WILEY | Hindawi

Research Article

Fingerprinting Indoor Positioning Method Based on Kernel Ridge Regression with Feature Reduction

Yanfen Le [ID], Shijialuo Jin, Hena Zhang, Weibin Shi, and Heng Yao

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 20093, China

Correspondence should be addressed to Yanfen Le; leyanfen@usst.edu.cn

An important goal of indoor positioning systems is to improve positioning accuracy as well as reduce power consumption. In this paper, we propose an indoor positioning method based on the received signal strength (RSS) fingerprint. The proposed method used a certain criterion to select fixed access points (FPs) in an offline phase instead of an online phase for location estimation. Principal component analysis (PCA) was applied to reduce the features of the RSS measurements but retain the most information possible for establishing the positioning model. Then, a kernel-based ridge regression method was used to obtain the nonlinear relationship between the principal components of the RSS measures and the position of the target. We thoroughly investigated the performance of the proposed method in realistic wireless local area network (WLAN) and wireless sensor network (WSN) indoor environments and made comparisons with recently developed methods. The experimental results indicated that the proposed method was less dependent on the density of the reference points and had higher positioning accuracy than the commonly used positioning methods, and it adapts to different application environments.

1. Introduction

Recent advances in information science have made it practical and accessible to provide indoor positioning services (IPS), for applications such as indoor personal navigation, healthcare, and environmental monitoring [1–3]. The key task for such systems is to determine the position of the user or portable device. Accurate positioning in complicated indoor environments is quite challenging and has received wide attention in the past decade [4–6].

A variety of wireless technologies have been used for indoor positioning. Some of them achieve high accuracy in the order of tens of centimeters, which usually requires additional specialized hardware [7–10]. Techniques based on radio frequency (RF) signals are considered as the most common cost-effective solutions for indoor positioning. Among them, the fingerprinting technique, which makes use of only received signal strength (RSS) values, has gained significant interest [11–14]. Generally, a predeployment site survey process or offline phase is required during which a radio map is constructed by collecting RSS samples (fingerprint) from location fixed access points (FP), which are access points in

a WLAN or anchor nodes in a WSN at different reference points (RP) over the whole positioning area. Then, the target is located during the online phase by matching the online RSS values with the prestored fingerprints. In general, such methods require neither the location of the FPs nor a wireless channel propagation model. The fingerprinting method was first proposed in [15], where the Euclidean distance between online and offline RSS values was used to select the matched RPs and then weights the positions of the RPs to obtain the estimated target position; this technique is also known as weighted K-nearest neighbor (WKNN) algorithm. The concept of the fingerprinting method is based on the fact that the RSS value at any given point is determined primarily by the surrounding environment and the location of the FPs and is, thus, unique. However, some shortcomings of RSS measurement, such as its vulnerability to environmental changes and random fluctuations, prevent the positioning accuracy from reaching a satisfactory level [16].

Various pattern recognition algorithms have been applied for better positioning performance [17, 18]. One aim of these algorithms is to improve the positioning accuracy even if there are anomalous or missing RSS

measurements. Classification algorithms such as support vector machine (SVM) [1, 19, 20], decision tree [21], or even neural networks classifier [22] have been used for indoor localization. In general, the number of classifier models increases with the number of RPs when indoor localization is considered as a multiclass classification problem, which incurs high time and memory consumption. Some models use regression algorithms with features extracted from the RSS values to build a robust and adaptive relationship between the positions and the RSS measures. The sparse recovery algorithm, the least absolute shrinkage and selection operator (LASSO) algorithm, kernel ridge regression (KRR), and Elastic-Net algorithm have been used to model linear or nonlinear relations for better performance [23–26].

Feature extraction is a commonly used method for data reduction and noise elimination [27, 28]. The new method we proposed uses feature reduction before a relationship is established between the locations and the RSS values. The proposed technique uses principal component analysis (PCA) to find a new presentation of data in terms of least square [29, 30]. Thus, a new set of variables called principal components (PCs) instead of RSS values is obtained. The advantages of using PCA to reduce features are as follows: first, PCs are compact descriptions of the measured RSS values. Choosing a proper number of PCs eliminates redundancy but retains maximal information, which is more conducive to establishing the positioning model. The significance of PCA also lies in the reduction of the computational complexity. In general, the dimension of RSS features, or the number of FPs, is an important factor in determining computational complexity during the entire positioning process. Instead of choosing a subset of FPs as reported in previous works [23, 24, 31], PCA selects a subset of PCs but retains information from all the FPs with low dimensionality. Then, a kernel-based model is used to explore a proper linear combination of kernels with selected PCs to represent the position. In the experiments, the localization system is evaluated thoroughly using the collected RSS data from two actual indoor environments, including a corridor with WLAN signals and a hall with WSN signals.

The rest of this paper is organized as follows: the details of the proposed positioning algorithm are described in Section 2. The experimental setup and a discussion of the results are introduced in Section 3. Finally, the conclusion is presented in Section 4.

2. Proposed Algorithm Based on PCA and KRR

This section describes the proposed fingerprinting method as shown in Figure 1, which consists of two phases: the offline phase and the online phase. In the offline phase, a site survey is carried out to collect the RP's information and the corresponding RSS values, then spatial filter based on FP coverage and FP selection according to certain criterion is used to select the proper set of FPs to build a radio map, and then the model training with the proposed PCA and KRR algorithm is performed. In the online phase, the localization process using the PCA reduction and KRR is performed to get the target position.

2.1. Radio Map Construction with FP Selection.
Consider an indoor environment where RF signals from FPs can be received throughout the area. A number of RPs, denoted as $[P_1, P_2, \cdots, P_N]$, are set, and their coordinates $p(P_i) = [x_i, y_i]$ are recorded. The RF signals are scanned at each RP, and RSS values collected from the same FP are averaged and then stored as a fingerprint into radio map in a vector $F_i = [f_{i1}, \cdots, f_{ij}, \cdots, f_{iK}]$, where f_{ij}, $i = 1, \cdots, N, j = 1, \cdots, K$, holds the time average RSS values from FP_j at location P_i. Let $\chi = \{FP_1, \cdots, FP_K\}$ be the set of FPs. Because not all FPs detected at the site are available at each RP, a certain value, -100 dBm for WSN-based anchors and 0 dBm for WLAN-based APs in our experiments, is experimentally set to imply an FP's unavailability.

It is also worth mentioning that in an indoor environment, especially in a WLAN-based positioning scenario, up to dozens of APs can usually be detected throughout the site. For example, in our experiments, 198 APs in total were detectable on one floor, which is far more than the number of FPs required for positioning. The signal from each FP could provide some information about the location, while too many FPs used for positioning increases the cost of storage and the computation of the algorithm. In addition, FPs having a large variance of RSS values are not suitable for positioning and could contribute to large errors. Thus, a certain criterion should be used to select a particular set of FPs that present the characteristics of the signal distribution effectively.

In our work, we considered three FP section methods to find a balance between the number of FPs and the positioning accuracy they can achieve. Because the FP selection is applied in the offline phase in our method, the radio map constructed with the selected FPs is to be used for further positioning model setup. Therefore, it is necessary that FPs detected in the online phase be as consistent as possible with those selected in the offline phase. Therefore, a spatial filter based on the coverage is first applied to find the FPs with the largest coverage. We define the FP coverage as

$$C_j = \sum_{i=1}^{N} c_{ij}, j = 1, 2, \cdots, K, \tag{1}$$

where $c_{ij} = 1$ if the RSS value from FP_j is available at RP_i and is 0 otherwise. The FP_j with $C_j \geq C_0$ is selected as the basis of the subsequent FP section, where C_0 is a threshold varying with the number of RPs in different scenarios. This value is set to 9 in one of our implementations, which means at least 9 RPs can receive RSS data from each filtered FP. After the spatial filter, a group of $|H|$, $H \subset \chi$, FPs are selected. FP selection is then carried out within this group.

Three FP selection criteria, denoted as Strongest FPs, Least-variance FPs, and Combined criterion, are evaluated in our work.

(1) *Strongest FPs*. This criterion assigns a score to an FP according to the strength of its signal throughout the site, which is defined as:

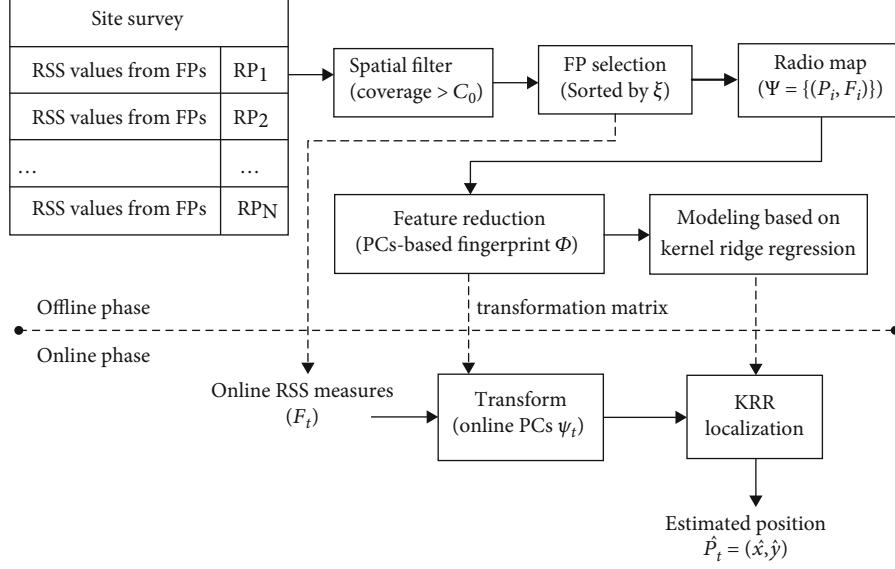$$\xi_j^1 = \sum_{i=1}^{N} f_{ij}, j \in H, \tag{2}$$

FIGURE 1: Flow diagram of the proposed indoor positioning method.

where f_{ij} is the time-average RSS value. The idea behind this criterion is that bigger RSS values can reveal more features of the site. The FPs are sorted in decreasing order according to their scores, and then a subset of $|S^1|$, $S^1 \subseteq H$, FPs with the largest ξ^1 are selected.

(2) *Least-variance FPs.* In this criterion, scores are also assigned to FPs but with focus on the variance of the RSS signals throughout the collection time. The score of FP_j is calculated as follows:

$$\xi_j^2 = \sum_{i=1}^{N} \Delta_{ij}, j \in H, \tag{3}$$

where Δ_{ij} is the unbiased estimated variance of the RSS readings from FP_j at location P_i. An FP with a low variance indicates that it can provide stable signals over time, which is conducive to a correct match between the online RSS and the fingerprints. It is important to note that some RPs cannot receive the signal from FP_j as mentioned earlier; the variance of these RPs needs to be set to a large value (e.g., 100 in our implementation). The FPs are sorted in increasing order according to their scores, and then a subset of $|S^2|$, $S^2 \subseteq H$, FPs with the lowest scores are selected.

(3) *Combined criterion selected FPs.* The criterion accounts for both the spatial distribution of RSS across all RPs and the RSS variance over the collecting time. According to this criterion, each FP is assigned a score calculated as follows:

$$\xi_j^3 = \frac{\sum_{i=1}^{N} \left(f_{ij} - \bar{f}_j \right)^2}{\sum_{i=1}^{N} \Delta_{ij}}, j \in H, \tag{4}$$

where

$$\bar{f}_j = \frac{1}{N} \sum_{i=1}^{N} f_{ij}, j \in H. \tag{5}$$

A higher score means better stability of RSS over time as well as greater discriminability of FPs across the site. A subset of $|S^3|$, $S^3 \subseteq H$, FPs with the highest ξ^3 are selected.

Now assume that K' FPs are retained after selection to build the radio map Ψ, which is denoted by:

$$\Psi = \{(P_i, F_i)\} = \begin{bmatrix} p_1 & F_1 \\ p_2 & F_2 \\ \cdots & \cdots \\ p_N & F_N \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & f_{1,1} & \cdots & f_{1,K'} \\ x_2 & y_2 & f_{2,1} & \cdots & f_{2,K'} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_N & y_N & f_{N,1} & \cdots & f_{N,K'} \end{bmatrix}, \tag{6}$$

The total fingerprint $F = [F_1, F_2, \cdots, F_{N'}]^T$ can be considered prior knowledge of the signal characteristics of the surveyed site. In the online phase, the target gets a new F_t to determine its coordinate $\hat{p}_t$, which is the key task of the indoor positioning. In this way, a radio map Ψ is crucially important for the proper training of the regression model.

2.2. Feature Reduction with PCA. PCA is a fast and efficient technique of dimensionality reduction used widely [32]. Instead of directly using all the selected FPs, our approach replaces the RSS values with a subset of PCs, which is obtained by a transformation based on the PCA. Given a transformation between vector F and ψ as $\psi = FW$, where $F \in \Re^{1*K'}$, $\psi \in \Re^{1*K}$, and $k < K'$. F should be reconstructed from ψW^T and labeled as $\hat{F}$. PCA seeks to solve the problem

of minimizing the mean square reconstruction error, which can be expressed by:

$$\underset{W}{\arg\min} \sum_{i=1}^{N} \left\| \hat{F}_i - F_i \right\|_2^2 = \underset{W}{\arg\min} \sum_{i=1}^{N} \left\| \psi_i W^T - F_i \right\|_2^2, \quad (7)$$

where W is the transformation matrix and N is the number of training samples. This optimization problem has been solved well, and W can be obtained by the following steps. First, calculate the covariance matrix of all RSS samples defined by:

$$C_\sigma = \sum_{i=1}^{N} \left(F_i - \bar{F} \right) \left(F_i - \bar{F} \right)^T, \quad (8)$$

where $\bar{F} = 1/N \sum_{i=1}^{N} F_i$ is the mean of the samples, C_σ is a positive-semidefinite symmetric matrix, and its eigenvalues are easy to compute and sort in descending order as $\{\lambda_1, \lambda_2, \cdots, \lambda_K{'}\}$. The corresponding normalized eigenvectors $\{w_1, w_2, \cdots, w_K{'}\}$ are geometrically orthonormal and statistically uncorrelated. Then, the $K' \times k$ transformation matrix W has the form:

$$W = [w_1 \cdots w_k]^T. \quad (9)$$

Here k is the number of PCs selected for the best transmission from RSS values to PCs with reduced dimension for maximal extraction of signal features. Then, the proper value of k is determined by cross-validation with the sampled fingerprints in the radio map during the offline phase. When the matrix W is available, the new PC-based fingerprinting Φ transformed from the RSS-based F is given by:

$$\Phi = FW = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{bmatrix} = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1k} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_{N1} & \varphi_{N2} & \cdots & \varphi_{Nk} \end{bmatrix}. \quad (10)$$

A regression model can be built between the new Φ and the locations.

During the positioning stage, the online PCs can be directly extracted from the online RSS measurements using the trained matrix W:

$$\psi_t = F_t W = [\varphi_{t1}, \varphi_{t2}, \cdots, \varphi_{tk}]. \quad (11)$$

With the extracted Ψ_t, the regression model can be used to determine the position of the target.

2.3. KRR Approach Based on PCs. A brief review is given for a general understanding of the theory of following adopted kernel ridge regression. Given a training set $\{(\psi_1, p_1), \cdots, (\psi_N, p_N)\}$, where N is the number of samples, each $\psi_i \in \mathbb{R}^k$ is a row vector in Φ denoting an input sample with

a corresponding out $p_i \in \mathbb{R}$. The ridge regression algorithm entails solving the following optimization problem [33].

$$\hat{\omega} = \arg \underset{\omega}{\min} \left[(P - \omega)^T (P - \Phi\omega) + \lambda \|\omega\|^2 \right], \quad (12)$$

where $\omega \in \mathfrak{R}^k$ is the weight vector, $\lambda \geq 0$ is a regularization parameter tuned to control the compromise between the training error and the complexity of the solution. Then, a regression model is found to describe the linear relationship between input vector ψ and output p, such that

$$p = \omega^T \psi, \quad (13)$$

where $\omega = \Phi^T (\Phi\Phi^T + \lambda I_k)^{-1} P$, and I_k is a $k \times k$ identity matrix.

Generally, in nonlinear cases, a kernel-based method is introduced to solve the problems, which map the samples into a higher dimensional feature space where the problems become linear separable, such that

$$\phi : \psi \in \mathbb{R}^k \longrightarrow \phi(\psi) \in \mathbb{R}^\Lambda, \quad (14)$$

where superscript Λ stands for a higher-dimensional space. The mapping ϕ which is chosen to convert nonlinear relation between the output and the independent input variables into linear relation is not necessary to know. Then, the regression can be constructed in the feature space, and the solution to the regression problem only depends on the dot product in the feature space. The kernel function satisfying Mercer's condition is introduced as a format of a dot product, $K(\psi_i, \psi_j) = \phi(\psi_i) \cdot \phi(\psi_j)$. Now, the weight vector can be rewritten as:

$$\omega = \Phi^T (K + \lambda I_N)^{-1} P = \Phi^T \alpha = \sum_{i=1}^{N} \alpha_i \psi_i, \quad (15)$$

where K is an $N \times N$ kernel matrix with the element of $\kappa(\psi_i, \psi_j)$, P is the coordinate sets of the RPs, and $\alpha = (K + \lambda I_k)^{-1} P$ can be determined by the PC-based fingerprinting Φ.

In the online phase, the target's position can be directly estimated as

$$\hat{p} = \omega^T \psi_t = \sum_{i=1}^{N} \alpha_i \psi_i^T \psi_t = \sum_{i=1}^{N} \alpha_i \kappa(\psi_t, \psi_i), \quad (16)$$

In our method, a Gaussian kernel is used:

$$\kappa\left(\psi_i, \psi_j \right) = \exp \left(-\frac{\left\| \psi_i - \psi_j \right\|^2}{2\sigma^2} \right), \quad (17)$$

where σ is the bandwidth of the Gaussian kernel. Considering that our output is two coordinates, $\hat{x}$ and $\hat{y}$, two
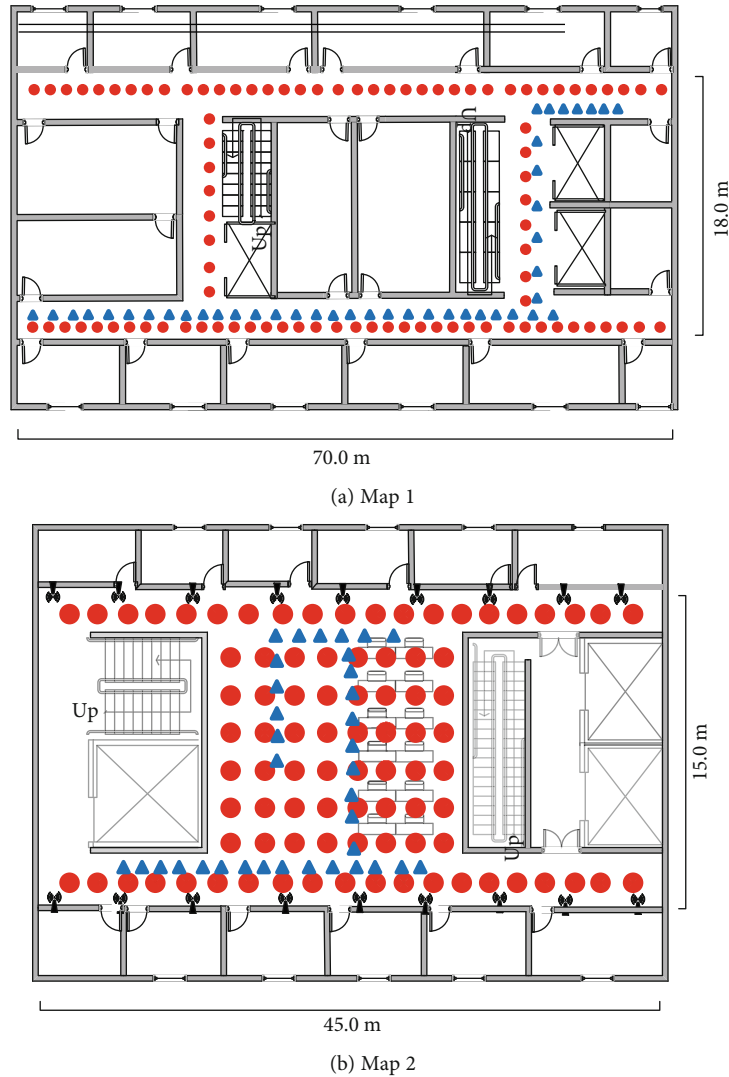
(a) Map 1

(b) Map 2

Figure 2: Maps of experimental sites. The red dots indicate the RPs, and the blue triangles represent the test points.

parameters, σ_x and σ_y, should be set and determined by the training RSS data.

3. Experimental Results and Analysis

In this section, we introduce the experimental setups and evaluate the performance of the proposed algorithm by comparing it with other algorithms. The performance of the proposed method is measured by the average positioning error (AE) of all test points and the empirical cumulative distribution function (CDF) of errors. The former is intended to calculate the Euclidean distance between the estimated and the actual location of the test point, while the latter indicates the maximum and minimum errors.

3.1. Experimental Setups. The experiments were conducted at two office sites of the Optical Engineering and Technology building at the University of Shanghai for Science and Technology, China, as shown in Figure 2. The first site labeled as Map 1 includes four hallways, where 92 locations are arranged as PRs with a grid spacing of 1.8 m. The BSSID (MAC address) and RSS data of the available WLAN APs were collected with a scanning interval of 3 seconds using a HUAWEI Android cell phone within 27 seconds at each RP. A total of 198 APs, all in the rooms or from the next floors, were detected throughout the site. The RSS data of the desired APs were averaged to construct the radio map. The variances of the RSS data were also obtained. Online RSS measures were collected on different days using the same device at 42 test points. The second site labeled as Map 2 includes two hallways and one lobby, which is equipped with WSN nodes based on TI CC2430 including 90 RPs spaced 1.8 m apart and 18 anchors working as FPs. The RSS measurements were obtained at each RP by moving node with sniffer software. During the offline phase, the RSS data were recorded for a period of 120 seconds (2 samples per second) over each RP. The online observations were collected in the same way with 36 selected test points.

Figure 3 shows the number of received FPs at each RP of the two sites in the offline phase. It can be seen from Figure 2
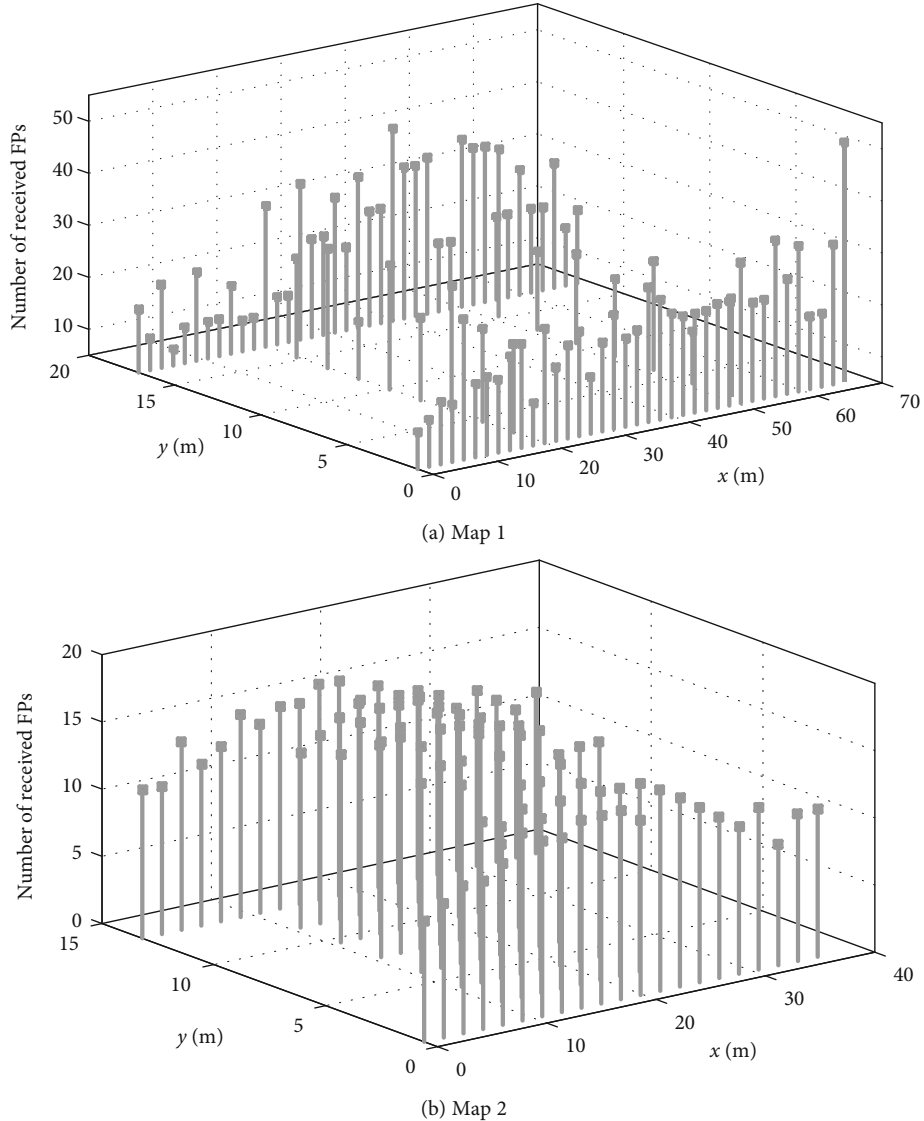
(a) Map 1

(b) Map 2

Figure 3: The number of received FPs at each RP in the offline phase.

that most of the RPs receive 10~30 FPs on Map 1, while on Map 2, this number is 10~15. No RP can receive all the FPs' signals on both maps. In these two typical indoor environments, especially that of Map 1, the dense distribution of rooms and walls greatly attenuated or obstructed the radio signals. Although a large number of FPs could be detected in the whole area, the FPs available at each RP were greatly reduced.

3.2. Experimental Results with Different FP Selections. Localization performance is always related to the number of FPs (K') used for positioning. In our method, FPs with certain coverage areas were first selected using a spatial filter. The threshold C_0 varied with the layout of the map, as well as the distributed RPs and FPs. Figure 4 shows the CDF of the positioning error when different thresholds were applied to filter the FPs. On Map 1, although there are up to 198 FPs, only 52 FPs can be received by more than 15 RPs out of a total of 92. The number K' increases to 89 when the thresh-

old drops to 9. With the FP selection criterion, 80 FPs were used for positioning, and almost no performance difference was observed as shown in Figure 4(a). Figure 4(b) shows the experimental results for Map 2. When $C_0 \leq 50$, all 18 FPs were received by all RPs. This may be attributed to the regular distribution of WSN anchor nodes and an indoor environment without many barriers. Considering this, no spatial filter was used in the following experiments for Map 2.

Figure 5 depicts the average location errors under the three different FP selection criteria, namely, the Strongest FPs, Least-variance FPs, and Combined criterion, when a proper subset of PCs is used for KRR-based localization.

It can be seen that no matter which FP selection criterion is adopted, the positioning method with more selected FPs obtains higher positioning accuracy at both sites. As shown in Figure 5(a), when 80 FPs are adopted, that is, 80 FPs are selected from the spatial filtered 98 FPs according to the FP selection criterion, the AE under the three criteria is all
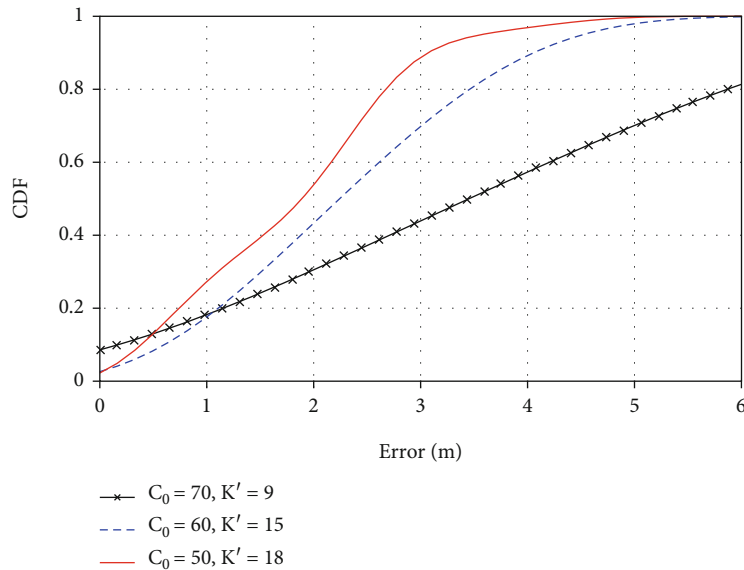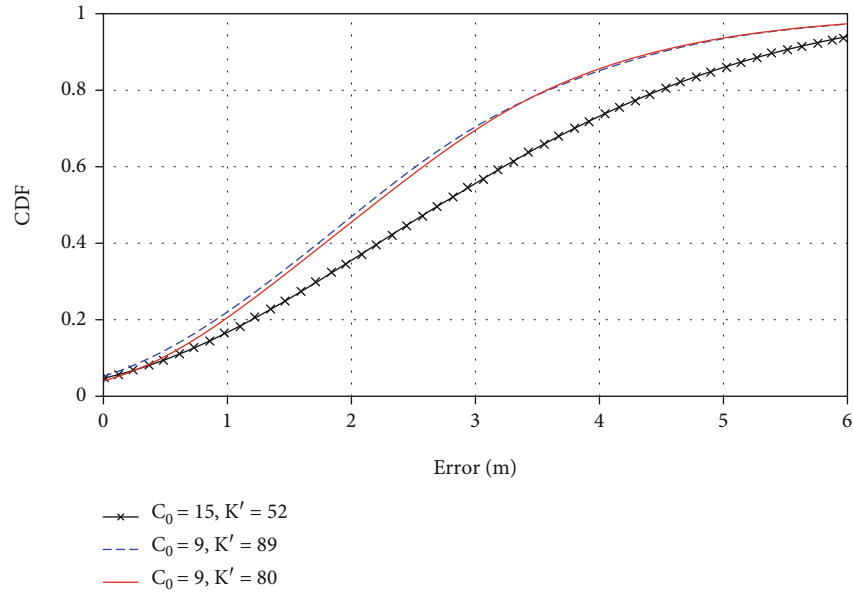
(a) Map 1

(b) Map 2

FIGURE 4: Positioning error of the proposed method with different thresholds for the spatial filter.

approximately 2.4 m. Almost the same result can be observed in Map 2. It is worth mentioning that even if we choose 80 FPs in Map 1 according to a certain FP selection criterion, such as the strongest FPs, only 18 of the RPs throughout the whole site can receive signals from more than 20 FPs. The situation is even worse when only 40 FPs are used; only one of the 92 RPs can receive signals from more than 20 FPs, which may explain the relationship between the positioning error and the number of FPs in Figure 5(a). For the positioning method that needs to establish the model between positions and signals in the offline phase, it is necessary to collect enough FP RSS measurements to extract features, especially in indoor environments with rooms and walls that greatly attenuate or completely block signal propagation.

This is a completely different approach from those that apply FP selection in the online phase, as proposed in [23, 24, 31].

3.3. Analysis of the Number of PCs. The proper choice of k (the number of PCs) is important to the proposed method. A larger value of k seems to retain more information after PCA transformation, but it may also retain redundant information and noise, which can be seen from the experimental results. Figure 6 compares the average positioning errors for the different number of FPs versus the number of PCs at both sites.

As illustrated in Figure 6, no matter how many FPs are used in the experiments, the positioning accuracy varies with the number of PCs, and there is always an optimal value k_p
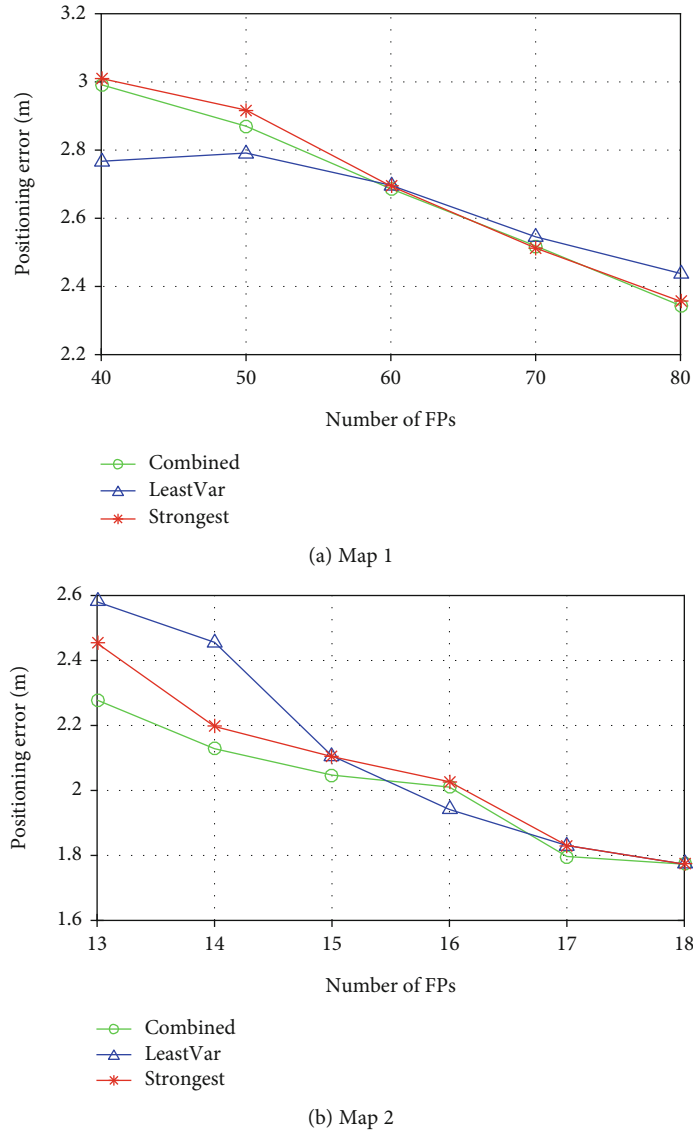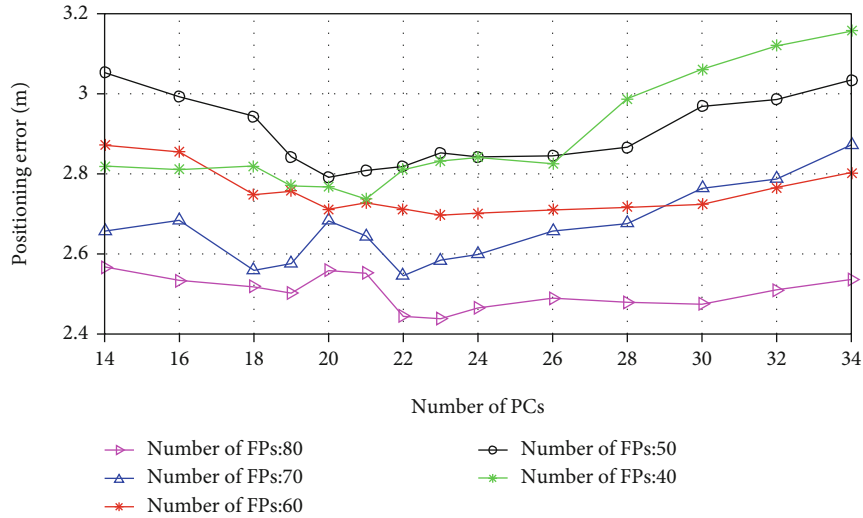
(a) Map 1

(b) Map 2

FIGURE 5: Average positioning error of the proposed method using three different FP selection criteria.

for the highest positioning accuracy, but this value is certainly not the maximal one. For example, in the case of the 80 FPs adopted in Map 1, the optimal k_p is 23. Further research found that this value k_p ensured that the selected PCs retained more than 85% of the total information if eigenvalues were used to quantify the information contribution of each PC, which provides a relatively simple way to determine this optimal value. It is also worth mentioning that no matter how many PCs are extracted, the premise is to have sufficient FPs' information for high positioning accuracy. From Figure 6(a), the best performance in the case of 40 FPs is 2.77 m, while with 80 FPs, the worst performance is 2.56 m. The same is true for Map 2, where the corresponding values for the best and worst performance are 2.28 m and 1.88 m. It can also be inferred from Figure 5 that the number of FPs has a greater impact on performance compared with the number of PCs, which means considering enough FPs before choosing the right number of PCs.

3.4. *Reduction in the Number of RPs.* Figure 7 shows the comparison of positioning error among Euclidian-WKNN, LASSO, KRR, and the proposed PCA-KRR method with respect to the number of RPs. For Map 1, a total number of 80 FPs selected with the Combined criterion were used with the optimal value of k_p. The number of FPs is 18 for Map 2. From the results displayed in Figure 7, we can see that the accuracy of the positioning gradually drops for both maps as the number of RPs used for positioning decreases, while the proposed PCA-KRR method achieves significantly better accuracy for both maps than the other methods regardless of the number of RPs. For instance, when the number of PRs used was 92 and a quarter of RPs (23) in Map 1, the AE of the proposed PCA-KRR method is, respectively, 2.34 m and 2.88 m, which are lower than the 2.71 m and 3.93 m of WKNN and the 2.78 m and 3.20 m of KRR. It is worth noting that the LASSO algorithm shows a different positioning performance for each of the two maps. For Map 1, the reduction

(a) Map 1
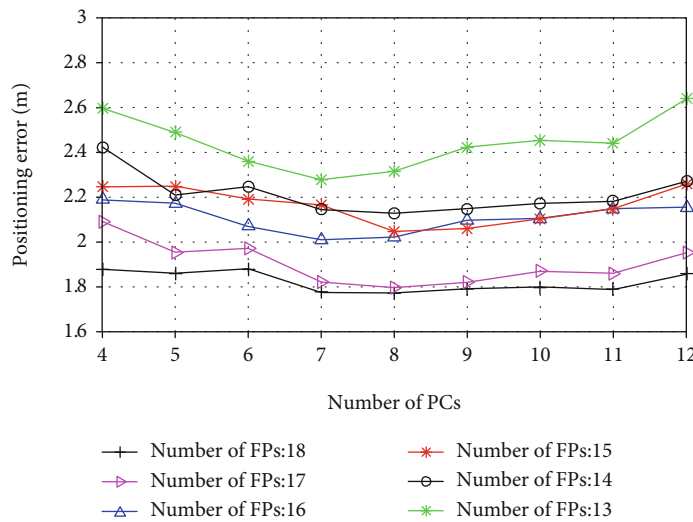
(b) Map 2

FIGURE 6: Average positioning error versus the number of PCs. (a) FPs were selected with Least-variance FP selection. (b) FPs were selected with Combined criterion FP selection.

in the number of RPs does not affect the positioning accuracy. While for Map 2, the result is different. The reduction of RPs introduces a large positioning error, indicating that LASSO is not robust enough for various environments. Furthermore, from these experimental results, it can be deduced that the density of RP does affect the positioning accuracy of our method, but the influence is limited. For example, when only a quarter of the RPs are used for positioning, i.e., the space between the RPs is as large as 7.2 m, the AE of the PCA-KRR method still reaches 2.88 m and 2.13 m, respectively, in Map 1 and Map 2 when enough FPs are guaranteed. Therefore, it is suggested that the number of FPs should be as high as possible while the density of RPs can be appropriately reduced to ensure the positioning accuracy while reducing the cost of labor.

3.5. Tuning the Parameters. There are several ways to tune the regularization parameter λ and the bandwidth of the Gaussian kernel σ involved in the proposed method. One effective approach for setting these parameters is the well-known cross-validation (CV). Another approach is to use the training data to find the best values that make the positioning algorithm the most accurate, which has been applied in this paper. We use some of the collected RSS data at the test points for parameter training. The experimental results show that the best value of λ changes only with the positioning site and does not change with the number of RPs and PCs used in the method. This value is 0.001 and 0.0001 in Map 1 and Map 2, respectively. As mentioned earlier, there are also two kernel widths σ_x and σ_y for two coordinates. Table 1 shows the positioning error for the different σ values used for positioning in the two maps. When σ_x and σ_y take the same value, there is no significant impact on the positioning accuracy, which can be observed from the results for both maps. Whether the two parameters choose the same value
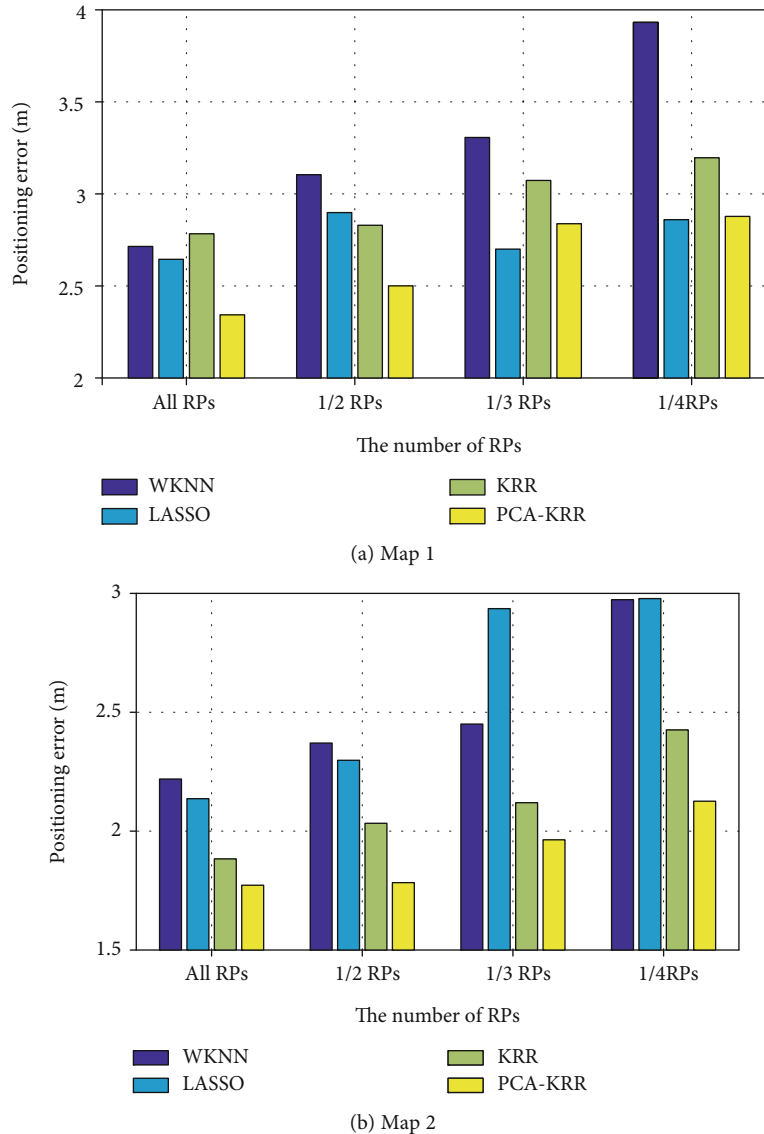
(a) Map 1

(b) Map 2

FIGURE 7: Average positioning error versus the number of RPs of WKNN, LASSO, KRR, and the proposed method. 1/2, 1/3, and 1/4 RPs mean, respectively, a half, a third, and a quarter of the total RPs in the maps.

TABLE 1: Positioning error with different bandwidth of the Gaussian kernel σ used in the method.

Site	Number of RPs N	Number of PCs k	σ_x	σ_y	AE/m
Map 1	92	22	11	11	2.34
			10	7	2.30
	46	25	12	12	2.50
			13	11	2.49
Map 2	90	8	14	14	1.77
			15	8	1.75
	45	9	15	15	1.78
			11	15	1.74

or not affects the positioning accuracy within 2%. In our experimental results above, σ_x and σ_y share the same value for the simplification of parameter turning.

4. Conclusion

In this paper, we have proposed a new RSS fingerprinting-based positioning method for indoor localization. We showed that PCA-based feature reduction can efficiently extract the RSS features related to positions, and the nonlinear ridge regression can build a proper model between the features of RSS and positions. The proposed positioning method, PCA-KRR, provides high positioning accuracy in various indoor environments. Extensive experimental results have demonstrated that the performance of the proposed method is sensitive to the number of APs or anchor nodes, while the sparsity of the RPs did not reduce the accuracy of

the method much. This is conducive to practical application for reducing the labor cost of fingerprints collection. We also have shown that the proposed method can efficiently and effectively locate a target through WLAN signals or WSN signals.

Data Availability

All data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Pei, J. Liu, R. Guinness, Y. Chen, H. Kuusniemi, and R. Chen, "Using LS-SVM based motion recognition for smartphone indoor wireless positioning," *Sensors*, vol. 12, no. 5, pp. 6155–6175, 2012.

[2] M. D. Rodriguez, J. Favela, E. A. Martinez, and M. A. Munoz, "Location-aware access to hospital information and services," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 4, pp. 448–455, 2004.

[3] Y. Wu, P. Chen, F. Gu, X. Zheng, and J. Shang, "HTrack: an efficient heading-aided map matching for indoor localization and tracking," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 3100–3110, 2019.

[4] A. K. M. M. Hossain and W. Soh, "A survey of calibration-free indoor positioning systems," *Computer Communications*, vol. 66, pp. 1–13, 2015.

[5] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1974–2002, 2017.

[6] X. Tian, R. Shen, D. Liu, Y. Wen, and X. Wang, "Performance analysis of RSS fingerprinting based indoor localization," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2847–2861, 2017.

[7] K. Joshi, S. Hong, and S. Katti, "Pinpoint: localizing interfering radios," in *Proceeding of Usenix Conference on Networked systems Design and Implementation*, pp. 241–253, Lombard, IL, 2013.

[8] J. Xiong and K. Jamieson, "Arraytrack: a fine-grained indoor location system," in *Proceeding of Usenix Conference on Networked systems Design and Implementation*, pp. 71–84, Lombard, IL, 2013.

[9] N. A. Alsindi, B. Alavi, and K. Pahlavan, "Measurement and modeling of ultrawideband TOA-based ranging in indoor multipath environments," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1046–1058, 2009.

[10] Q. Fan, B. Sun, Y. Sun, and X. Zhuang, "Performance enhancement of MEMS-based INS/UWB integration for indoor navigation applications," *IEEE Sensors Journal*, vol. 17, no. 10, pp. 3116–3130, 2017.

[11] W. Xue, K. Yu, X. Hua, Q. Li, W. Qiu, and B. Zhou, "AP's virtual positions-based reference point clustering and physical distance-based weighting for indoor Wi-Fi positioning," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3031–3042, 2018.

[12] Y. Le, Z. Tang, C. Sheng, and W. Shi, "Fast and resource efficient method for indoor localization based on fingerprint with varied scales," *Journal on Communications*, vol. 40, 2019.

[13] C. Kumar and K. Rajawat, "Dictionary-based statistical fingerprinting for indoor localization," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8827–8841, 2019.

[14] M. M. Atia, A. Noureldin, and M. J. Korenberg, "Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1774–1787, 2013.

[15] P. Bahl and V. Padmanabhan, "Radar: an in-building RF-based user location and tracking system," in *Proceeding of Infocom Nineteenth Joint Conference of the IEEE Computer & Communications Societies*, pp. 775–784, Tel Aviv, Israel, 2000.

[16] X. Fang, Z. Jiang, L. Nan, and L. Chen, "Optimal weighted K-nearest neighbour algorithm for wireless sensor network fingerprint localisation in noisy environment," *IET Communications*, vol. 12, no. 10, pp. 1171–1177, 2018.

[17] J. Xue, J. Liu, M. Sheng, Y. Shi, and J. Li, "A WiFi fingerprint based high-adaptability indoor localization via machine learning," *China Communications*, vol. 17, pp. 247–259, 2020.

[18] J. Yoo, "Change detection of RSSI fingerprint pattern for indoor positioning system," *IEEE Sensors Journal*, vol. 20, no. 5, pp. 2608–2615, 2020.

[19] G. Sun and W. Guo, "Robust mobile geo-location algorithm based on LS-SVM," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 3, pp. 1037–1041, 2005.

[20] C. Chen, Y. Wang, Y. Zhang, and Y. Zhai, "Indoor positioning algorithm based on nonlinear PLS integrated with RVM," *IEEE Sensors Journal*, vol. 18, no. 2, pp. 660–668, 2018.

[21] Y. Chen, Q. Yang, J. Yin, and X. Chai, "Power-efficient access-point selection for indoor location estimation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 7, pp. 877–888, 2006.

[22] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: a deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 763–776, 2016.

[23] C. Feng, W. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Transactions on Mobile Computing*, vol. 11, no. 12, pp. 1983–1993, 2012.

[24] A. Khalajmehrabadi, N. Gatsis, D. J. Pack, and D. Akopian, "A joint indoor WLAN localization and outlier detection scheme using LASSO and elastic-net optimization techniques," *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2079–2092, 2017.

[25] S. Mahfouz, F. Chehade, P. Honeine, and J. Farah, "Non-parametric and semi-parametric RSSI/distance modeling for target tracking in wireless sensor networks," *IEEE Sensors Journal*, vol. 16, no. 7, pp. 2115–2126, 2016.

[26] Y. F. Le, H. Zhang, W. Shi, and H. Yao, *RSS-based Indoor Positioning Algorithm Using Advanced Clustering and Kernel Ridge regression*, Frontiers of Information Technology & Electronic Engineering, 2020.

[27] Z. Zhou, Q. M. J. Wu, Y. Yang, and X. Sun, "Region-level visual consistency verification for large-scale partial-duplicate image

search," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 2, pp. 1–25, 2020.

[28] Z. Zhou, Y. Mu, and Q. M. J. Wu, "Coverless image steganography using partial-duplicate image retrieval," *Soft Computing*, vol. 23, no. 13, pp. 4927–4938, 2019.

[29] S. Fang and T. Lin, "Principal component localization in indoor WLAN environments," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 100–110, 2012.

[30] A. Li, J. Fu, H. Shen, and S. Sun, "A cluster principal component analysis based indoor positioning algorithm," *IEEE Internet of Things Journal*, vol. 14, pp. 1–10, 2021.

[31] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Structured group sparsity: a novel indoor WLAN localization, outlier detection, and radio map interpolation scheme," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6498–6510, 2017.

[32] X. Liang, Z. Tang, X. Xie, J. Wu, and X. Zhang, "Robust and fast image hashing with two-dimensional PCA," *Multimedia Systems*, pp. 1–13, 2020.

[33] M. Maalouf and D. Homouz, "Kernel ridge regression using truncated Newton method," *Knowledge-Based Systems*, vol. 71, pp. 339–344, 2014.