r+1),AP_m}^{vir} & L & \eta_{(i,j^2-r+1),AP_m}^{vir} \\ \eta_{(i,2),AP_m}^{vir} & \eta_{(i,r+2),AP_m}^{vir} & L & \eta_{(i,j^2-r+2),AP_m}^{vir} \\ M & M & L & M \\ \eta_{(i,r),AP_m}^{vir} & \eta_{(i,2r),AP_m}^{vir} & L & \eta_{(i,j^2),AP_m}^{vir} \end{bmatrix}_{r \times r} \quad (30)$$

Figure 4 shows the effect of regenerated fingerprint signal strength under the condition of $\Delta_{vir} = 0.8m$. The effect of Figure 4 is that the AP and the RP positions in the green area in Figure 3 are selected for deployment.

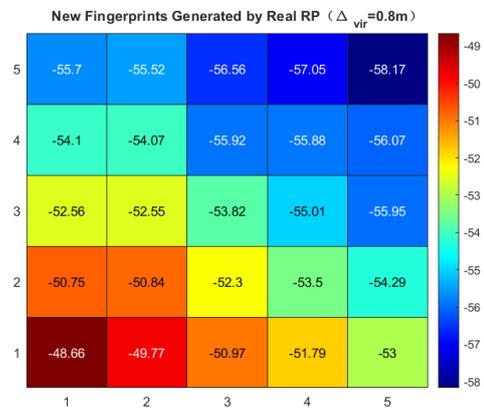


Fig. 4. Fingerprint Signal Strength of Simulated Area with Virtual Fingerprint Interval of 0.8m.

If you want to improve the granularity of the regenerated fingerprint, you need to adjust the number of regenerated RPs n_{vir}^2 to build an offline fingerprint database.

Similar to Figure 4, Figure 5 is a schematic diagram of the effect of adjusting the fingerprint signal strength generated by $\Delta_{vir} = 0.1m$.

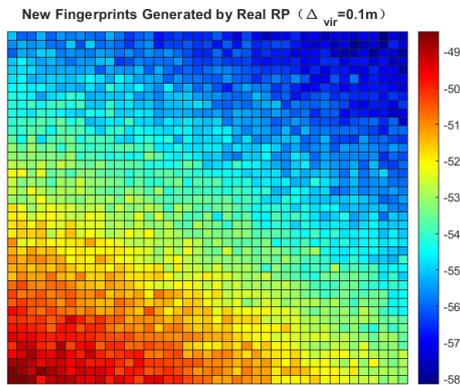


Fig. 5. Fingerprint Signal Strength of Simulated Area with Virtual Fingerprint Interval of 0.1m.

In summary, by combining the newly generated virtual fingerprints with the matching virtual fingerprint location information, the offline fingerprint database B_{off} can be expressed as:

$$B_{off} = \begin{bmatrix} \hat{\eta}_{1,AP1}^{vir} & \hat{\eta}_{1,AP2}^{vir} & L & \hat{\eta}_{1,APM}^{vir} \\ \hat{\eta}_{2,AP1}^{vir} & \hat{\eta}_{2,AP2}^{vir} & L & \hat{\eta}_{2,APM}^{vir} \\ M & M & L & M \\ \hat{\eta}_{N,AP1}^{vir} & \hat{\eta}_{N,AP2}^{vir} & L & \hat{\eta}_{N,APM}^{vir} \end{bmatrix}_{N \times M} \quad (31)$$

3.3. Online Data Processing

3.3.1. Processing of Online Data

During the actual testing process, the situation was extremely complex, such as the RSS sampling is lost due to equipment instability and slow entry into working state. Directly deleting the data will cause waste. As mentioned in Section 3.1.1. above, the data is difficult to detect when the sampling value is lower than -95dBm. As shown in Figure 6, the packet loss data is represented by [] and filled with -95dBm. For the same AP, if the

number of recorded sampling values not greater than -95dBm exceeds half of the total number of samples, the AP is regarded as an abnormal AP, and its weight is set to 0 before positioning, marking it as untrustworthy. These abnormal APs will seriously affect the positioning accuracy and even cause positioning failure. Therefore, the information of these abnormal APs will not be considered when generating online fingerprints. Finally, the processed data will be applied with RPCA and dual metric purification algorithms to generate online fingerprints.

Therefore, similar to equation(31), the fingerprint database generated by online data can be expressed as:

$$B_{on} = \begin{bmatrix} \eta_{1,AP1}^{on} & \eta_{1,AP2}^{on} & L & \eta_{1,APM}^{on} \\ \eta_{2,AP1}^{on} & \eta_{2,AP2}^{on} & L & \eta_{2,APM}^{on} \\ M & M & L & M \\ \eta_{T,AP1}^{on} & \eta_{T,AP2}^{on} & L & \eta_{T,APM}^{on} \end{bmatrix}_{T \times M} \quad (32)$$

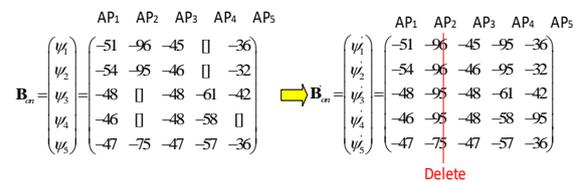


Fig. 6. Online Data Preprocessing.

3.3.2. AP and RP Screening

Based on the previous section, this paper found that the credibility of each AP in the actual measurement may not be the same, that is, the AP signals received by the target user are all reliable. In general, the farther the distance between the AP and the target user or the greater the interference, the greater the fluctuation of the received signal. Secondly, this paper observes that even in complex indoor environments, the

stability of signals received by the same RP from different APs varies greatly; Similarly, even when the same AP transmits signals to different RPs, there are significant differences in stability. Therefore, this paper constructs a set of confidence matrices by collecting the variance of the sampling signals from each AP in each RP, and calculates the confidence weights of each RP in each AP.

After completing the offline data preprocessing, each confidence matrix composed of variances can be expressed as:

$$\mathbf{v}_{off} = \begin{bmatrix} \mathbf{v}_1^{off} \\ \mathbf{v}_2^{off} \\ \mathbf{M} \\ \mathbf{v}_N^{off} \end{bmatrix} = \begin{bmatrix} v_{1,AP_1}^{off} & v_{1,AP_2}^{off} & L & v_{1,AP_M}^{off} \\ v_{2,AP_1}^{off} & v_{2,AP_2}^{off} & L & v_{2,AP_M}^{off} \\ \mathbf{M} & \mathbf{M} & L & \mathbf{M} \\ v_{N,AP_1}^{off} & v_{N,AP_2}^{off} & L & v_{N,AP_M}^{off} \end{bmatrix}_{N \times M} \quad (33)$$

In equation(33), v_n ($n \in 1, 2, \dots, N$) represents the weight vector of the n^{th} RP across all APs. $v_n^{AP_m}$ is the variance collected from the m^{th} AP at the n^{th} RP. Among them, $v_n^{AP_m}$ can be expressed as:

$$v_{n,AP_m}^{off} = \text{var}(rss_{n,AP_m}^{vir}(1), rss_{n,AP_m}^{vir}(2), \dots, rss_{n,AP_m}^{vir}(S)) \quad (34)$$

Then the confidence weight matrix can be expressed as:

$$\mathbf{W}_{off} = \begin{bmatrix} w_1^{off} \\ w_2^{off} \\ \mathbf{M} \\ w_N^{off} \end{bmatrix} = \begin{bmatrix} w_{1,AP_1}^{off} & w_{1,AP_2}^{off} & L & w_{1,AP_M}^{off} \\ w_{2,AP_1}^{off} & w_{2,AP_2}^{off} & L & w_{2,AP_M}^{off} \\ \mathbf{M} & \mathbf{M} & L & \mathbf{M} \\ w_{N,AP_1}^{off} & w_{N,AP_2}^{off} & L & w_{N,AP_M}^{off} \end{bmatrix}_{N \times M} \quad (35)$$

Among them, $w_{off}^{AP_M, n}$ is the confidence weight of the M^{th} AP at the n^{th} RP point, which can be expressed as:

$$w_{n,AP_m}^{off} = \frac{1}{v_{n,AP_m}^{off} + 1} / \sum_{j=1}^N \left(\sum_{i=1}^M \frac{1}{v_{j,AP_i}^{off} + 1} \right) \quad (36)$$

If $w_{off}^{AP_M, n} = 0$ occurs, it means that the information collected by the n^{th} RP point from the M^{th} AP is unreliable and will not participate in subsequent position estimation. Finally, the confidence weight corresponding to the regenerated fingerprint proposed in the paper is the same as the weight of the original RP point of the regenerated fingerprint.

In the online phase, the reliability of APs is also different. Therefore, this paper additionally calculated the variance of data collected from each AP at TP as the confidence weight for online APs. After completing the online data preprocessing, the variance of the sampled data of each AP can be expressed as:

$$v_{on} = (v_{TP,AP_1}^{on}, v_{TP,AP_2}^{on}, L, v_{TP,AP_M}^{on}) \quad (37)$$

In the above equation(37) is the variance of the data sampled from the M^{th} AP, which can be expressed as:

$$v_{TP,AP_m} = \text{var}(rss_{TP,AP_m}(1), rss_{TP,AP_m}(2), \dots, rss_{TP,AP_m}(S)) \quad (38)$$

Then, the AP weight of the online fingerprint can be obtained as:

$$\mathbf{W}_{on} = (w_{TP,AP_1}^{on}, w_{TP,AP_2}^{on}, \dots, w_{TP,AP_M}^{on}) \quad (39)$$

Among them, $w_{on}^{AP_M}$ is the confidence weight of the M^{th} AP, which can be expressed as:

$$w_{AP_m}^{on} = \frac{1}{v_{TP,AP_m}^{on} + 1} / \sum_{i=1}^M \frac{1}{v_{TP,AP_i}^{on} + 1} \quad (40)$$

3.4. WKNN Position Estimation

The specifications of regenerated fingerprints created by location information differ in the offline fingerprint database. Therefore, in order

to better illustrate the algorithm in this paper, this section uses the traditional offline fingerprint database location estimation to illustrate. Each RP's fingerprint in the offline fingerprint database is compared to the user's online fingerprint throughout the online phase, and the distance calculated using Euclidean geometry between them is computed as a similarity metric. The obtained similarity vector \mathbf{d} can be expressed as:

$$\mathbf{d} = \begin{bmatrix} d_{1,AP_1} & d_{1,AP_2} & L & d_{1,AP_M} \\ d_{2,AP_1} & d_{2,AP_2} & L & d_{2,AP_M} \\ M & M & L & M \\ d_{N,AP_1} & d_{N,AP_2} & L & d_{N,AP_M} \end{bmatrix}_{N \times M} \quad (41)$$

In equation(41), $d_{n,M}$ represents the weighted Euclidean distance between the online fingerprint generated by the M^{th} AP and the offline fingerprint collected by the n^{th} RP. Among them, $d_{n,M}$ can be expressed as:

$$d_{n,AP_m} = \sqrt{w_{n,AP_m}^{off} w_{AP_m}^{on} \left(\hat{\mathbf{q}}_{n,AP_m}^{vir} - \hat{\mathbf{q}}_{T,AP_m}^{on} \right)^2} \quad (42)$$

The position of TP typically is estimated using K RPs with the least Euclidean distance in conventional KNN position estimation methods. Based on historical data, the K number is often set at two or three. The approximate location of TP may be represented as:

$$\hat{p}(x, y) = \frac{1}{k} \sum_{i=1}^k p_i(x_i, y_i) \quad (43)$$

On the basis of traditional KNN position estimation, in order to handle the situation where the TP position is not at the center of gravity of RP, the WKNN algorithm uses the weighted mean of K RPs to estimate the user position. The

estimated position of TP can be expressed as:

$$\hat{p}(x, y) = \frac{\omega_i}{k} \sum_{i=1}^k p_i(x_i, y_i) \quad (44)$$

Among them, ω_i in the above equation (44) is:

$$\omega_i = \frac{M}{\sum_{j=1}^M \frac{1}{d_{i,AP_j}}} \bigg/ \sum_{i=1}^k \left(\frac{M}{\sum_{j=1}^M \frac{1}{d_{i,AP_j}}} \right) \quad (45)$$

For the selection of k value in equation (45), the common k value selection is

$$k \in (1, 2, 3, 4) \quad (46)$$

For, both KNN and WKNN use a fixed K value, but this strategy may reduce the positioning performance in some cases. Figure 7 shows several cases of K value selection during fingerprint positioning. For example, for Figure 7 (a), K value 1 can achieve the best effect. Figure 7 (b) selects K as 2 and Figure 7 (c) uses K as 3 or 4 to further improve the positioning accuracy.

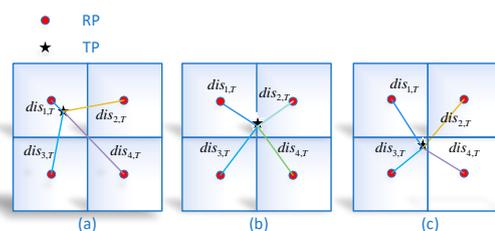


Fig. 7. Several Scenarios of K Value Selection.

In the test area, K is usually between the 4 RPs. In the similarity vector \mathbf{d} obtained in equation(41), the four minimum values are selected to correspond to $dis_{1,T}$, $dis_{2,T}$, $dis_{3,T}$ and $dis_{4,T}$ in Figure 7 (in the figure, $dis_{1,T} < dis_{2,T} < dis_{3,T} < dis_{4,T}$). Now we take the difference between their

pairwise similarities, which can be expressed as:

$$\begin{cases} \Delta_{1,2} = \text{abs}(dis_{1,T} - dis_{2,T}) \\ \Delta_{1,3} = \text{abs}(dis_{1,T} - dis_{3,T}) \end{cases} \quad (47)$$

In this paper, depending on the real similarity scenario, the threshold ε is established, and the selection of K value is determined according to the relationship in equation(47), In this paper, the threshold ε is set according to the actual similarity situation, and the selection of K value is determined according to the relationship in equation(47), which can be expressed as:

$$k = \begin{cases} 1 & \Delta_{1,2} \geq \varepsilon \\ 2 & \Delta_{1,2} < \varepsilon \ \& \ \Delta_{1,3} \geq \varepsilon \\ 3 & \text{else} \end{cases} \quad (48)$$

4. RESULTS AND DISCUSSION

4.1. Simulation Configuration and Parameter Settings

The simulation layout diagram is shown in Figure 8, and the testing area space is set to 40m×20m. The testing area will be divided into 50 identical grids with a size of 4m×4m, and 10 APs and 50 RPs will be deployed at the positions shown in the diagram. The generated RSS data used in this simulation is given by equation(24). In addition, during the data collection process, Sharp isolated outliers that differ from the norm are present. In response to this situation, this paper randomly inserts larger sharp noise based on equation(24), which can be expressed as:

$$RSS' = RSS - X_{\sigma'} \\ = \left(P_r - P_{d_0} - 10n \log \left(\frac{d}{d_0} \right) - X_{\sigma} \right) - X_{\sigma'} \quad (49)$$

In the above equation(49), $X_{\sigma'}$ is a Gaussian noise with a mean of 0 and a standard deviation of 15dBm, and the other parameters are the same as equation(24). According to the characteristics of the measured data, outlier noise generally accounts for 10%. Therefore, in the offline phase, among the 1000 data obtained by each RP from each AP, 90% of the data is generated by equation(24) and the remaining 10% of the data is generated by equation(49). In addition, the 10% data representing outlier noise should be randomly distributed in the total sampled data. For the grid of the regenerated fingerprint, n_{vir} takes a value of 41, and all the data is generated by equation(24).

This paper compares three methods with the algorithm proposed in this paper.

- 1) KNN (Ilias et al., 2016): A classic indoor positioning solution that estimates the target position by calculating the Euclidean distance between the target fingerprint and RP.
- 2) WKNN (Yen et al., 2017): Based on KNN, WKNN additionally considers the weight of RP.
- 3) WED-WKNN (Wang et al., 2020): Based on WKNN, WED-WKNN introduces signal weights to balance the RSS difference and Euclidean distance by assigning weights to the RSS of different APs.

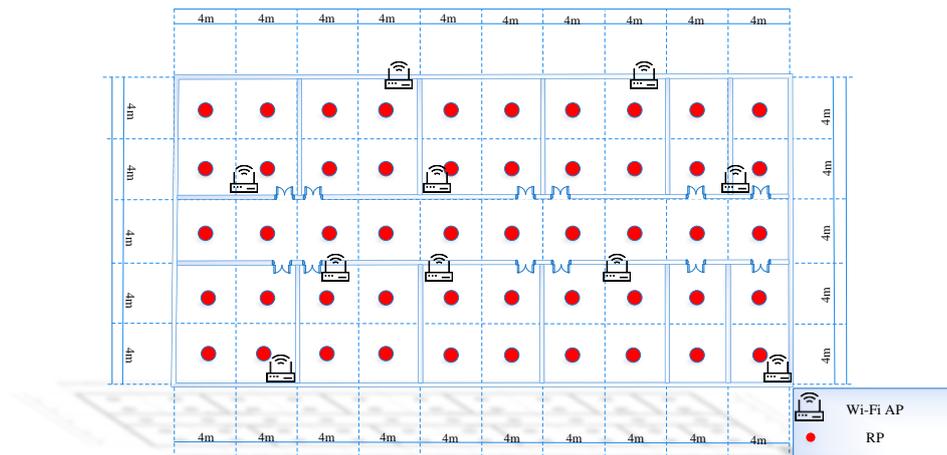


Fig. 8. Schematic Diagram of Simulation Layout.

4.2. Simulation Result Analysis

Table 2 shows the average fingerprint data of 5 APs for all RPs, in dBm. The reference fingerprint is the fingerprint data without noise calculated by equation (24), while the real fingerprint is the fingerprint data with sparse noise calculated by equation(49). The last two

rows show the fingerprint data after RPCA denoising and dual metric purification respectively. Compared with directly using RPCA denoising, the fingerprint data generated by the algorithm proposed in this paper is closer to the reference value, thereby effectively improving the purity of the offline fingerprint database.

Table 2. Mean fingerprint values of all reference points (dBm).

Fingerprint Type	AP1	AP2	AP3	AP4	AP5
Theoretical value	-60.5	-58.7	-55.5	-57.8	-53.8
Actual measurement value	-63.4	-61.4	-58.4	-60.6	-56.5
RPCA fingerprint	-59.5	-57.6	-54.4	-56.7	-52.7
Fingerprint of this paper	-60.8	-59.1	-55.8	-57.9	-54.1

In general, the offline phase and the online phase often use the same parameters, such as the same random noise parameters. However, the RSS signal changes over time, resulting in certain differences in environmental conditions. The actual situation is that the time of online and offline phases is usually not consistent. Figure 9 shows the

positioning performance results under different test noise conditions.

As shown in the figure below, the positioning error of the proposed algorithm increases with the increasing test noise. When the test noise is 8dBm, the positioning errors of the proposed algorithm, WEDWKNN algorithm, WKNN algorithm and KNN algorithm

are 0.64m, 1.13m, 1.24m and 1.62m respectively. Compared to alternative positioning systems, the one recommended algorithm's positioning error is significantly superior. The above results show that the proposed algorithm has certain noise resistance performance.

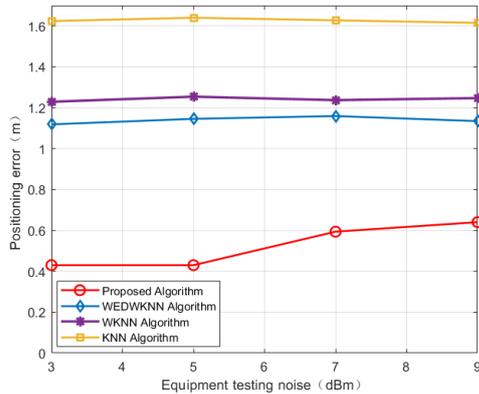


Fig. 9. Impact of Online Fingerprint Noise Changes on Positioning Results.

In addition, 1000 test points TP were selected for testing in an experimental environment with $\sigma = 5 \text{ dBm}$, test noise of 15 dBm, and grid size of 4m. The Cumulative Distribution Function (CDF) of the positioning error was plotted in Figure 10. The probability of the positioning accuracy of the methods proposed in this paper, WEDWKNN algorithm, WKNN algorithm, and KNN algorithm being less than 1m is 81%, 49%, 45%, and 30%, respectively. Secondly, 90% of the TP positioning errors in this scheme are less than 1.27m, while the positioning accuracies of other schemes are 63%, 57%, and 43%, respectively. In addition, besides the selection of AP, the algorithm in this paper can adaptively select K

values based on the matching results, and the algorithm in this paper may choose appropriate RP to reduce positioning errors.

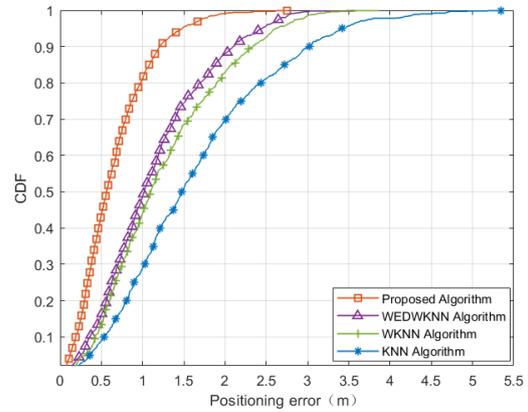


Fig. 10. Simulation of Positioning Errors of Three Algorithms.

4.3. Measured Configuration and Parameter Settings

To confirm the effectiveness of the RPCA regeneration fingerprint-based sparse environment placement technique disclosed in this study, this study was tested in the A103 conference room of the Information Technology Building of Nanjing University of Information Science and Technology. The model diagram of the test space is shown in Figure 11. The size of the test environment is 8m × 10.7m. The portable computer used for data processing is configured with Intel Core i7 10750H CPU and 16GB RAM, and the operating environment is MATLAB R2024a. Other detailed configurations are shown in Table 3.

Table 3. Test configuration information.

Target Object	Purpose
Test Area Dimensions	8m × 10.7m
Number of APs	5
Number of RPs	6
Collection Software	CSI Tool
Intel 5300 NIC (Three Antennas)	Network card used to build AP/RP industrial host
X30A	Industrial host that implements network card sending and receiving functions
HP Pavilion Gaming Laptop 15	Implement the positioning system described in this paper

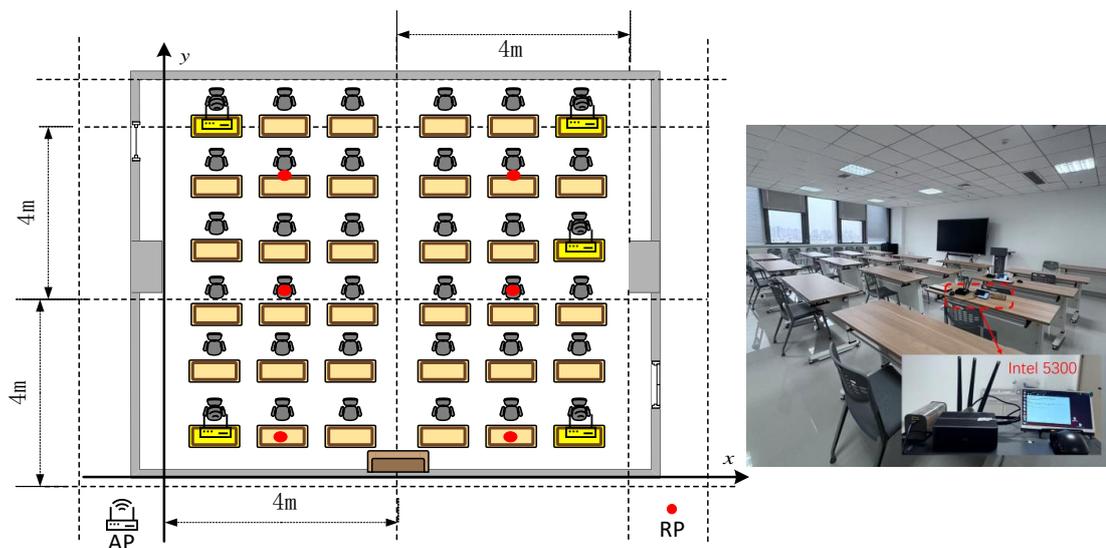


Fig. 11. AP and RP Layout and Actual Test Scenario.

4.4. Measured Performance Analysis

Based on the above settings, the collected data are positioned using WEDWKNN, WKNN, KNN and the algorithm proposed in this paper. As shown in Figure 11, the AP and RP

positions are deployed. Then the coordinate system is established, and 30 points are selected as TPs in the test area. Among them, Figure 12 shows the positioning error plots under the three methods.

As shown in Figure 12, this paper uses four methods for positioning. The average positioning errors of the proposed algorithm, WEDWKNN, WKNN and KNN are 1.38m, 1.47m, 1.47m and 2.01m respectively, and the average error of the proposed algorithm is the smallest. The maximum error is also lower than that of other methods, and the minimum error performance is slightly worse than that of the other two algorithms. In summary, for 30 TPs, the positioning performance of the proposed method in this paper has a more obvious performance improvement compared with the first two methods in a sparse fingerprint environment.

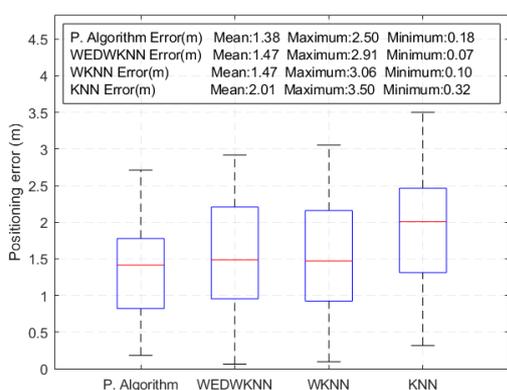


Fig. 12. Positioning Error Diagram of Different Methods under Actual Measurement Environment.

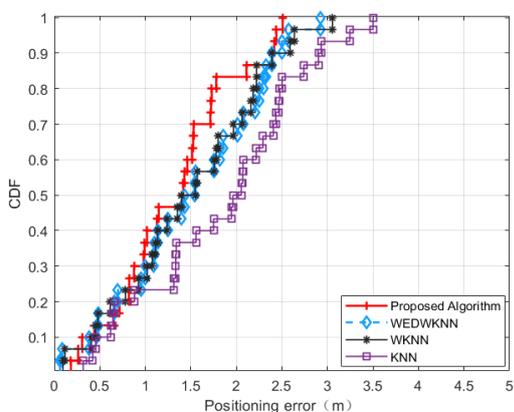


Fig. 13. CDF Plots of Three Tested Methods.

The CDF plots of the positioning errors of the recommended approach are displayed in Figure 13, WEDWKNN algorithm, WKNN algorithm, and KNN algorithm in the actual test scenario. According to Figure 13, the probability of the positioning accuracy of the proposed algorithm, WEDWKNN, WKNN, and KNN being lower than 2m is 83.3%, 66.7%, 66.7%, and 50%, respectively. In actual testing scenarios, compared to the other three algorithms, the positioning accuracy of the algorithm proposed in this paper is higher. Therefore, for sparse fingerprint scenarios, this paper's proposed algorithm has a high positioning accuracy.

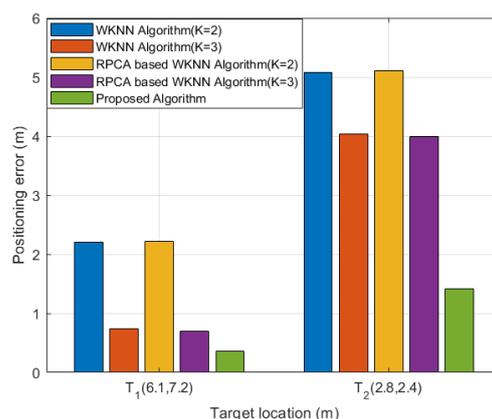


Fig. 14. Target Positioning Effect Diagram.

Figure 14 takes two target points as an example to analyze the performance of the algorithm proposed in this paper, the WKNN algorithm under different K values, and the WKNN algorithm after RPCA processing. The results show that the algorithm proposed in this paper has the smallest positioning error at the two target points. While the WKNN algorithm's positioning mistakes and the WKNN algorithm based on RPCA processing both decrease as K values

grow, the positioning accuracy is still limited. At the same time, the WKNN algorithm based on RPCA processing has a significant improvement compared with the WKNN algorithm alone, which shows that RPCA processing effectively improves the accuracy of the model at these two target points.

5. CONCLUSION

In a sparse fingerprint environment, the traditional Wi-Fi indoor fingerprint positioning method based on matching leads to unsatisfactory positioning results due to signal measurement noise and the long RP spacing or insufficient RP number. To address this issue, this paper proposes a positioning method based on RPCA regeneration fingerprint to solve

this problem. First, the RSS fingerprint data is filtered and denoised by combining the RPCA algorithm with the dual-metric purification algorithm to reduce the impact of noise. Then, according to the RP position information of the actual scene, the virtual fingerprint is regenerated using distance simulation and stored in the offline fingerprint database. Next, the online data is also processed similarly, and then the AP and RP are screened. Finally, the WKNN improved by the algorithm in this paper is used for positioning to obtain the positioning result. The simulation results show that the regenerated fingerprint positioning system can significantly improve positioning accuracy in a sparse real fingerprint environment or with a long RP interval.

REFERENCES

- Abdellatif, A. G., Salama, A. A., Zied, H. S., Elmahallawy, A. A., & Shawky, M. A. (2023). An improved indoor positioning based on crowd-sensing data fusion and particle filter. *Physical Communication*, 61, 102225.
- Abraha, A. T., & Wang, B. (2024). A Survey on Scalable Wireless Indoor Localization: Techniques, Approaches and Directions. *Wireless Personal Communications*, 136(3), 1455-1496.
- Asaad, S. M., & Maghdid, H. S. (2022). A comprehensive review of indoor/outdoor localization solutions in IoT era: Research challenges and future perspectives. *Computer Networks*, 212, 109041.
- Bazin, S., & Navaie, K. (2023). Improving Indoor Positioning Accuracy Using RIS-based RSS Optimization. 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit),
- Bonthu, B., & Mohan, S. (2023). Combining Wi-Fi Fingerprinting and Pedestrian Dead Reckoning to Mitigate External Factors for a Sustainable Indoor Positioning System. *Sustainability*, 15(14), 10943.
- Duong, T. H., Trinh, A. V., & Hoang, M. K. (2024). Efficient and Accurate Indoor Positioning System: A Hybrid Approach Integrating PCA, WKNN, and Linear Regression. *Journal of Communications*, 19(1).
- Ezhumalai, B., Song, M., & Park, K. (2021). An efficient indoor positioning method based on Wi-Fi RSS fingerprint and classification algorithm. *Sensors*, 21(10), 3418.

- Gismalla, M. S. M., Azmi, A. I., Salim, M. R. B., Abdullah, M. F. L., Iqbal, F., Mabrouk, W. A., Othman, M. B., Ashyap, A. Y., & Supa'at, A. S. M. (2022). Survey on device to device (D2D) communication for 5GB/6G networks: Concept, applications, challenges, and future directions. *IEEE Access*, *10*, 30792-30821.
- Hou, C., Xie, Y., & Zhang, Z. (2022). FCLoc: A novel indoor Wi-Fi fingerprints localization approach to enhance robustness and positioning accuracy. *IEEE Sensors Journal*, *23*(7), 7153-7167.
- Ilias, B., Shukor, S. A. A., Adom, A. H., Abd Rahim, N., Ibrahim, M. F., & Yaacob, S. (2016). Indoor mobile robot localization using KNN. 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE),
- Isaia, C., & Michaelides, M. P. (2023). A review of wireless positioning techniques and technologies: From smart sensors to 6G. *Signals*, *4*(1), 90-136.
- Ji, Y., Zhao, X., Wei, Y., & Wang, C. (2022). Generating indoor Wi-Fi fingerprint map based on crowdsourcing. *Wireless Networks*, *28*(3), 1053-1065.
- Lan, T., Wang, X., Chen, Z., Zhu, J., & Zhang, S. (2022). Fingerprint augment based on super-resolution for WiFi fingerprint based indoor localization. *IEEE Sensors Journal*, *22*(12), 12152-12162.
- Lee, S.-H., Kim, W.-Y., & Seo, D.-H. (2022). Automatic self-reconstruction model for radio map in Wi-Fi fingerprinting. *Expert Systems with Applications*, *192*, 116455.
- Leitch, S. G., Ahmed, Q. Z., Abbas, W. B., Hafeez, M., Laziridis, P. I., Sureephong, P., & Alade, T. (2023). On indoor localization using WiFi, BLE, UWB, and IMU technologies. *Sensors*, *23*(20), 8598.
- Li, H., Qian, Z., Tian, C., & Wang, X. (2020). TILoc: Improving the robustness and accuracy for fingerprint-based indoor localization. *IEEE Internet of Things Journal*, *7*(4), 3053-3066.
- Li, X., Pang, H., Li, G., Jiang, J., Zhang, H., Gu, C., & Yuan, D. (2024). Wireless Positioning: Technologies, Applications, Challenges, and Future Development Trends. *CMES-Computer Modeling in Engineering & Sciences*, *139*(2), 1135-1166.
- Lin, H., Purmehdi, H., Fei, X., Zhao, Y., Isac, A., Louafi, H., & Peng, W. (2023). Two-stage clustering for improve indoor positioning accuracy. *Automation in Construction*, *154*, 104981.
- Liu, H., Yang, J., Sidhom, S., Wang, Y., Chen, Y., & Ye, F. (2013). Accurate WiFi based localization for smartphones using peer assistance. *IEEE Transactions on Mobile Computing*, *13*(10), 2199-2214.
- Liu, W., Zhang, Y., Deng, Z., & Zhou, H. (2023). Low-cost indoor wireless fingerprint location database construction methods: A review. *IEEE Access*, *11*, 37535-37545.
- Ma, W., Fang, X., Liang, L., & Du, J. (2024). Research on indoor positioning system algorithm based on UWB technology. *Measurement: Sensors*, *33*, 101121.
- Moradbeikie, A., Keshavarz, A., Rostami, H., Paiva, S., & Lopes, S. I. (2021). GNSS-free outdoor localization techniques for resource-constrained IoT architectures: A literature review. *Applied Sciences*, *11*(22), 10793.
- NSC, J., Wahab, N., Sunar, N., Ariffin, S., Wong, K., & Aun, Y. (2022). Indoor Positioning System: A Review. *International Journal of Advanced Computer Science and Applications*, *13*(6).

- Shu, M., Chen, G., Zhang, Z., & Xu, L. (2022). Indoor geomagnetic positioning using direction-aware multiscale recurrent neural networks. *IEEE Sensors Journal*, 23(3), 3321-3333.
- Song, J., Li, W., Duan, C., Wang, L., Fan, Y., & Zhu, X. (2024). An Optimization-Based Indoor-Outdoor Seamless Positioning Method Integrating GNSS RTK, PS, and VIO. *IEEE Transactions on Circuits and Systems II: Express Briefs*.
- Sun, K., Sun, L., & Chen, T. (2024). Indoor Positioning Model Based on Principal Component Extraction of Support Vector Regression. 2024 36th Chinese Control and Decision Conference (CCDC),
- Tang, C., Sun, W., Zhang, X., Zheng, J., Wu, W., & Sun, J. (2023). A novel fingerprint positioning method applying vision-based definition for wifi-based localization. *IEEE Sensors Journal*, 23(14), 16092-16106.
- Tinh, P. D., Bui, H. H., & Nguyen, D. C. (2023). A genetic based indoor positioning algorithm using Wi-Fi received signal strength and motion data. *IAES International Journal of Artificial Intelligence*, 12(1), 328.
- Wang, B., Gan, X., Liu, X., Yu, B., Jia, R., Huang, L., & Jia, H. (2020). A novel weighted KNN algorithm based on RSS similarity and position distance for Wi-Fi fingerprint positioning. *IEEE Access*, 8, 30591-30602.
- Wang, X., Zhu, C., Tian, Y., Wang, X., Lin, Y., & Li, W. (2023). Optimization and performance analysis of Wi-Fi indoor fingerprint localization algorithm. 2023 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC),
- Wei, Y., & Zheng, R. (2021). Efficient Wi-Fi fingerprint crowdsourcing for indoor localization. *IEEE Sensors Journal*, 22(6), 5055-5062.
- Xu, J., Li, Z., Zhang, K., Yang, J., Gao, N., Zhang, Z., & Meng, Z. (2023). The principle, methods and recent progress in RFID positioning techniques: A review. *IEEE Journal of Radio Frequency Identification*, 7, 50-63.
- Yen, L., Yan, C.-H., Renu, S., Belay, A., Lin, H.-P., & Ye, Y.-S. (2017). A modified WKNN indoor Wi-Fi localization method with differential coordinates. 2017 International Conference on Applied System Innovation (ICASI),
- Zekavat, S., Buehrer, R. M., Durgin, G. D., Lovisolo, L., Wang, Z., Goh, S. T., & Ghasemi, A. (2021). An overview on position location: Past, present, future. *International Journal of Wireless Information Networks*, 28, 45-76.
- Zhou, R., Yang, Y., & Chen, P. (2021). An RSS transform – Based WKNN for indoor positioning. *Sensors*, 21(17), 5685.
- Zhuang, Y., Zhang, C., Huai, J., Li, Y., Chen, L., & Chen, R. (2022). Bluetooth localization technology: Principles, applications, and future trends. *IEEE Internet of Things Journal*, 9(23), 23506-23524.