

Article

Indoor Localization Algorithm Based on Information Gain Ratio and Affinity Propagation Clustering

Rencheng Jin , Di Zhang , Xiao Tian and Jianping Ma

Key Laboratory for Micro/Nano Technology and System of Liaoning Province, Dalian University of Technology, Dalian 116024, China; zd_dlut@mail.dlut.edu.cn (D.Z.); xiaot@mail.dlut.edu.cn (X.T.); majianping6890@mail.dlut.edu.cn (J.M.)

* Correspondence: rcjin@dlut.edu.cn

Abstract

In indoor positioning systems, it is common to use existing AP deployments within buildings to build a fingerprint database, providing positioning information during the online phase. However, AP layouts inside buildings often contain a large number of redundant APs, which leads to the improvement in positioning accuracy leveling off as the number of redundant APs increases, while also increasing the computational load of indoor positioning services. To address this problem, the thesis proposes a method for calculating the AP location discrimination capability and combines the location discrimination capability with coverage to eliminate redundant APs. Experiments conducted in real indoor scenarios, as well as on the Crowdsourced dataset and the SODIndoorLoc dataset, validate the results. The results show that the redundant AP removing strategy ensures that the average positioning accuracy fluctuates by no more than 5% compared to the unfiltered case, while significantly reducing the number of APs in the fingerprint database—by 64.43%, 72.78%, and 59.62%, respectively. In the position estimation phase, this paper uses affinity propagation clustering for coarse positioning and combines Bayesian methods for fine positioning. Compared with GMM, K-Means, and the pointwise algorithm, the average positioning error of the proposed method is reduced by 11% to 39%.

Keywords: indoor positioning; redundant AP elimination; affinity propagation clustering; location resolution

1. Introduction

Global Positioning System (GPS) signals are often severely obstructed and reflected by the structural walls of indoor buildings, leading to significant signal attenuation or even unstable reception, making it difficult to provide reliable location services [1]. As a result, indoor localization technologies have emerged as a prominent area of research. Wi-Fi, with its broad coverage, established infrastructure, and low additional hardware modification costs, offers distinct advantages [2]. The existing access points (APs) in indoor environments can directly serve as positioning infrastructure, which is why Wi-Fi signals are widely adopted in practical applications.

Wi-Fi indoor localization technology consists of two major classifications: the log-distance model (LDPLM) indoor localization technology and fingerprint-based indoor localization technology. LDPLM indoor localization technology estimates the line-of-sight (LoS) propagation distance from the Target Positions (TPs) to APs based on the LDPLM. In non-line-of-sight propagation paths, there are significant multipath effects



Academic Editor: Zhiyong Wang

Received: 12 December 2025

Revised: 10 January 2026

Accepted: 16 January 2026

Published: 19 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and shadow effects, causing discrepancies between the calculated and actual propagation distances, which weaken the stability of the propagation model and reduce positioning accuracy [3]. Fingerprinting is currently the most generally employed method in Wi-Fi indoor localization [4]. Fingerprinting comprises two stages: offline and online. Researchers gather Wi-Fi RSSIs and other signal features at the reference points (RPs) along with their actual coordinates during the offline stage. The data undergoes preprocessing, including filtering, feature extraction, clustering analysis, and the removal of redundant APs, resulting in the construction of a fingerprint database. During the online stage, researchers compare instantaneous RSSI features with the fingerprint database to infer the true position of TPs [5]. Unlike methods that rely on idealized propagation models, fingerprinting establishes an empirical relationship between “signal features and spatial locations,” making it more robust in complex indoor environments characterized by non-line-of-sight propagation and shadowing effects [6].

2. Related Work

At the 19th IEEE Conference on Computer Communications (INFOCOM), Bahl et al., from Microsoft Research [7], first proposed the RADAR system. The system constructs a radio map inside the building based on reference points (RPs), and, during the online stage, indoor localization is achieved by matching the instantaneous RSSI vector with fingerprints in the fingerprint database.

Mainstream research efforts aimed at improving the quality of fingerprint databases, localization accuracy, and system stability encompass various strategies, including the optimization of fingerprint data, refinement of fingerprint features, and enhancement of the efficiency in fingerprint database construction. In terms of optimizing fingerprint data and feature processing, Fang et al. [8] proposed utilizing Principal Component Analysis (PCA) to reduce the scale of RSSI data, replacing the original high-dimensional data with a few principal components. Lee and Han [9] employed nonlinear autoencoder technology to reduce the dimensionality of fingerprint map data, extracting more discriminative low-dimensional features. Luo and Fu [10] proposed using Kernel Principal Component Analysis (KPCA) for nonlinear mapping of RSSI features to eliminate data redundancy in the fingerprint database while retaining useful location information.

In terms of AP selection and improving the efficiency of fingerprint database construction, Cheng et al. [11] used Support Vector Regression (SVR) to simulate unsampled or missing fingerprint data, enhancing the system’s noise resistance to environmental changes such as occlusions or human traffic. Youssef et al. [12] proposed the MaxMean method, selecting several APs with the highest RSSI at the target location to form a cluster classification mechanism based on AP selection. Pan et al. [13] combined Building Information Modeling (BIM) with a multi-layer, multi-wall signal propagation model to generate virtual fingerprints, constructing an offline fingerprint database and achieving reliable positioning accuracy. Li et al. [14] mapped Channel State Information (CSI) to a two-dimensional amplitude feature map and used amplitude-feature deep convolutional generative adversarial networks to extend the fingerprint database, significantly reducing the burden of manual data collection. Reyes et al. [15] proposed an innovative deep neural model to identify long-term time-varying data. The proposed model uses only a small amount of fingerprint information instead of the complete fingerprint database. Xiao et al. [16] combined a region-based IPS with a noise map to track the movement trajectory of objects.

These methods minimize the overheads associated with fingerprint database construction and improve feature discrimination capability; however, they do not resolve the issue of varying location resolutions across different APs. The paper proposes an algorithm for eliminating redundant APs based on location resolution evaluation, which converts

the original signal strength information into a grid $Grid_{r \times c}$ and calculates the information gain ratio of each signal access point (AP). The algorithm uses information gain ratio and location coverage as ability indicators to evaluate the geographical location resolution of APs, significantly reducing the quantity of APs stored in the fingerprint dataset while ensuring positioning accuracy. In the location evaluation phase, combining affinity propagation clustering and the Bayesian algorithm, the positioning accuracy improves to varying degrees compared to KNN, the group matching method (GMM) [17], K-Means, pointwise [18], and other algorithms.

3. Offline Phase

3.1. AP Optimization via Location Resolution

Excessive and varying-quality AP information in complex indoor positioning buildings introduces signal noise, interferes with the matching process, and reduces positioning stability. Additionally, high-dimensional fingerprint databases significantly increase the cost of storage and retrieval [19]. Thus, eliminating redundant APs to create a compact and high-quality fingerprint database is crucial for improving the accuracy and maintainability of indoor Wi-Fi localization systems. Based on these considerations, this paper proposes an algorithm for eliminating redundant APs based on a location resolution assessment. The algorithm treats the signals from each access point (AP) as features, using the information gain rate as the AP location resolution capability indicator to eliminate redundant APs with lower location resolution capabilities.

3.1.1. Definition of Location Resolution Ability

Assume that in a certain indoor positioning deployment, the amount of APs installed is n , and the detected signal strength is categorized into P clusters: $C_1, C_2, \dots, C_P$. The total number of RPs is m , and these RPs are clustered into V clusters: $U_1, U_2, \dots, U_V$ utilizing the K-Means algorithm based on their geographical locations. The K-Means algorithm groups geographically close RPs into the same cluster U_i . Within the same cluster, the difference between each RP is minimized, while the geographical differences between different clusters U_i and U_j are maximized.

The fingerprint data collected from each AP feature measurement at all reference measurement points (RPs) can be transformed into a 2D grid of size $row \times col$. The information gain ratio for each signal access point (AP) is then calculated. The calculation method is as follows:

1. Compute the Shannon entropy of the location area, denoted as $Entropy(U)$.

$$Entropy(U) = - \sum_{v=1}^V \frac{|U_v|}{|U|} \log_2 \frac{|U_v|}{|U|} \quad (1)$$

where $|U_v|$ represents the number of RPs in cluster U_v and $|U|$ represents the total number of RPs.

2. Calculate the conditional entropy $ConditionalEntropy(C|U)$ of each access point (AP) as a feature.

$$\begin{aligned} ConditionalEntropy(C|U) &= \sum_{v=1}^V \frac{|U_v|}{|U|} \left(- \sum_{p=1}^P \frac{|D_{v,p}|}{|U_v|} \log_2 \frac{|D_{v,p}|}{|U_v|} \right) \\ &= - \sum_{v=1}^V \sum_{p=1}^P \frac{|D_{v,p}|}{|U|} \log_2 \frac{|D_{v,p}|}{|U_v|} \end{aligned} \quad (2)$$

where $D_{v,p}$ represents the union of C_p and U_v , and $|D_{v,p}|$ represents the number of signal reference points (RPs) in $D_{v,p}$.

3. Calculate the information gain $InfoGain(C, U)$ of the AP as a feature.

$$\begin{aligned} InfoGain(C, U) &= Entropy(U) - ConditionalEntropy(C|U) \\ &= - \sum_{v=1}^V \frac{|U_v|}{|U|} \log_2 \frac{|U_v|}{|U|} + \sum_{v=1}^V \sum_{p=1}^P \frac{|D_{v,p}|}{|U|} \log_2 \frac{|D_{v,p}|}{|U_v|} \end{aligned} \quad (3)$$

4. Calculate the information gain ratio $InfoGainRatio(C, U)$ of each AP as a feature.

$$\begin{aligned} InfoGainRatio(C, U) &= \frac{InfoGain(C, U)}{SplitInfo(C, U)} \\ &= \frac{- \sum_{v=1}^V \frac{|U_v|}{|U|} \log_2 \frac{|U_v|}{|U|} + \sum_{v=1}^V \sum_{p=1}^P \frac{|D_{v,p}|}{|U|} \log_2 \frac{|D_{v,p}|}{|U_v|}}{- \sum_{p=1}^P \frac{|C_p|}{|C|} \log_2 \frac{|C_p|}{|C|}} \end{aligned} \quad (4)$$

where C_p represents the set of signal strength categories for a given AP as a feature, $|C_p|$ represents the number of location RPs in the set C_p , and $SplitInfo(C, U)$ represents the split information of the signal strength classification.

The calculation formula for split information $SplitInfo(C, U)$ is given by Equation (5):

$$SplitInfo(C, U) = - \sum_{p=1}^P \frac{|C_p|}{|C|} \log_2 \frac{|C_p|}{|C|} \quad (5)$$

3.1.2. Removing Redundant APs

In Section 3.1.1, the location resolution capabilities of each AP are first calculated and then ranked in descending order. The redundant APs, whose location resolution is in the bottom $\lambda\%$ and whose RP coverage is in the bottom $\gamma\%$, are filtered out. The remaining APs are then used in the subsequent location estimation phase to calculate the actual position of TPs, thereby improving the overall performance and efficiency of the indoor positioning system.

3.2. AP Optimization via Correlation Clustering

In the indoor positioning environment, numerous access points (APs) are deployed. For instance, the SODIndoorLoc dataset includes approximately 552 APs, with signal strength data collected for Spearman correlation analysis. The histogram of correlation statistics for each AP is presented in Figure 1, revealing that certain APs exhibit strong correlations. The higher the similarity between two APs, the stronger their interchangeability. Grouping APs with high similarity for AP optimization can reduce the feature dimension of the fingerprint database.

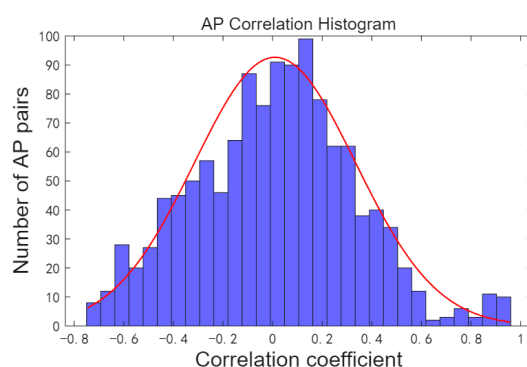


Figure 1. AP correlation histogram.

3.2.1. Region Partitioning Based on K-Means

As outlined in Section 3.1.1, calculating the information gain as an indicator of AP location resolution requires dividing the area into distinct location zones and forming a two-dimensional grid. The K-Means algorithm groups location reference points that are geographically close to the same cluster. In this study, the publicly available dataset published by Lohan et al. [20] is used to demonstrate the K-Means clustering results. The dataset was collected from a four-floor university building in Tampere, Finland, between January and August 2017, and includes 687 training fingerprints and 3951 test fingerprints. Figure 2 shows the results of location reference point clustering using the K-Means method for the entire location estimation dataset, while Figure 3 displays the regional division for each floor (first to fourth) of the building. As illustrated in Figures 2 and 3, the K-Means algorithm effectively clusters the location reference points (RPs) in the dataset based on their geographical locations, which is then applied in Section 3.2.3 for the calculation of location resolution.

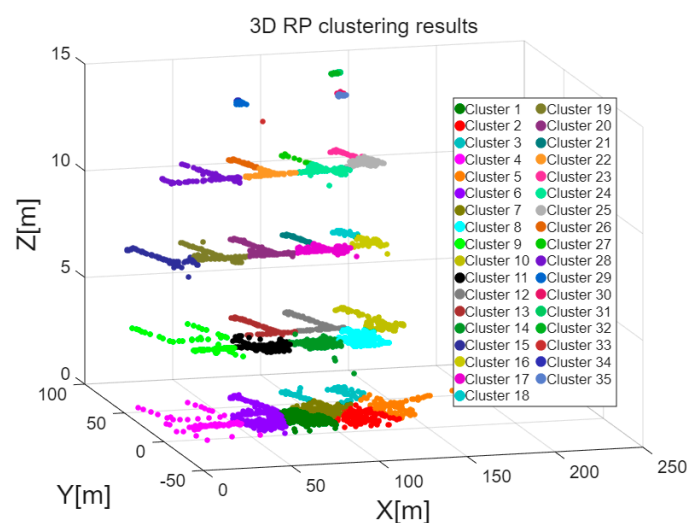


Figure 2. Three-dimensional RP clustering results.

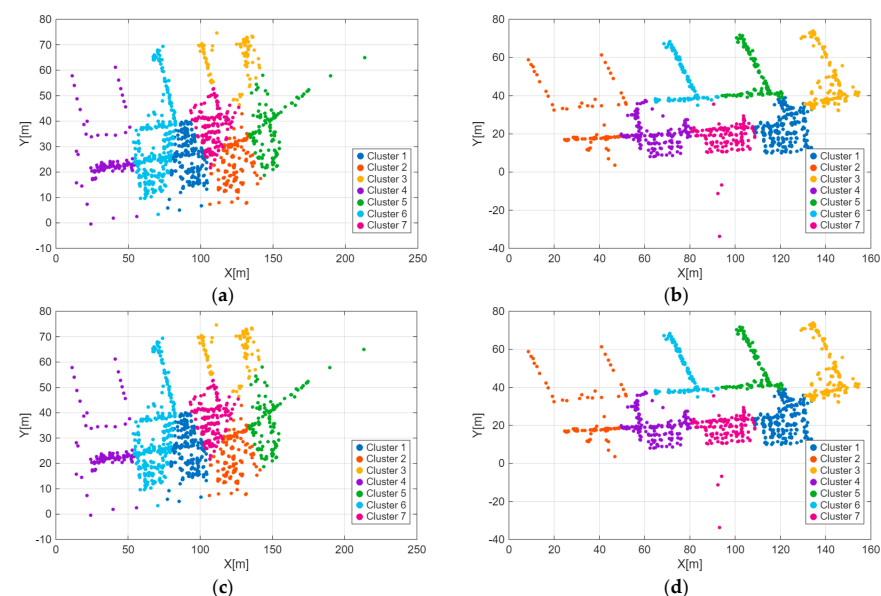


Figure 3. Clustering results of RPs on different floors. (a) Clustering results of RPs on the 1st floor. (b) Clustering results of RPs on the 2nd floor. (c) Clustering results of RPs on the 3rd floor. (d) Clustering results of RPs on the 4th floor.

3.2.2. AP Correlation Clustering

Assume that the signal strength vector $V_p = (rssi_1^p, rssi_2^p, rssi_3^p, \dots, rssi_m^p)$ of the signal access point AP_p consists of the signal strength measurements at the RP, and the signal strength vector $V_q = (rssi_1^q, rssi_2^q, rssi_3^q, \dots, rssi_m^q)$ of the signal access point AP_q consists of the signal strength measurements at each RP. After normalizing the signal vectors V_p and V_q , the maximum mutual information between AP_p and AP_q is calculated to obtain their similarity. The calculation steps are as follows:

1. Collecting Wi-Fi Fingerprint Signals

For any two signal access points, AP_p and AP_q , a signal sample set $D = \{(rssi_p^i, rssi_q^i) | i = 1, 2, \dots, N\}$ is constructed, $D = \{(rssi_p^i, rssi_q^i) | i = 1, 2, \dots, N\}$, $V_p = (rssi_1^p, rssi_2^p, rssi_3^p, \dots, rssi_m^p)$, and $V_q = (rssi_1^q, rssi_2^q, rssi_3^q, \dots, rssi_m^q)$, where N denotes the quantity of samples.

2. Computation of Maximum Mutual Information

The signals of V_p and V_q are converted into two-dimensional scatter plots, which are then divided into $r \times c$ smaller regions, satisfying the constraint Formula (6).

$$r \times c < B(N) \quad (6)$$

where $B(N)$ represents the threshold for the number of region divisions, typically set as $B(N) = N^{0.6}$ [21] and N represents the total data volume.

Calculate the mutual information $I(AP_p; AP_q | G)$ for each region division scheme. Different partitioning schemes exist for the same grid size. Mutual information is normalized based on the grid partitioning scale, and the maximum mutual information value is selected as the MIC. The calculation formula is as follows:

$$I(AP_p; AP_q) = \max_{r \times c < B(N)} \left\{ \frac{I(AP_p; AP_q | G)_{r,c}}{\log_2 \min\{r, c\}} \right\} \quad (7)$$

where $I(AP_p; AP_q | G)_{r,c}$ represents normalized mutual information values for different grid partitioning scales.

3. Signal Access Point Clustering

The maximum mutual coefficient for all pairs of signal access points AP is used to form a fuzzy relation matrix M . The fuzzy equivalence matrix M^* and its λ -cut matrix M_λ^* are then computed. M_λ^* satisfies Equations (8) and (9).

$$M_\lambda^* = \begin{bmatrix} \tilde{m}_{11}^* & \tilde{m}_{12}^* & \cdots & \tilde{m}_{1n}^* \\ \tilde{m}_{21}^* & \tilde{m}_{22}^* & \cdots & \tilde{m}_{2n}^* \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{m}_{n1}^* & \tilde{m}_{n2}^* & \cdots & \tilde{m}_{nn}^* \end{bmatrix} \quad (8)$$

$$\tilde{m}_{ij}^* = \begin{cases} 1, \tilde{m}_{11} \geq \beta \\ 0, \tilde{m}_{11} < \beta \end{cases} \quad (9)$$

Clustering of signal access points (APs) is performed based on the λ -cut matrix. The signal access points are assigned to groups $S_1, S_2, \dots, S_K$.

3.2.3. Removing Redundant APs

In the same clustering group S_k , the signal access points (APs) exhibit high similarity and can be interchangeable. Therefore, a signal access point with higher position resolution is selected within S_k to construct the fingerprint features. In this section, the position resolution metric from Section 3.1.1 is also used as the evaluation criterion for resolution among APs within the same cluster. The AP with the highest position resolution is selected from any similar AP cluster S_k to represent S_k and participate in constructing the fingerprint database.

4. Online Phase

Common clustering methods used in indoor positioning systems include K-Means clustering [22,23], variants of K-Means clustering [24–26], and fuzzy clustering [27]. The affinity propagation clustering method, proposed by Frey and Dueck [28], distinguishes itself from other clustering methods by not requiring the pre-selection of initial cluster points or the number of clusters. In this paper, during the position evaluation phase, affinity propagation clustering is applied to cluster the location reference points (RPs), forming clusters $S_1, S_2, \dots, S_K$, which complete the coarse positioning. The top N most similar clusters, $S_k, S_{k+1}, \dots, S_{k+n}$, are selected as the clusters used in the fine positioning phase. Gaussian probability estimation is performed on the fingerprint dataset contained within the clusters, and Bayesian position estimation is applied based on the Gaussian probabilities to determine the final location of the point to be located. The computational complexity of the core algorithm for position prediction is $O(M \times (C + N))$, where M represents the number of test samples, N represents the number of training samples, and C represents the number of clusters. To present this process more clearly, Algorithm 1 shows the pseudocode for the key parts of the position estimation phase, which facilitates a better understanding of the implementation details of the method proposed in this paper.

4.1. Affinity Propagation-Based Coarse Positioning

Affinity propagation clustering treats each sample data point as a potential cluster center candidate. A similarity matrix $Q_{n \times n}$ is constructed between the sample data points, where the similarity $q_{i,j}$ between each pair of data points is determined by the logarithmic Gaussian distance $L_{i,j}$. The formula for calculating $L_{i,j}$ is as follows:

$$L_{i,j} = \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(rssi_i - rssi_j)^2}{2\sigma^2} \right) \right) \quad (10)$$

where σ^2 represents the shadow variance of Wi-Fi signal strength and $rssi_i$ represents the RSSI vector of the i-th RP.

In affinity propagation clustering, responsibility $r_{i,j}$ and availability $a_{i,j}$ are used to measure the suitability of a data point as a cluster representative for other data points. These values are iteratively updated until convergence is reached. For the i-th sample data point, the values of $r_{i,j}$ and $a_{i,j}$ are exchanged to maximize $r_{i,j} + a_{i,j}$. A sample data point j is selected as the cluster center if it satisfies the condition $r_{i,j} + a_{i,j} > 0$.

4.2. Bayesian-Based Fine Positioning

Using affinity propagation clustering, all location reference points (RPs) are divided into K clusters, and the cluster centers of each cluster are obtained. These cluster centers are then used to assess the top n best clusters $S_k, S_{k+1}, \dots, S_{k+n}$ in the coarse positioning stage. Based on the logarithmic Gaussian distance and prior conditions, Bayes' theorem is used to update the posterior distribution and estimate the localization of TPs.

Algorithm 1 Indoor Localization Prediction

Input: Training fingerprints $D = \{(r_i, y_i)\}_{i=1..m}$, AP set $A = \{a_1..a_n\}$
Output: Estimated position $\hat{y}$
// ----- Offline: AP discriminability evaluation and redundancy filtering -----

1: Cluster RP coordinates $\{y_i\}$ into V clusters via KMeans $\rightarrow C(y_i)$ // region partition
2: Compute region entropy $H(C)$ // entropy of location clusters
3: for each AP a in A do // treat each AP as a feature
4: Discretize RSSI values $\{r_i[a]\}$ into P bins $\rightarrow S_a$ // RSSI category set for AP a
5: Compute conditional entropy $H(C \mid a)$ using S_a
6: $IG(a) \leftarrow H(C) - H(C \mid a)$ // information gain
7: $SI(a) \leftarrow \text{SplitInfo}(S_a)$ // penalty term for many bins
8: $IGR(a) \leftarrow IG(a)/SI(a)$ // InfoGainRatio as discriminability
9: end for
10: Rank APs by $IGR(a)$ in descending order // higher = better location resolution
11: Keep top- K APs (or threshold-based) $\rightarrow A_{\text{sel}}$; discard the rest as redundant // build compact fingerprint DB
12: Build reduced training set D_{sel} using only A_{sel} // dimension-reduced fingerprint library
// ----- Online: Position evaluation (coarse + fine localization) -----
13: Compute similarity matrix $S(i,j)$ on D_{sel} using log-Gaussian distance // similarity for clustering
14: Run Affinity Propagation on $D_{\text{sel}} \rightarrow$ clusters G_k with exemplars e_k // coarse clusters, no K preset
15: For test r^* , compute $\text{sim}(r^*, e_k)$ for all exemplars, select top- N clusters // coarse localization
16: Candidate RPs $\Omega \leftarrow$ union of RPs in the selected clusters // restrict search space
17: for each RP i in Ω do
18: $p_i \leftarrow \text{GaussianLikelihood}(r^* \mid r_i, \sigma)$
// Bayes likelihood from RSSI gap
19: $\text{score}_i \leftarrow \log(p_i)$
// posterior proxy score
20: end for
21: Select top- M RPs by $\text{score}_i \rightarrow \Omega_M$
// best-matching reference points
22: $\hat{y} \leftarrow \text{weighted_average}(\{y_i \mid i \in \Omega_M\}, \text{weights} = \text{score}_i)$
// final position estimation
23: return $\hat{y}$

In the Bayesian position estimation phase, the training set is defined as the RPs and the likelihood distribution of the RPs. The difference vector between the position of the TPs and RPs in the training set is used to calculate the Gaussian PDF. The matching probability $P(rssi_{TP}|loc_{RP})$ of the RSSI vectors between the TPs and the RPs is then calculated, as shown in Formula (11).

$$P(rssi_{TP}|loc_{RP}) = \prod_m \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(rssi_{TP,m} - rssi_{RP,m})^2}{2\sigma^2}\right) \quad (11)$$

where rss_{iTP} represents the RSSI vector detected at the TP, $rss_{iRP,m}$ represents the RSSI vector detected at the RP, $rss_{iTP,m}$ represents the RSSI detected at the TP from AP_m , loc_{RP} represents the position vector at the location reference point (RP), and σ^2 represents the shadow variance of the RSSI.

The logarithm of the matching probability $P(rss_{iTP}|loc_{RP})$ between the TP and the RP is taken as the proxy posterior score for the position of the RP. The top N RPs with the highest scores are selected as candidate RPs. The final position of the TP is calculated by performing a weighted average of the localization of the candidate RPs.

5. Model Estimation and Experimental Results

This section validates the superiority of the model proposed in the paper compared to other indoor positioning methods introduced in recent years. Experiments were conducted in a long corridor on the sixth floor of the Zhifang Building at Dalian University of Technology to measure positioning errors. To facilitate comparison with other positioning methods proposed in recent years, the performance of the model was also evaluated using the Tempare and SODIndoorLoc datasets.

The Zhifang Building at Dalian University of Technology, located at longitude 121.52 and latitude 38.89, has a total of nine floors, with each floor's architectural layout being quite similar. This study conducted the positioning effect validation experiment on the sixth floor of the Zhifang Building, and a 2D floor plan of the sixth floor is shown in Figure 4. Some experimental scenes are shown in Figure 5. The spacing between each RP is approximately 0.6 m.

The Crowdsourced dataset [20] is an indoor fingerprint dataset collected via crowdsourcing from January to August 2017 within a four-story building at Tampere University of Technology in Finland. The building covers an area of approximately 22,570 square meters. The dataset consists of two subsets: the training set, which contains about 687 fingerprint samples, and the testing set, which contains about 3951 fingerprint samples. The quantity of APs is approximately 992.

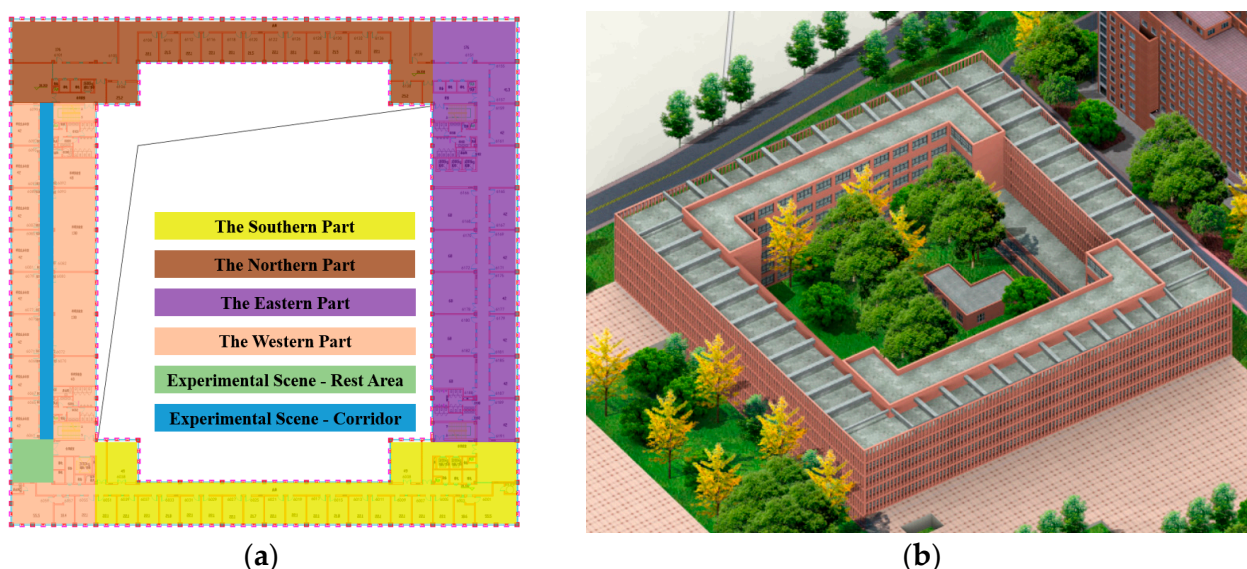


Figure 4. Environmental layout of the sixth floor of the Zhifang Building. (a) Interior layout of the Zhifang building. (b) Exterior shape of the Zhifang building.

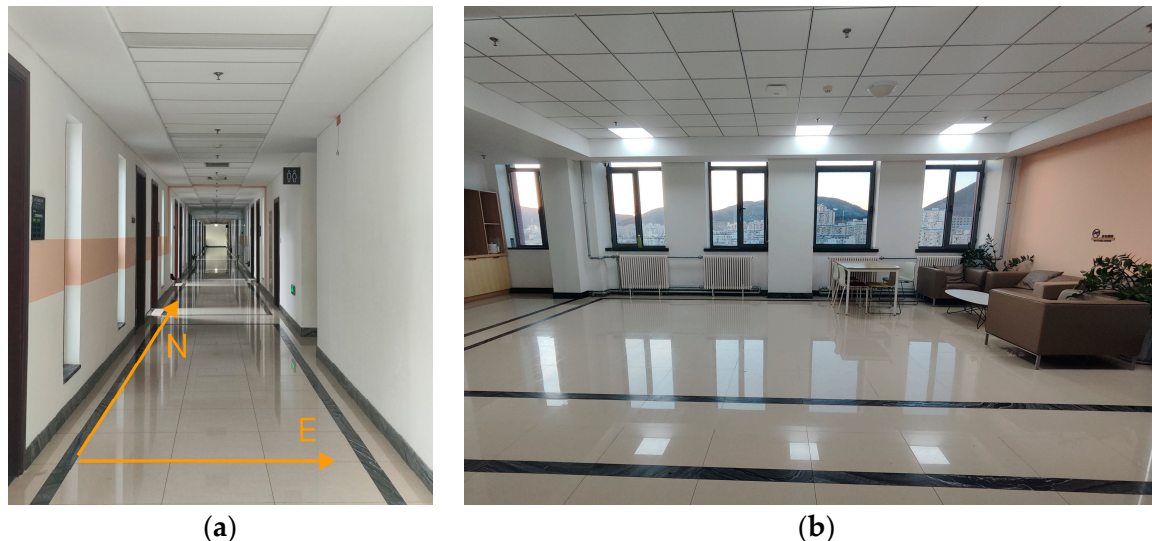


Figure 5. Environmental scenarios. (a) Long corridor experimental scene. (b) Rest area experimental scene.

The SODIndoorLoc dataset [29] is an extension of the classic UJIIndoorLoc dataset, covering three buildings: CETC331, HCXY, and SYL, with a total covered area of approximately 8000 square meters. The indoor positioning scenarios include various building environments, such as offices, corridors, and meeting rooms. The dataset contains about 23,925 fingerprint samples, with the training set consisting of approximately 21,205 samples and the testing set containing about 2720 samples. The number of signal access points is around 762. The spacing between each RP in the building numbered SYL and the building numbered HCXY is approximately 1.2 m, while the spacing between each RP in the CETC building is approximately 0.5 m.

The experimental scenario involves factors such as movement of individuals and door openings and closings, as well as obstacles like sofas, tables, chairs, electrical appliances, and walls. The fingerprint database for the experiment is collected in this environment, where algorithm validation is performed, meeting the dynamic and complex conditions required for indoor positioning. The Crowdsourced dataset constructs a fingerprint database for a multi-story building, utilizing crowdsourcing to collect fingerprint data from 21 different mobile devices and users. The data collection period for the Crowdsourced dataset spans approximately 7 months, fulfilling the temporal dynamics requirement for indoor positioning algorithm validation. The SODIndoorLoc dataset covers various indoor environments, including offices, meeting rooms, and corridors, with around 20,000 data samples. The indoor environments contain diverse obstacles, meeting the complexity requirements for indoor positioning algorithm validation.

5.1. Comparison of AP Optimization Effects

In Section 3, two AP optimization schemes were proposed: Scheme 1—AP Redundancy Optimization Algorithm Based on Position Resolution Evaluation, and Scheme 2—AP Optimization Algorithm Based on Correlation Grouping and Position Resolution Filtering. This section compares the positioning results of these two schemes in terms of AP reduction ratio, positioning error distribution, average positioning error, and other performance metrics.

In Scheme 1, after filtering the signal access points (APs) based on position coverage of $\lambda\%$ and position resolution of $\gamma\%$, the MSE of indoor localization on the fingerprint database is shown in Figure 6. Figures 6–8 illustrate that as λ and γ decrease, the average positioning error shows a decreasing trend, with the rate of decrease gradually slowing

down. The more APs there are, the more available information the fingerprint database provides for positioning services, leading to a gradual reduction in positioning errors. However, once the quantity of APs reaches a certain threshold, additional APs result in only marginal gains in positioning accuracy.

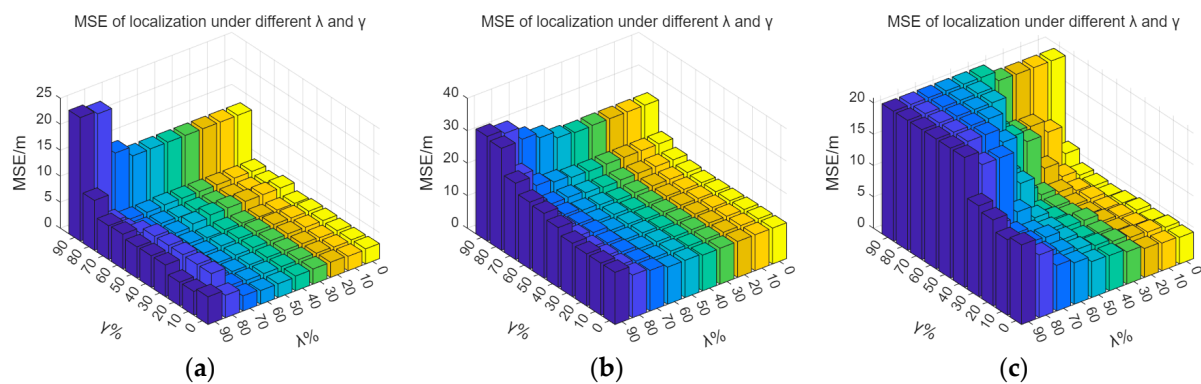


Figure 6. MSE of localization under different λ and γ . (a) Average localization error in experimental scenario under different λ and γ . (b) Average localization error in the Crowdsourced dataset under different λ and γ . (c) Average localization error in SODIndoorLoc dataset under different λ and γ .

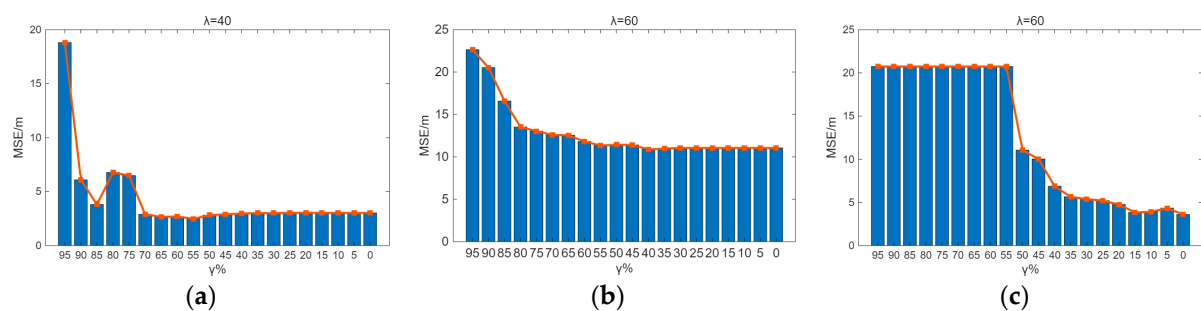


Figure 7. MSE of localization under different γ with fixed λ . (a) Average localization error in experimental scenario under different γ with fixed optimal λ . (b) Average localization error in Crowdsourced dataset under different γ with fixed optimal λ . (c) Average localization error in SODIndoorLoc dataset under different γ with fixed optimal λ .

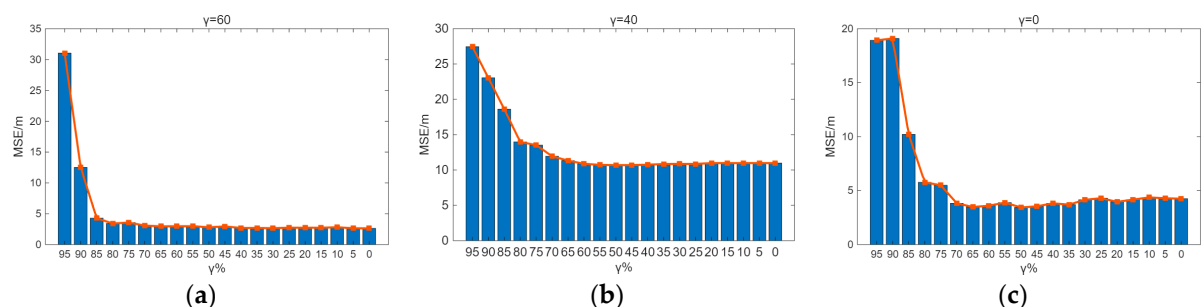


Figure 8. MSE of localization under different λ with fixed γ . (a) Average localization error in experimental scenario under different λ with fixed optimal γ . (b) Average localization error in Crowdsourced dataset under different λ with fixed optimal γ . (c) Average localization error in SODIndoorLoc dataset under different λ with fixed optimal γ .

Sensitivity analysis was performed on the parameters λ and γ . From Figures 9 and 10, it can be observed that the indoor positioning mean square error gradually decreases as λ and γ decrease, with the rate of decrease slowing down. In the experimental scenario, the indoor positioning mean square error remains almost unchanged at $\gamma = 60\%$ and $\lambda = 50\%$.

In the Crowdsourced dataset and SODIndoorLoc dataset, the indoor positioning mean square error remains almost unchanged at $\gamma = 50\%$ and $\lambda = 30\%$, and $\gamma = 30\%$ and $\lambda = 30\%$, respectively. Figures 9 and 10 further validate that when the number of APs increases to a certain level, the improvement in positioning accuracy due to the increase in the number of APs becomes limited.

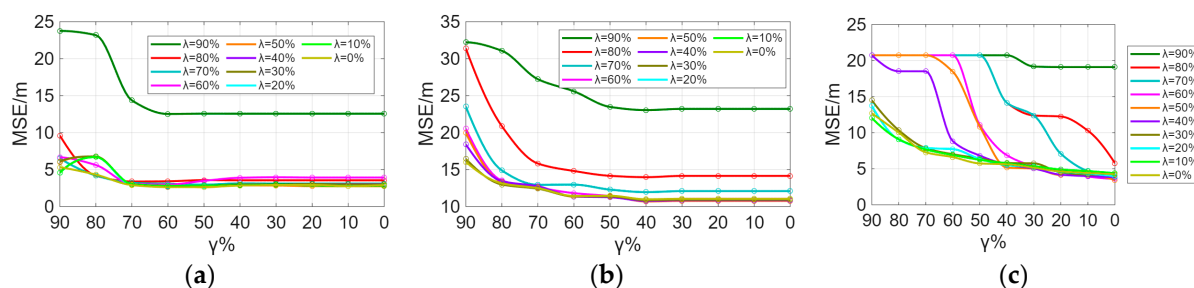


Figure 9. γ sensitivity analysis. (a) Sensitivity analysis of γ in the experimental scenario. (b) γ sensitivity analysis of the Crowdsourced dataset. (c) γ sensitivity analysis of the SODIndoorLoc dataset.

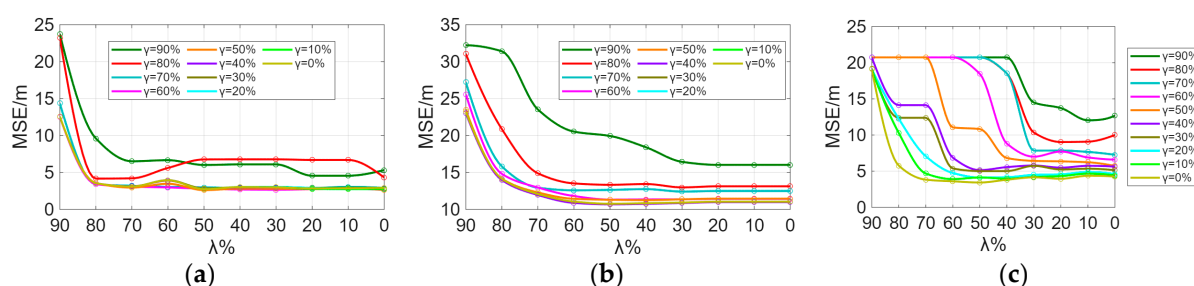


Figure 10. λ sensitivity analysis. (a) Sensitivity analysis of λ in the experimental scenario. (b) λ sensitivity analysis of the Crowdsourced dataset. (c) λ sensitivity analysis of the SODIndoorLoc dataset.

By traversing λ and γ , it is found that in the real indoor experimental scenario, the Crowdsourced dataset, and the SODIndoorLoc dataset, the minimum number of APs can be achieved with $\lambda = 40$, $\gamma = 60$; $\lambda = 60$, $\gamma = 40$; and $\lambda = 60$, $\gamma = 0$, respectively. Furthermore, the increase in the MSE of indoor localization compared to the original MSE of indoor localization does not exceed 5%. After the original fingerprint database undergoes AP redundancy processing and is combined with KNN for position estimation, the mean positioning errors for the experimental scenario, the Crowdsourced dataset, and the SODIndoorLoc dataset are 2.69 m, 10.88 m, and 3.60 m, respectively.

Scheme 2 groups APs with high similarity into the same cluster S_k and filters out APs with lower position resolution within S_k to construct the fingerprint database. The Spearman correlation statistics for the APs in the three indoor positioning fingerprint datasets are shown in Figure 11. The proportion of APs with strong correlation (correlation coefficient η satisfying $|\eta| > 0.8$) is relatively small.

The quantity of APs used and the MSE of indoor positioning in the positioning experiments conducted on the three datasets are illustrated in Table 1. The number of strongly correlated APs that can be excluded in Scheme 2 is far fewer than in Scheme 1. Scheme 1 can use fewer APs while maintaining similar positioning accuracy. Specifically, with Scheme 1, the number of APs used was reduced by 64.43%, 72.78%, and 59.62% for the three datasets, respectively, while Scheme 2 only reduced the number of APs by approximately 1.34%, 0.40%, and 61.54%.

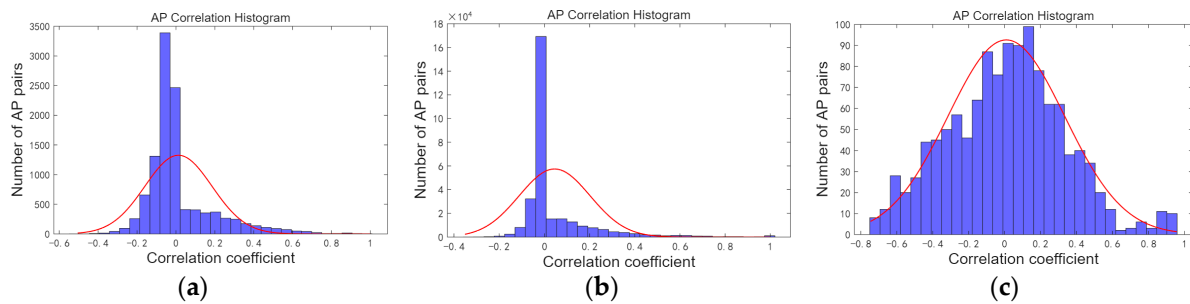


Figure 11. AP correlation histogram. (a) Correlation histogram of all AP signals in experimental scenario. (b) Correlation histogram of all AP signals in Tempare dataset. (c) Correlation histogram of all AP signals in SODIndoorLoc dataset.

Table 1. MSE and amount of APs using Schemes 1 and 2.

	Methods	Scheme	Error/m	Number of APs	Initial Number of APs
Experiment	KNN	Scheme 1	2.69	53	149
		Scheme 2	2.71	147	149
	WKNN	Scheme 1	2.90	53	149
		Scheme 2	2.79	147	149
	GMM	Scheme 1	2.73	53	149
		Scheme 2	2.33	147	149
Tempare	KNN	Scheme 1	10.88	270	992
		Scheme 2	11.04	988	992
	WKNN	Scheme 1	10.45	270	992
		Scheme 2	10.58	988	992
	GMM	Scheme 1	11.16	270	992
		Scheme 2	11.22	988	992
SODIndoorLoc	KNN	Scheme 1	3.60	21	52
		Scheme 2	3.63	20	52
	WKNN	Scheme 1	3.63	21	52
		Scheme 2	3.66	20	52
	GMM	Scheme 1	5.66	21	52
		Scheme 2	6.02	20	52

The heatmaps of the AP location resolution for the initial fingerprint database, Schemes 1 and 2 are shown in Figure 12. The closer the color is to warm tones, the higher the AP location resolution. As shown in Figure 12, the fingerprint database in Scheme 1 exhibits higher AP location resolution.

The error CDF images for the initial fingerprint database, Schemes 1 and 2 in the three experimental scenarios are shown in Figure 13. Both Schemes 1 and 2 can still ensure positioning accuracy similar to the initial fingerprint database, despite the reduced quantity of APs, and the increase in positioning error is kept within 5% compared to before processing. In the experimental scenario, the Crowdsourced dataset, and the SODIndoorLoc dataset, Scheme 1 slightly outperforms Scheme 2 in terms of positioning accuracy.

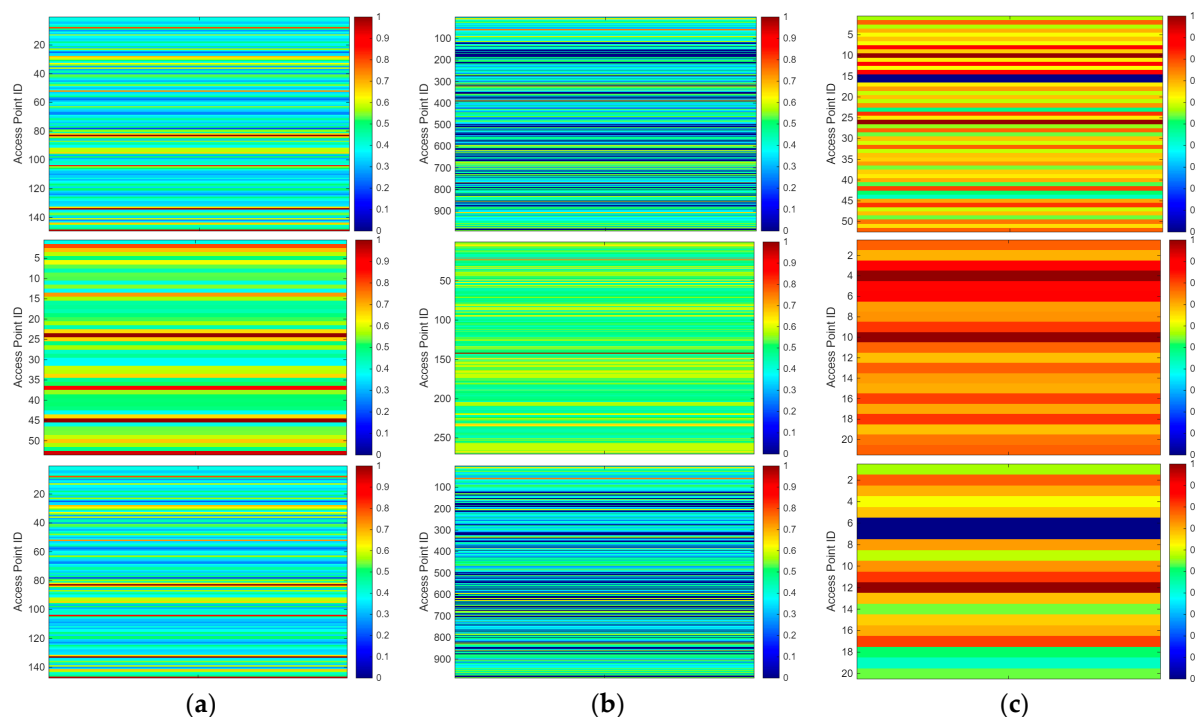


Figure 12. Heatmap of AP location resolution (initial, Schemes 1 and 2). (a) Raw fingerprint database signal heatmap. (b) Fingerprint database signal heatmap after processing with Scheme 1. (c) Fingerprint database signal heatmap after processing with Scheme 2.

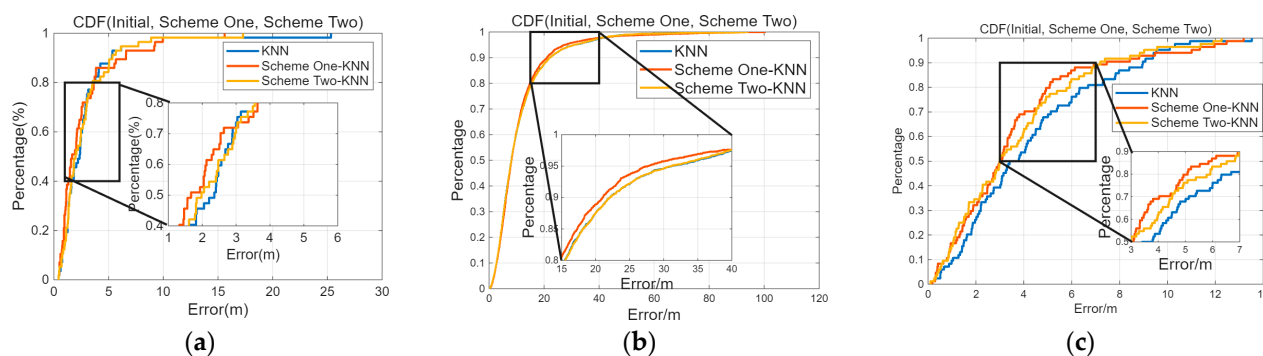
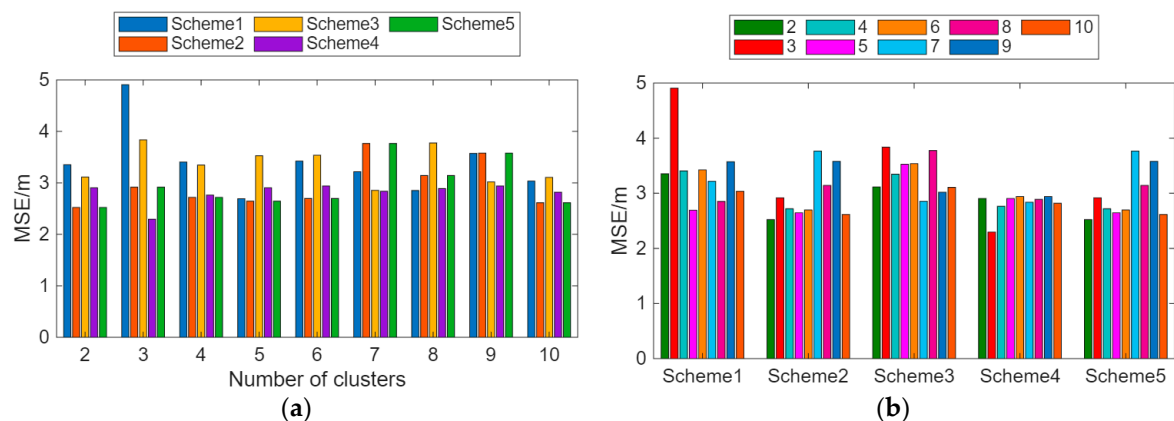


Figure 13. CDF (initial, Schemes 1 and 2). (a) Localization error CDF on the 6th floor of experimental scenario under different schemes. (b) Localization error CDF in the Crowdsourced dataset under different schemes. (c) Localization error CDF in the SODIndoorLoc dataset under different schemes.

Scheme 1 uses the K-Means clustering method to divide the RPs into V clusters and the RSSI values into P clusters. This paper also investigates the impact of five different RSSI strength partitioning schemes and the number of RP clusters on positioning accuracy. Table 2 presents the five different RSSI strength partitioning schemes, and Figure 14 displays the average positioning errors under different partitioning methods. From Figure 14a, it can be observed that Schemes 4 and 5 in the RSSI partitioning schemes have relatively small positioning errors. Figure 14b indicates that in the experimental scenario, clustering numbers between Schemes 4 and 6 can achieve higher positioning accuracy.

Table 2. Different RSSI strength partitioning schemes.

Scheme	Details					
	Excellent	Good	Fair	Poor	Very Poor	No Signal
Scheme 1	≥ -55	$[-77, -55)$	$[-88, -77)$	$[-100, -88)$	-	< -100
Scheme 2	≥ -50	$[-60, -50)$	$[-70, -60)$	$[-80, -70)$	-	< -80
Scheme 3	≥ -45	$[-55, -45)$	$[-65, -55)$	$[-75, -65)$	$[-85, -75)$	< -85
Scheme 4	-	≥ -60	$[-70, -60)$	< -70	-	-
Scheme 5	≥ -50	$[-60, -50)$	$[-70, -60)$	$[-80, -70)$	-	< -80

**Figure 14.** Average positioning error under different partitioning methods. (a) Average positioning error under different RSSI strength partitioning schemes. (b) Average positioning error for different numbers of clusters.

In summary, Scheme 1 can ensure that the increase in the average indoor positioning error does not exceed 5% while reducing the number of APs used by approximately 50%, demonstrating superior performance compared to Scheme 2. Therefore, Scheme 1 is chosen as the technical approach for fingerprint database processing in this study.

5.2. Comparison of Indoor Positioning Errors

The indoor positioning algorithm proposed in the paper identifies the position reference point clusters to which the TP belongs during the coarse localization stage. It then selects the top n most similar RP clusters $S_k, S_{k+1}, \dots, S_{k+n}$ to the TP. The results of affinity propagation clustering on the experimental scenario and the SODIndoorLoc-CETC331 building dataset are shown in Figure 13. It can be seen that affinity propagation clustering effectively clusters the location reference points (RPs) based on signal strength similarity. During the fine positioning stage, the RPs in clusters $S_k, S_{k+1}, \dots, S_{k+n}$ are used as candidate reference points for Bayesian position estimation.

Table 3 presents the floor identification accuracy of the proposed method on two public fingerprint datasets, achieving rates of 90.00% and 95.24%, respectively, which demonstrates high performance in floor determination.

Table 3. Correct floor identification rate for different positioning methods.

Dataset	Method	KNN	WKNN	GMM	K-Means	Pointwise	Proposed
Tempare		81.12%	81.83%	82.11%	82.59%	86.64%	90.00%
SODIndoorLoc		98.81%	98.81%	97.62%	98.81%	97.62%	95.24%

Table 4 presents the MSE of the algorithm proposed in the paper compared to other positioning methods. Figure 15 illustrates the distribution of positioning errors for different methods. Compared to the KNN, WKNN, GMM [17], K-Means clustering-Bayesian [30], and Pointwise Coverage Area [18] methods, the proposed method exhibits a smaller average positioning error. Figure 16 displays the CDF statistics of positioning errors for these methods on the experimental scenario, Crowdsourced dataset, and SODIndoorLoc-CETC331 building dataset. Due to differences in fingerprint collection density, the positioning error distributions vary across the datasets.

A large number of scholars [4,6,13,31,32] have studied the impact of RP collection density on indoor positioning accuracy. Shang and Wang [4] emphasized the importance of RP density for positioning accuracy, noting that an increased number of RPs allows for a more detailed signal distribution, but also increases collection costs and computational overheads. Ayub et al. [32] pointed out that high-density RP collection can improve the accuracy of indoor positioning service systems. The research on RP density by previous scholars is already quite comprehensive, so this paper does not further investigate the impact of RP collection density on indoor positioning errors.

Table 4. MSE of positioning for different methods. (Unit: meters.)

	KNN	WKNN	GMM	K-Means	Pointwise	Proposed
Experiment	2.90	2.90	2.32	2.68	2.73	2.07
Crowdsourced	11.05	10.57	11.22	11.56	10.04	8.79
SODIndoorLoc	4.25	4.26	5.36	3.77	4.30	3.27

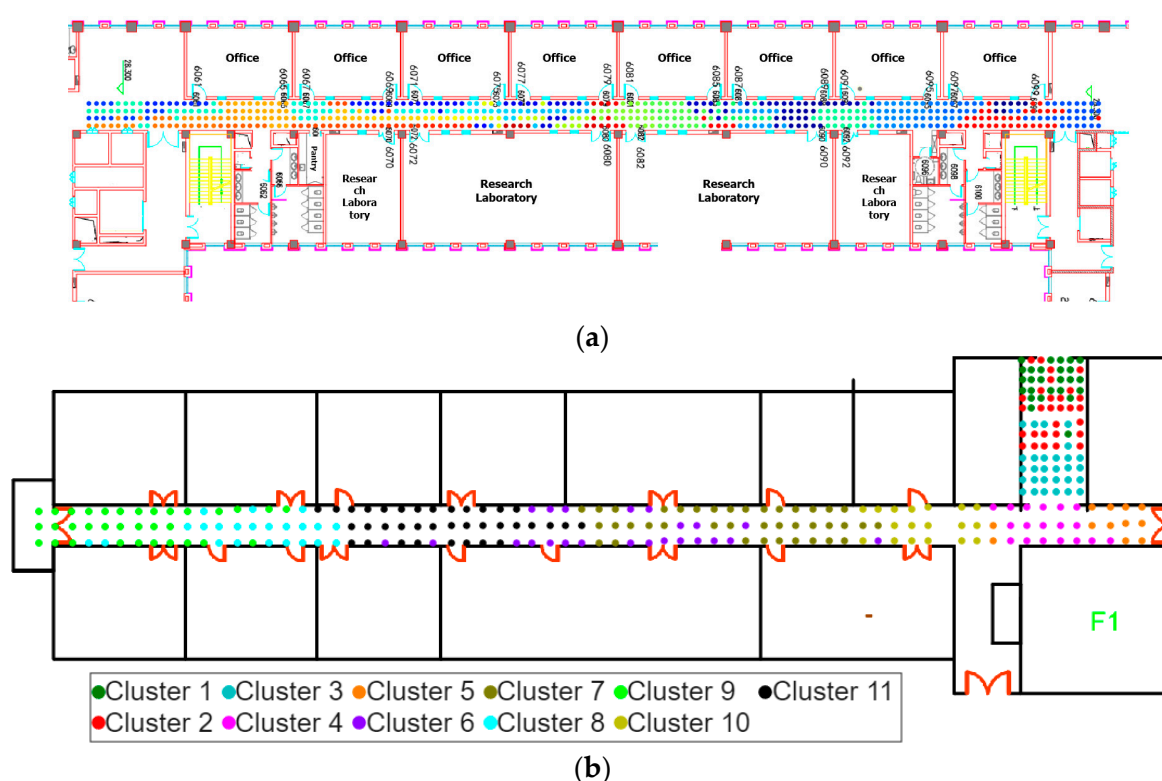


Figure 15. Cont.

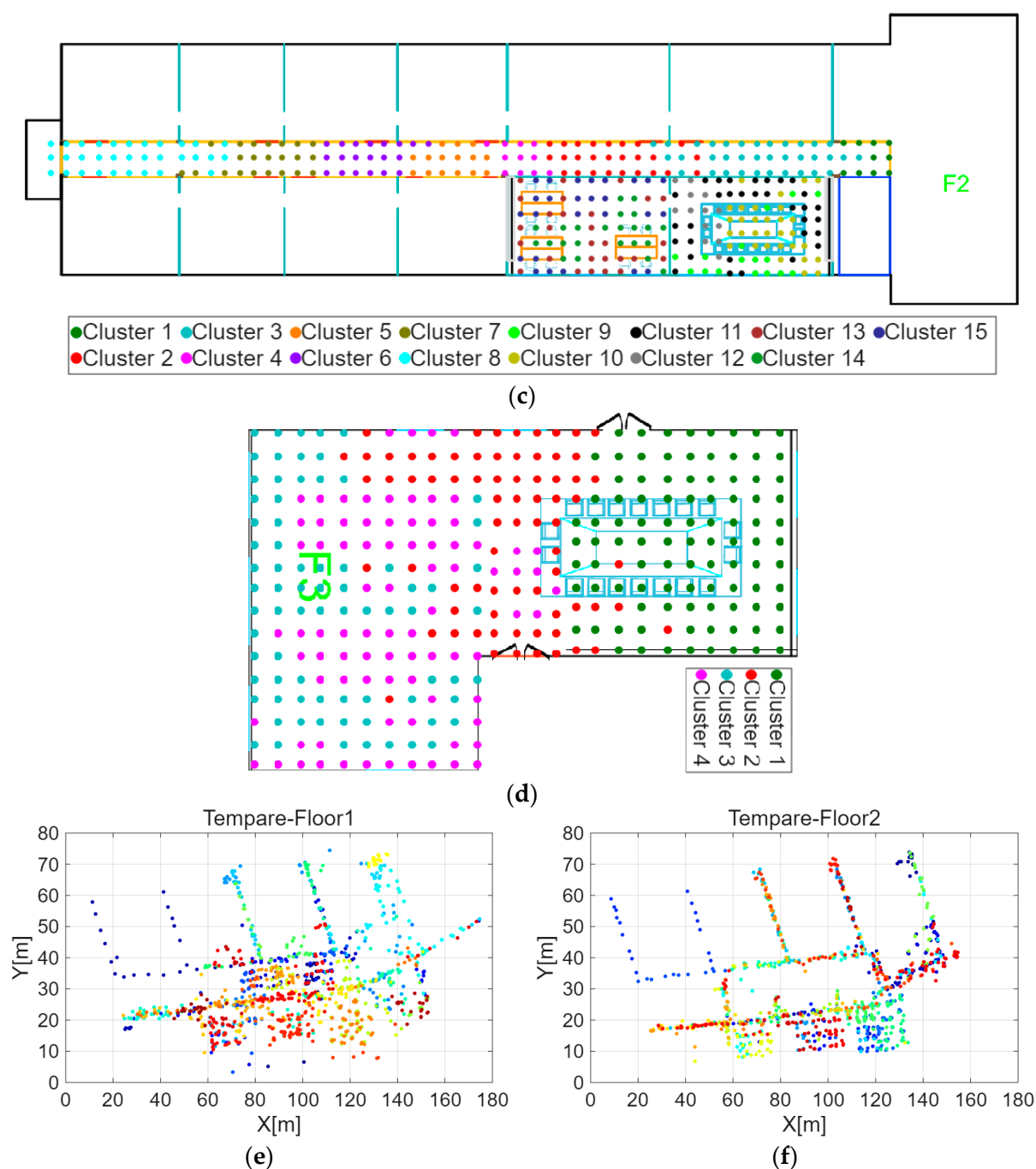


Figure 15. Affinity propagation clustering results. (a) RP clustering results on the experimental scenario. (b) RP clustering results on the 1st floor of CETC331 building. (c) RP clustering results on the 2nd floor of CETC331 building. (d) RP clustering results on the 3rd floor of CETC331 building. (e) RP clustering results on the 1st floor of Crowdsourced dataset. (f) RP clustering results on the 2nd floor of Crowdsourced dataset.

Using the proposed model, for the experimental scenario, 80% of the positioning errors are within 3.25 m, and 95% are within 5.44 m. For the Crowdsourced dataset, 80% of the positioning errors are within 12.46 m, and 95% are within 24.45 m. For the SODIndoorLoc dataset, 80% of the positioning errors are within 4.56 m, and 95% are within 10.19 m. The Crowdsourced dataset and SODIndoorLoc dataset are both multi-story building fingerprint database datasets. The method proposed in this paper achieves floor correct identification

rates of 90.00% and 95.24% for the Crowdsourced and SODIndoorLoc datasets, respectively, which basically meet the service requirements of indoor positioning systems.

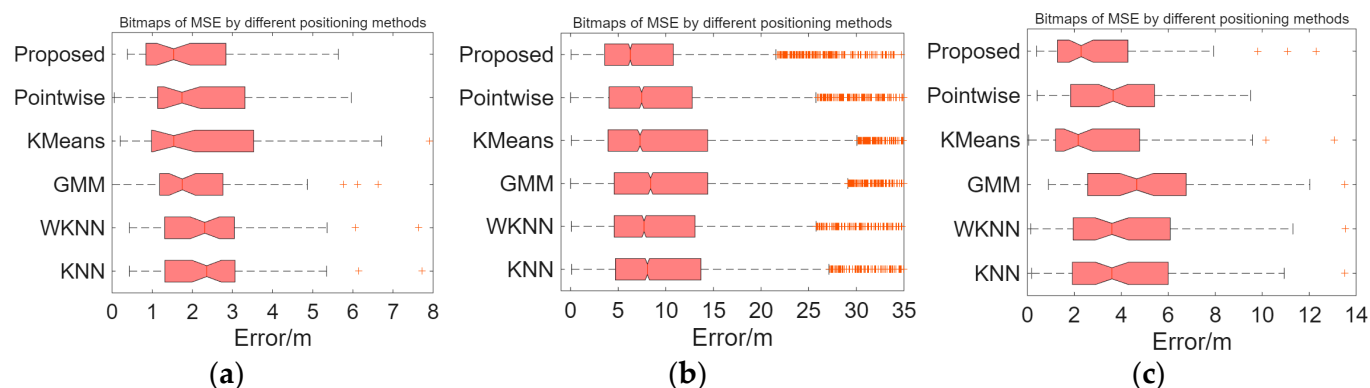


Figure 16. Boxplot of positioning errors for various indoor positioning methods. (a) Average localization error boxplot for the experimental scenario. (b) Average localization error boxplot for the Crowdsourced dataset. (c) Average localization error boxplot for the SODIndoorLoc dataset.

Figures 17 and 18 show the average positioning errors for different positioning models on the experimental scenario, Crowdsourced dataset, and SODIndoorLoc-CETC331 building dataset. The proposed positioning model demonstrates higher positioning accuracy. Compared to GMM positioning, the positioning errors for the experimental scenario, Crowdsourced dataset, and SODIndoorLoc-CETC331 building dataset are reduced by approximately 11.02%, 21.65%, and 39.10%, respectively. Compared to the K-Means clustering method, the positioning errors are reduced by 22.73%, 23.94%, and 13.31%, respectively. Compared to the pointwise method, the positioning errors are reduced by 24.09%, 12.41%, and 24.12%, respectively.

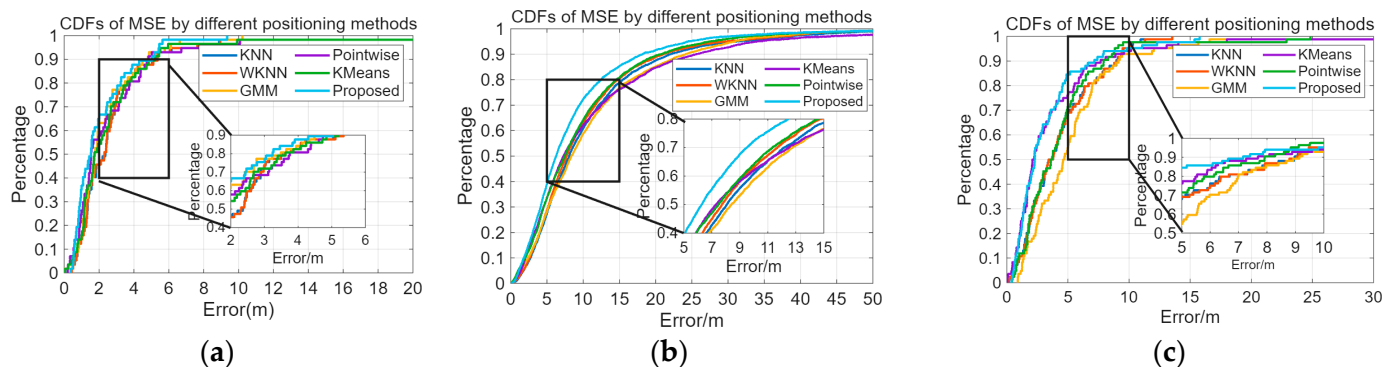


Figure 17. CDF statistics of positioning errors for various indoor positioning methods. (a) CDF of localization error for different localization methods on experimental scenario. (b) CDF of localization error for different localization methods in Crowdsourced dataset. (c) CDF of localization error for different localization methods in SODIndoorLoc dataset.

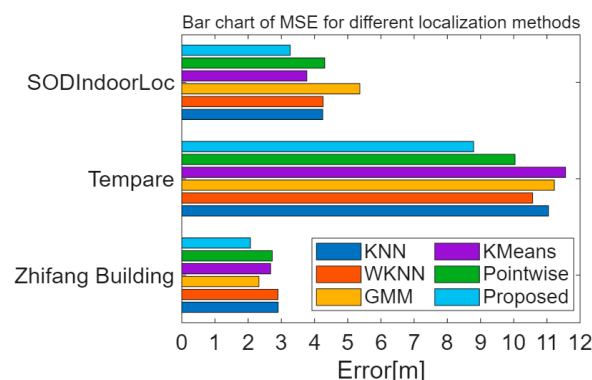


Figure 18. Bar chart of MSE for different localization methods.

6. Conclusions

The paper proposes a Wi-Fi fingerprint indoor positioning strategy based on the information gain ratio and affinity propagation clustering. A method for calculating the position resolution of APs is defined, and redundant APs are removed based on position resolution and position coverage. The effectiveness of the proposed method was validated through experiments conducted using the Tampere dataset, the SODIndoorLoc dataset, and a real-world experimental scenario.

The experimental results demonstrate that the proposed AP redundancy removal strategy effectively decreases the quantity of APs stored in the fingerprint dataset, while preserving positioning accuracy. The number of APs is reduced by 64.43%, 72.78%, and 59.62%, respectively. In the position estimation phase, coarse positioning is performed using affinity propagation clustering, and fine positioning is achieved through Bayesian methods.

Compared to GMM positioning, the positioning errors in the experimental scenario, Crowdsourced dataset, and SODIndoorLoc-CETC331 building dataset are reduced by approximately 11.02%, 21.65%, and 39.10%, respectively. When compared to the K-Means clustering method, positioning errors decrease by 22.73%, 23.94%, and 13.31%, respectively. Compared to the pointwise method, positioning errors are reduced by 24.09%, 12.41%, and 24.12%, respectively.

Author Contributions: Conceptualization, D.Z.; methodology, D.Z.; software, D.Z., X.T., and J.M.; validation, D.Z., X.T., J.M., and R.J.; formal analysis, X.T. and J.M.; investigation, X.T.; data curation, D.Z. and J.M.; writing—original draft preparation, D.Z.; writing—review and editing, R.J.; visualization, D.Z.; supervision, R.J.; project administration, R.J.; funding acquisition, R.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities (DUTZD25114).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to restrictions. The data presented in this study are available on request from the corresponding author due to the policies and confidentiality agreements followed by our laboratory. Additionally, we used data from the publicly available Zenodo dataset at <https://zenodo.org/records/1001662>, reference number 10.5281/zenodo.889798, and github at <https://doi.org/10.1186/s43020-022-00086-y>, reference number 10.1186/s43020-022-00086-y. These data were derived from the following resources available in the public domain: Crowdsourced dataset: <https://zenodo.org/records/1001662> (accessed on 8 November 2022). SODIndoorLoc dataset: <https://github.com/bijingxue/SODIndoorLoc> (accessed on 12 September 2017).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

APs	Access Points
RP	Reference Points
KNN	K-Nearest-Neighbor
WKNN	Weighted K-Nearest Neighbor
CDF	Cumulative Distribution Function
RSSI	Received Signal Strength Indication
MIC	Maximum Information Coefficient

References

1. Hailu, T.G.; Guo, X.; Si, H.; Li, L.; Zhang, Y. Theories and Methods for Indoor Positioning Systems: A Comparative Analysis, Challenges, and Prospective Measures. *Sensors* **2024**, *24*, 6876. [\[CrossRef\]](#)
2. Xie, T.; Jiang, H.; Zhao, X.; Zhang, C. A Wi-Fi-Based Wireless Indoor Position Sensing System with Multipath Interference Mitigation. *Sensors* **2019**, *19*, 3983. [\[CrossRef\]](#)
3. Yang, Z.; Zhou, Z.; Liu, Y. From RSSI to CSI: Indoor Localization via Channel Response. *ACM Comput. Surv.* **2013**, *46*, 25. [\[CrossRef\]](#)
4. Shang, S.; Wang, L. Overview of WiFi Fingerprinting-Based Indoor Positioning. *IET Commun.* **2022**, *16*, 725–733. [\[CrossRef\]](#)
5. Xia, S.; Liu, Y.; Yuan, G.; Zhu, M.; Wang, Z. Indoor Fingerprint Positioning Based on Wi-Fi: An Overview. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 135. [\[CrossRef\]](#)
6. Zholamanov, B.; Saymbetov, A.; Nurgaliyev, M.; Bolatbek, A.; Dosymbetova, G.; Kuttybay, N.; Orynbassar, S.; Kapparova, A.; Koshkarbay, N.; Beyca, Ö.F. RSSI Fingerprint-Based Indoor Localization Solutions Using Machine Learning Algorithms: A Comprehensive Review. *Smart Cities* **2025**, *8*, 153. [\[CrossRef\]](#)
7. Bahl, P.; Padmanabhan, V.N. RADAR: An in-Building RF-Based User Location and Tracking System. In Proceedings of the IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064), Tel Aviv, Israel, 26–30 March 2000; Volume 2, pp. 775–784.
8. Fang, S.-H.; Lin, T. Principal Component Localization in Indoor WLAN Environments. *IEEE Trans. Mob. Comput.* **2012**, *11*, 100–110. [\[CrossRef\]](#)
9. Lee, M.K.; Han, D.S. Dimensionality Reduction of Radio Map with Nonlinear Autoencoder. *Electron. Lett.* **2012**, *48*, 655–657. [\[CrossRef\]](#)
10. Luo, J.; Fu, L. A Smartphone Indoor Localization Algorithm Based on WLAN Location Fingerprinting with Feature Extraction and Clustering. *Sensors* **2017**, *17*, 1339. [\[CrossRef\]](#)
11. Cheng, Y.-K.; Chou, H.-J.; Chang, R.Y. Machine-Learning Indoor Localization with Access Point Selection and Signal Strength Reconstruction. In Proceedings of the 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, China, 15–18 May 2016; pp. 1–5.
12. Youssef, M.A.; Agrawala, A.; Udaya Shankar, A. WLAN Location Determination via Clustering and Probability Distributions. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003), Fort Worth, TX, USA, 26–26 March 2003; pp. 143–150.
13. Pan, H.; Huang, Z.; Zheng, W. Virtual Fingerprint Indoor Localization Based on Signal Propagation Model: RSS Prediction Using BIM. *J. Build. Eng.* **2025**, *113*, 113752. [\[CrossRef\]](#)
14. Li, Q.; Qu, H.; Liu, Z.; Zhou, N.; Sun, W.; Sigg, S.; Li, J. AF-DCGAN: Amplitude Feature Deep Convolutional GAN for Fingerprint Construction in Indoor Localization Systems. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *5*, 468–480. [\[CrossRef\]](#)
15. Reyes, J.M.R.; Ho, I.W.-H.; Mak, M.-W. Wi-Fi CSI Fingerprinting-Based Indoor Positioning Using Deep Learning and Vector Embedding for Temporal Stability. *Expert Syst. Appl.* **2025**, *264*, 125802. [\[CrossRef\]](#)
16. Xiao, L.; Ghafoorpoor Yazdi, P.; Thiede, S. A Zone-Based Wi-Fi Fingerprinting Indoor Positioning System for Factory Noise Mapping. *J. Intell. Manuf.* **2025**. [\[CrossRef\]](#)
17. Lin, H.; Purmehdi, H.; Fei, X.; Zhao, Y.; Isac, A.; Louafi, H.; Peng, W. Two-Stage Clustering for Improve Indoor Positioning Accuracy. *Autom. Constr.* **2023**, *154*, 104981. [\[CrossRef\]](#)
18. Koski, L.; Piche, R.; Kaseva, V.; Ali-Loytty, S.; Hannikainen, M. Positioning with Coverage Area Estimates Generated from Location Fingerprints. In Proceedings of the 2010 7th Workshop on Positioning, Navigation and Communication, Dresden, Germany, 11–12 March 2010; IEEE: Dresden, Germany, 2010; pp. 99–106.

19. Hailu, T.G.; Guo, X.; Si, H.; Li, L.; Zhang, Y. Ada-LT IP: Functional Discriminant Analysis of Feature Extraction for Adaptive Long-Term Wi-Fi Indoor Localization in Evolving Environments. *Sensors* **2024**, *24*, 5665. [\[CrossRef\]](#)
20. Lohan, E.S.; Torres-Sospedra, J.; Leppäkoski, H.; Richter, P.; Peng, Z.; Huerta, J. Wi-Fi Crowdsourced Fingerprinting Dataset for Indoor Positioning. *Data* **2017**, *2*, 32. [\[CrossRef\]](#)
21. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Zhong, Y.; Wu, F.; Zhang, J.; Dong, B. WiFi Indoor Localization Based on K-Means. In Proceedings of the 2016 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 11–12 July 2016; pp. 663–667.
23. Zhang, C.; Qin, N.; Xue, Y.; Yang, L. Received Signal Strength-Based Indoor Localization Using Hierarchical Classification. *Sensors* **2020**, *20*, 1067. [\[CrossRef\]](#)
24. Torres-Sospedra, J.; Quezada-Gaibor, D.; Mendoza-Silva, G.M.; Nurmi, J.; Koucheryavy, Y.; Huerta, J. New Cluster Selection and Fine-Grained Search for k-Means Clustering and Wi-Fi Fingerprinting. In Proceedings of the 2020 International Conference on Localization and GNSS (ICL-GNSS), Tampere, Finland, 2–4 June 2020; IEEE: Tampere, Finland, 2020; pp. 1–6.
25. Sunantasaengtong, P.; Chivapreecha, S. Mixed K-Means and GA-Based Weighted Distance Fingerprint Algorithm for Indoor Localization System. In Proceedings of the TENCON 2014–2014 IEEE Region 10 Conference, Bangkok, Thailand, 22–25 October 2014; IEEE: New York, NY, USA, 2015. [\[CrossRef\]](#)
26. Bai, S.; Wu, T. Analysis of K-Means Algorithm on Fingerprint Based Indoor Localization System. In Proceedings of the 2013 5th IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, Chengdu, China, 29–31 October 2013; pp. 44–48. [\[CrossRef\]](#)
27. Dong, B.; Wu, F.; Xing, J.; Zou, Y. Probabilistic Algorithm Based on Fuzzy Clustering for Indoor Location in Fingerprinting Positioning Method. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 155–159. [\[CrossRef\]](#)
28. Frey, B.J.; Dueck, D. Clustering by Passing Messages Between Data Points. *Science* **2007**, *315*, 972–976. [\[CrossRef\]](#)
29. Bi, J.; Wang, Y.; Yu, B.; Cao, H.; Shi, T.; Huang, L. Supplementary Open Dataset for WiFi Indoor Localization Based on Received Signal Strength. *Satell. Navig.* **2022**, *3*, 25. [\[CrossRef\]](#)
30. Pinto, B.; Barreto, R.; Souto, E.; Oliveira, H. Robust RSSI-Based Indoor Positioning System Using K-Means Clustering and Bayesian Estimation. *IEEE Sens. J.* **2021**, *21*, 24462–24470. [\[CrossRef\]](#)
31. Sun, H.; Zhu, X.; Liu, Y.; Liu, W. Construction of Hybrid Dual Radio Frequency RSSI (HDRF-RSSI) Fingerprint Database and Indoor Location Method. *Sensors* **2020**, *20*, 2981. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Ayub, A.; Abidin, Z.Z.; Alhammadi, A.; Soliman, N.F.; Khan, M.A.; Algarni, A.D. Comparative Study of Indoor Positioning Datasets Focusing on Localization Accuracy Success Rate and Floor Classification. *Sci. Rep.* **2025**, *15*, 34514. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.