IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 22, 2019, accepted December 3, 2019, date of publication December 13, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958957

Smartphone-Based Indoor Fingerprinting Localization Using Channel State Information

PENGPENG CHEN[1,2], FEN LIU[1], SHOUWAN GAO[1,2], PEIHAO LI[1], XU YANG[1,2], AND QIANG NIU[1,2]

[1]School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China
[2]Mine Digitization Engineering Research Center, Ministry of Education, Xuzhou 221116, China

Corresponding author: Qiang Niu (niuq@cumt.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2017QNA20.

ABSTRACT Indoor localization technology plays an important role in many indoor application scenarios. Existing WiFi-based indoor localization methods mainly obtain channel state information (CSI) through the personal computer, or obtain coarse-grained received signal strength (RSS) through the smartphone to finish the localization. Little work has been done on using smartphones to obtain fine-grained channel state information for localization. In this paper, we use the smartphone to collect fine-grained CSI that is more convenient and applicable, and propose a indoor fingerprinting localization. Compared with the CSI collected by the computer, the CSI signal collected by the smartphone fluctuates greatly. Hence, we corrects the CSI data through the signal processing technique and selects optimal subcarriers to obtain more stable and effective signals. In order to cope with the noisy WiFi environment, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method is used to remove abnormal sample points to reduce environmental interference. Moreover, the support vector machine multi-classification method is used for training and classification to achieve localization. Finally, we use the Google Nexus 5 smartphone to conduct experiments in two typical indoor environments. The localization accuracy is 91% and 86%, respectively, and both average localization errors are less than $0.5m$. Experimental results show that the proposed algorithm has higher localization accuracy compared with the typical algorithms.

INDEX TERMS Indoor localization, smartphone, channel state information, support vector machine.

I. INTRODUCTION

With the continuous development of wireless network technology [1], [2] and the rapid increase of mobile devices [3], [4], indoor localization and related applications have received extensive attention. Unlike outdoor localization systems (such as Global positioning system and Beidou satellite navigation system), indoor localization is subject to interference from some wireless signal propagation environments, including multipath effects, shadow fading and delay distortion. Therefore, outdoor localization technology is not suitable for indoor localization. How to provide good localization services in indoor environments has received extensive attention. WiFi-based indoor localization has become a research hotspot of indoor localization due to low cost and simple technology implementation without additional hardware equipment.

At present, WiFi-based indoor localization technology is mainly divided into fingerprint-based localization technology [5], [6], ranging-based localization technology [7], [8] and angle of arrival-based localization technology [9]–[11]. The fingerprint-based localization method has good robustness to the environment and high localization accuracy because it does not need to know the location information of the Access Point (AP) and the distance between the antennas in advance [12], [13], which has attracted the attention of many researchers.

Due to its simplicity and low hardware requirements, previous indoor localization systems used the Received Signal Strength (RSS) based localization method [14], [15]. However, this RSS-based approach has two main limitations. First, the RSS value usually has a high time-varying

The associate editor coordinating the review of this manuscript and approving it for publication was Anfeng Liu.

at a fixed position. Even for a fixed device, this high time variation can cause large position errors. Second, the RSS signal is easily interfered by multipath effects and noise in the indoor environment, and the RSS is the coarse-grained information. With the continuous development of network technology [16], channel state information (CSI) can now be obtained from some WiFi network interface cards (NICs), which is more fine-grained. It uses subcarriers [17], [18] based on Orthogonal Frequency Division Multiplexing (OFDM) technology to obtain more abundant multipath information. The indoor localization technology based on channel state information [19]–[21] has been widely used due to advantages of high processing efficiency, strong anti-interference and high localization accuracy. FILA is the first system to achieve indoor localization using fine-grained CSI instead of RSS [7]. It proposes the relationship between CSI value and distance, and applies trilateration to achieve localization. Jiang *et al.* [22] proposed a CSI-based indoor fingerprint localization system(FIFS), which used the diversity of time and space dimensions to process CSI to establish a fingerprint database, and combined Bayesian model to achieve position estimation. Wang *et al.* [23] proposed a indoor fingerprinting localization system (DeepFi) using deep learning and CSI. The weights of the deep network were trained as fingerprints using deep learning and the estimated position was obtained using a radial basis function based probabilistic method. Chapre *et al.* [24] proposed an indoor WiFi fingerprint system CSI-MIMO, which combined multiple input multiple output (MIMO) information and used the CSI amplitude and phase deviation between subcarriers to establish a fingerprint database. Qian *et al.* [25] established a theoretical model Widar, which can geometrically quantify the relationship between CSI changes and the human's position and moving velocity, thus achieving personnel localization and tracking. Mugahid and Yun [26] estimated the indoor distance for passive UHF RFID tag based on RSSI and RCS. Although the localization accuracy is improved, they may not work well in terms of scalability due to either high specialized infrastructure cost or extra device carrying. Whereas in fact, Smartphone-based approaches can efficiently utilize the available infrastructure for indoor localization. Schulz *et al.* [27] proposed that researchers' own WiFi test platforms with low-level mac and phy-access can be built by using firmware patches on mobile devices. By using the firmware patch on the smartphone, the Wi-Fi firmware is modified and CSI can be extracted on the smartphone [28]. And the Nexmon firmware patching framework is designed by Schulz *et al.* [29] to make MAC frame and physical-layer functionalities on Broadcom Wi-Fi chips accessible to researchers.

The existing WiFi-based indoor localization method has not yet appeared to use the smartphone to obtain fine-grained channel state information for localization and channel state information can only be obtained through the computer equipped with Intel 5300 NIC. However, the size of the computer is large, the battery life is short, and the mobility is poor, which makes people unable to carry it with them. And the computer is expensive and the operation is relatively complicated. Therefore, it is not suitable for some real indoor scenes. Nowadays, almost everyone uses smartphones. As smartphones play an increasingly important role in people's daily lives, people are increasingly inseparable from smartphones. Due to the convenience of smartphones, people carry smartphones almost everywhere. Compared with computers, it is more convenient to obtain fine-grained channel state information through smartphones. This paper uses the smartphone instead of the computer to collect CSI for indoor localization for the first time, which is more convenient and applicable. However, compared with the CSI collected by the computer, the CSI signal collected by the smartphone fluctuates greatly and the signal stability is poor, which is also an inherent problem of CSI signal collected by smartphone. In order to solve this challenge, this paper proposes the smartphone-based indoor fingerprinting localization using channel state information. This paper firstly processes the CSI data to correct the data, and obtains more stable and effective data to improve the localization accuracy by selecting the optimal subcarrier data and removing the abnormal sample points. The processed data is used as the final fingerprint, and then Support vector machine multi-classification algorithm is used to achieve localization. We conducted experiments in two real indoor environments, the corridor and laboratory. The localization accuracies are 91% and 86%, respectively, and both average localization errors are less than $0.5m$. Specifically, the main contributions of this paper are as follows:

- Realizing the inconvenience of network card-based solutions, we propose a smartphone-based indoor localization using CSI, which has the priority in pervasiveness in practical application rather than the computer. The fine-grained CSI is firstly collected by the smartphone and then used for localization.
- In order to solve the instability problem of signals collected by smartphones, some suitable pre-processing techniques are applied to correct the CSI data and an optimal subcarrier selection scheme is designed according to the characteristics of the CSI data collected by the smartphone.
- Extensive experiments have been conducted in multiple indoor environments, and the effects of methods used in this paper on the localization result have been studied. The experimental results show that the proposed algorithm outperforms state-of-the-art algorithms.

The rest of the paper is organized as follows: Section II introduces the basic theory of CSI. Section III presents the architecture of the system. Section IV reports the experimental design, evaluations and analyses. And section V concludes this paper and introduces future work.

II. CSI COLLECTION

A. CHANNEL STATE INFORMATION (CSI)

In the IEEE 802.11n/ac standard, orthogonal frequency division multiplexing (OFDM) technology is used to transmit

signals over multiple different orthogonal subcarriers. Today, detailed amplitude and phase information for different subcarriers can be provided in the form of CSI. CSI is also the channel property of the communication link. To a certain extent, it can describe the attenuation factor for each transmission path, revealing information such as signal scattering, mutipath fading and power attenuation.

As well known, the OFDM system with a narrowband flat fading channel in the frequency domain can be expressed as

$$y = Hx + n, \tag{1}$$

where y and x are the received and transmitted signal vectors, respectively, H and n are the channel information matrix and the additive white Gaussian noise (AWGN) vector, respectively.

According to Formula (1), the CSI of each subcarrier can be estimated as the below equation.

$$\hat{H} = \frac{y}{x}, \tag{2}$$

From here we can see that it is fine-grained and comes from the physical layer.

The CSI collected by the receiver can be divided into different subcarrier clusters, and the CSI matrix H can be expressed as

$$H = [H_1, H_2, \ldots, H_N], \tag{3}$$

where N is the number of subcarriers, $N = 56$ in the 20MHz bandwidth and $N = 114$ in the 40MHz bandwidth.

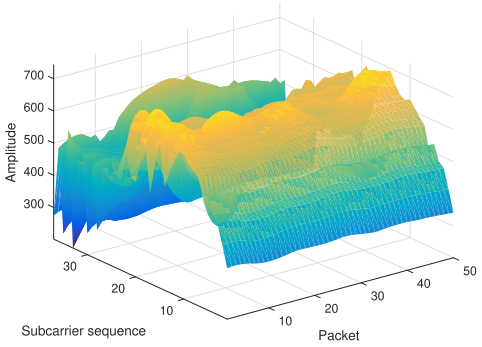
The signal can use the channel frequency response (CFR) to describe the multipath propagation of the signal from the amplitude frequency characteristic and the phase frequency characteristic respectively. CSI can be extracted in the frequency domain in the form of CFR, and CSI of a single subcarrier can be expressed as

$$H_i = |H_i| e^{j \sin\left\{\angle H_i\right\}}, \tag{4}$$

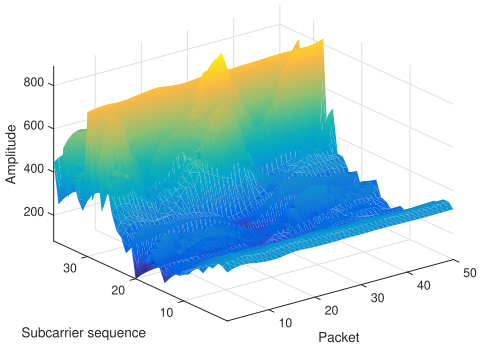
where $|H_i|$ and $\angle H_i$ are the amplitude and phase of the i-th subcarrier, respectively.

B. IMPACT OF SMARTPHONE LOCATION ON CSI

Since only the CSI amplitude information can be extracted through the smartphone, this paper only considers the CSI amplitude for fingerprint recognition, so as to solve the localization problem. The CSI amplitude value varies when the smartphone is in different positions. Hence this difference can be used to establish the mapping between location and CSI data. Fig. 1(a) and Fig. 1(b) show the amplitudes of the acquired subcarriers over packets when the smartphone is in two different positions. Obviously, the CSI data of these two positions show significant differences. Therefore, the localization problem can be solved by analyzing the CSI fingerprint characteristics.



(a) CSI data collected at location 1.



(b) CSI data collected at location 2.

FIGURE 1. CSI data at different locations.

III. SYSTEM ARCHITECTURE

In this paper, the CSI amplitude information of the reference point is collected by using the smartphone. First, the abnormal value processing, denoising and smoothing of the CSI signal are performed. Second, the optimal sub-carriers are selected and the abnormal sample points are removed. Then the processed data is used as the final CSI fingerprint information for building the fingerprint database. Finally, the support vector machine multi-classification algorithm is used to finish localization. The whole system architecture of this paper is shown in Fig. 2.

A. CSI SIGNAL PROCESSING

Compared with the CSI collected by the computer, the CSI signal collected by the smartphone fluctuates greatly, which is also an inherent problem of the CSI collected by the smartphone. To solve this problem, this paper processes the original CSI data to make them relatively stable and effective. First, the outliers are removed from the original CSI data. The main idea is to calculate the median value of the window composed of the sample data and its ten surrounding sample data, and estimate the standard deviation of each sample data with respect to the median value using the median absolute deviation. If the difference between the sample data and the median is greater than the three standard deviations, replace it with the median value. After removing the outliers, the noise

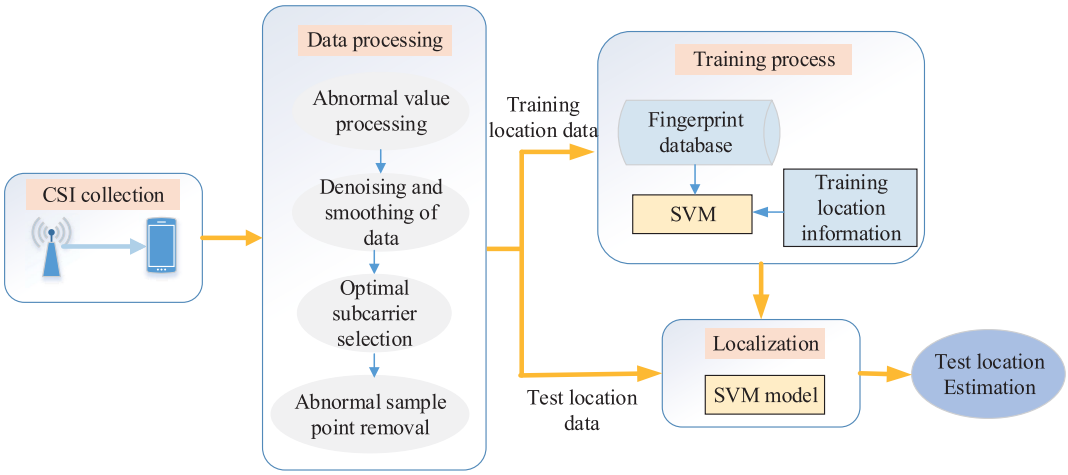


FIGURE 2. System architecture.

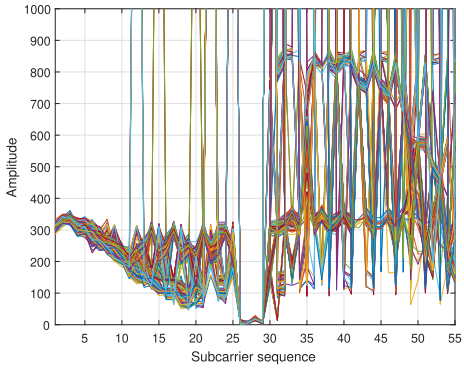
of CSI data is removed, which mainly uses the wavelet soft threshold denoising method to remove some noise existing in the environment. Then the signal is smoothed. The main step is the polynomial smoothing algorithm based on the least squares fitting principle. This paper uses some suitable filtering methods to realize the correction of CSI data.

Then the optimal subcarrier selection method is performed. In order to obtain more stable subcarriers, we remove the subcarriers with large amplitude fluctuations by using the variance selection method. The key idea is to set a suitable threshold. If the amplitude variance of a certain subcarrier is greater than the predefined threshold, it indicates that signal is unstable and should be removed. We first calculate the magnitude variance of each subcarrier over a period of time, and then set the second tertile of variances of all subcarriers as the variance threshold. Hence, only about one-third of the subcarriers will be removed, and the filtered features will not be large or small. By the optimal subcarrier selection method, some more stable subcarrier data are selected to further improve localization accuracy. In this paper, the information of 55 available subcarriers can be obtained from the smartphone and 37 optimal subcarrier information are selected by the variance selection method.

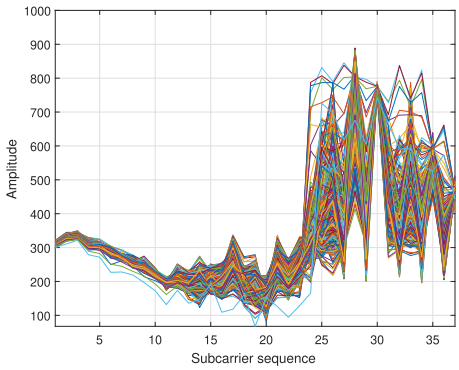
Fig. 3(a) shows the amplitude of the each original subcarrier that is continuously sampled 300 times at a certain location. A total of 55 subcarrier data can be obtained through the smartphone, and it can be seen that the CSI signal is unstable. Fig. 3(b) shows the CSI data processed by the above methods. The subcarrier data is corrected by the preprocessing method, and 37 optimal subcarrier data are selected by the optimal subcarrier selection method. The CSI data is normalized prior to classification to achieve higher precision and faster convergence. The CSI amplitude is mapped to [0, 1], and the formula is as follows:

$$CSI = \frac{CSI\text{-}CSI_{\min}}{CSI_{\max} - CSI_{\min}}, \tag{5}$$

where $CSI_{\min}$ represents the minimum CSI amplitude and $CSI_{\max}$ represents the maximum CSI amplitude.



(a) Original CSI data.



(b) Pre-processed CSI data.

FIGURE 3. CSI data before and after pre-processing.

Compared with the CSI collected by the computer, the CSI signal collected by the smartphone is very unstable, which is a challenge for the smartphone-based localization. After processing the signal by the aforementioned methods, there is still a lot of noise. If it is used directly for classification experiments, the error will be very large. Therefore, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method here is used to remove the abnormal sample points, and the sample points are automatically

classified into sub-classes of different sizes and shapes. In the final classification, the sample points become more efficient and relatively concentrated, thus the localization effect will be improved.

DBSCAN is a density-based clustering algorithm that divides the region of sufficient density into multiple clusters. In a spatial database with noise, clusters of any shape representing the largest set of density-connected points can be found. The closely packed points are grouped and the points in the low-density area are marked as outliers. There is not necessary to specify the number of clusters in advance and the outliers can be effectively detected and removed [30], [31]. In this paper, DBSCAN clustering algorithm is used to judge whether the sample point is an abnormal point, and the abnormal sample point is removed. The pseudo code is given in Algorithm 1. Where ε-neighborhood with sample point k as the core is $U_\varepsilon(\mathrm{k}) = [p \in H|\mathrm{dist}(k,p) \leq \varepsilon]$. $dist(k,p)$ is the distance between two sample points k and p in data set H, indicating similarity between samples, most commonly using Euclidean distance and Manhattan distance.

Algorithm 1 Abnormal Sample Point Removal With DBSCAN

Input: CSI data set H of all sample points, radius parameter ε, neighborhood density parameter *MinPts*;

Output: CSI data set H' for removing abnormal sample points;

```
Set all sample points in H as unvisited;
for Sample point K marked as unvisited in H do
    Set K as visited;
    Compute the number of sample points n in the ε-neighborhood of K;
    if n ≥ MinPts then
        Create a new cluster C and add K to C;
        Let S be the set of sample points in the ε-neighborhood of K;
        for Sample point K' marked as unvisited in S do
            Mark K' as visited;
            Calculate the number n of sample points in the ε-neighborhood of K';
            if n ≥ MinPts then
                Add all sample points in the ε-neighborhood of to S;
            end if
            if K' does not belong to any cluster then
                Add K' to C;
            end if
        end for
        Output C;
    else
        K is an abnormal sample point;
    end if
end for
Remove all abnormal sample points in H and output H'.
```

The DBSCAN algorithm has two important parameters ε and *MinPts*, where ε represents the neighborhood radius of the sample point and *MinPts* represents the neighborhood point threshold. However, these two parameters are not automatically estimated, usually based on the experimenter's experience. If the parameter value is not suitable, it is easy to lead to poor clustering results. Hence this paper uses the adaptive mesh method to estimate ε and *MinPts* according to the sample characteristics and experimental experience, which can reduce the clustering error. The optimal radius parameter ε and the neighborhood density parameter *MinPts* are set by the experiment. By removing the abnormal sample points, the noise is eliminated, the localization accuracy is improved. And the number of samples is reduced, so the operation rate is accelerated.

B. MODEL OF SVM LOCALIZATION ALGORITHM

The concept of Support Vector Machine (SVM) was originally proposed by Vapnik [32]. Nowadays, SVM methods have been widely used in classification and regression methods [33]–[35]. This paper employes the SVM classification method to achieve the aim of localization. The main idea is to assume that the sample set consists of two classes. After training with the samples marked with $y = 1$ and $y = -1$, a classifier is generated, and then correct labels of the unlabeled samples can be identified, so they can be correctly classified. By finding an optimal hyperplane, the SVM can distinguish between positive and negative marker samples as much as possible, and make the two types of samples farthest from the classification plane [36].

After obtaining the CSI amplitude information of the selected reference point, preprocessing the CSI data, extracting the feature values, and so on, the final CSI fingerprint information is obtained, and the fingerprint database is constructed. Let F_i, $F_i = [F_{i1}, F_{i2}, \ldots F_{in}]$, denotes the CSI fingerprint vector at the reference point i and n denotes the number of fingerprint attributes. The classification label corresponding to each fingerprint vector is $y_i = \pm 1$. After establishing the fingerprint database and category correspondence, the support vector machine classification method can be used to achieve localization.

In general, when SVM classification is performed, the sample set is not linearly separable. Therefore, the linear mapping, $\phi : F_i \rightarrow D$, is first used to map CSI fingerprint information from a low-dimensional space F_i to a higher-dimensional feature space D, thereby achieving linear separability in a high-dimensional space, where the classification hyperplane is

$$w^T \phi(F_i) + b = 0, \tag{6}$$

where w is a weight vector and b is a bias coefficient.

In order to maximize the distance between the points closest to the classification hyperplane in all sample points, the optimal hyperplane problem is transformed into solving the optimization problem, and the objective function is

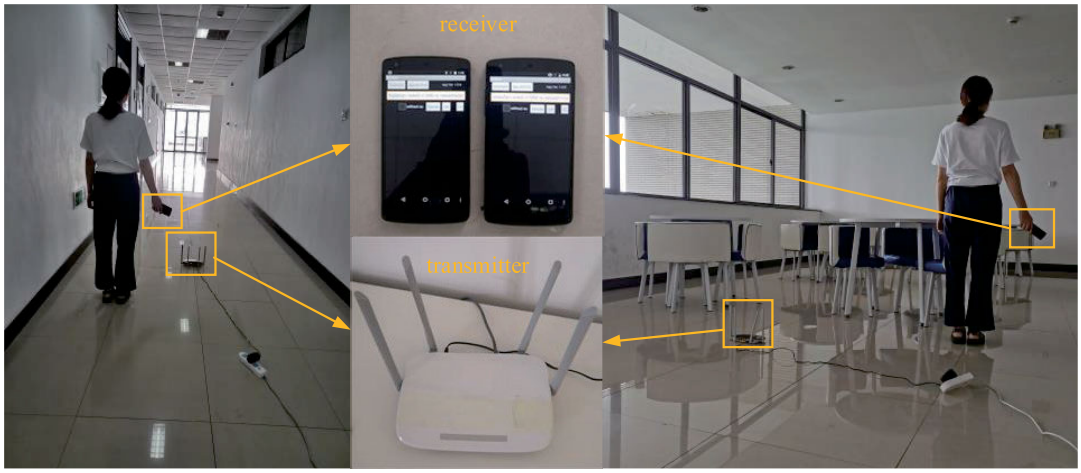


FIGURE 4. Experimental setup of the corridor and the laboratory.

solved as

$$\min \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{n}\xi_i, \tag{7}$$

where

$$\begin{cases} y_i\left(w^T\phi\left(F_i\right)+b\right) \geq 1-\xi_i \\ \xi_i \geq 0 \\ c > 0, \end{cases} \tag{8}$$

where ξ_i denotes the slack variable of the function interval allowed to control the sample point, and c denotes the penalty coefficient.

The above formula is a classic quadratic programming problem, and its Lagrangian polynomial is

$$L = \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{n}\xi_i - A - \sum_{i=1}^{n}\tau_i\xi_i, \tag{9}$$

where $A = \sum_{i=1}^{n}\alpha_i\left(y_i\left(w^T\phi\left(F_i\right)+b\right)-1+\xi_i\right)$, and α_i and τ_i are Lagrangian multipliers.

By solving, the above formula can be transformed into

$$L = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j k\left(F_i, F_j\right), \tag{10}$$

where $k\left(F_i, F_j\right) = \left\langle\phi\left(F_i\right), \phi\left(F_j\right)\right\rangle$ is the kernel function. By solving $\max L$, where $\sum_{i=1}^{n}\alpha_i y_i = 0$, $0 \leq \alpha_i \leq c$, $i = 1, \ldots, n$, and the optimal parameters, $\alpha^* = \left(\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*\right)$, can be obtained. The optimal weight vector and the optimal bias coefficient can be obtained by calculation, and the best discriminant function is obtained as follows:

$$\begin{aligned} f\left(F_i\right) &= \operatorname{sgn}\left[w^*\phi\left(F_i\right)+b\right] \\ &= \operatorname{sgn}\left[\sum_{j=1}^{n}\alpha_j^* y_j k\left(F_i, F_j\right) + b^*\right]. \end{aligned} \tag{11}$$

Therefore, the SVM classification algorithm can be implemented by the above formula, and the grid search method is used to select the optimal parameters c, gamma and kernel function. The support vector machine multi-classification algorithm mainly deals with multi-classification problems by constructing and combining multiple two-classifiers to realize the construction of multi-classifiers. This paper adopts one-against-one classification method, and the main idea is to construct a two-classifier between the training samples of any two categories to achieve multi-classification. If there are a total of k categories of training samples, k(k − 1)/2 two-classifiers can be constructed to classify the samples into the one with the largest number of samples.

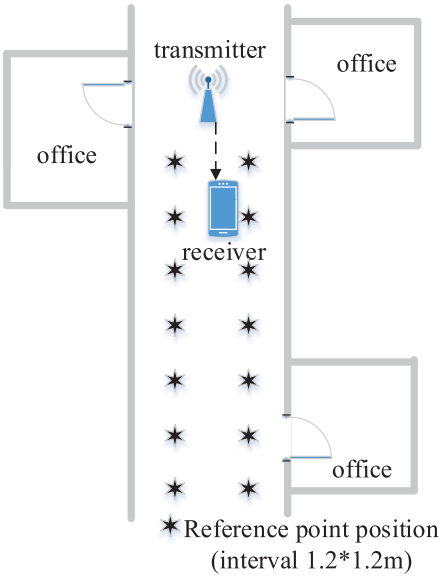


FIGURE 5. Floor-plan of the corridor.

IV. EXPERIMENT

A. EXPERIMENT SETUP

Experiments were conducted in the corridor and laboratory on the fifth floor of the academy of computer. The experimental environment of the corridor and laboratory is

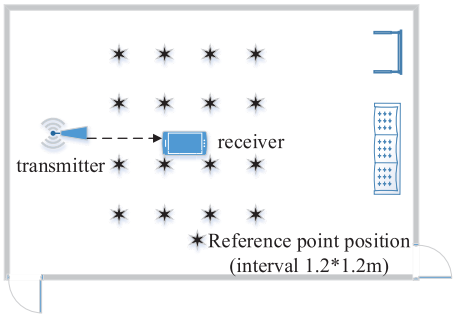


FIGURE 6. Floor-plan of the laboratory.

position	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	238	0	0	0	0	0	0	0	6	3	0	0	0	0
2	0	297	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	292	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	214	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	185	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	300	0	0	0	0	0	0	0	0
7	3	0	0	0	3	0	178	0	6	0	0	0	2	0
8	0	0	0	0	0	0	0	206	0	0	0	0	0	0
9	13	0	0	0	0	0	0	0	222	4	0	5	0	2
10	42	0	0	0	1	0	0	0	2	163	0	4	0	0
11	0	0	0	0	0	0	0	0	0	0	295	0	0	0
12	24	0	0	0	0	0	0	1	0	0	0	108	0	99
13	0	0	2	12	26	0	19	0	0	0	0	0	171	0
14	7	0	0	0	0	0	0	0	0	0	0	0	0	160

FIGURE 7. Confusion matrix of experimental results in the corridor.

position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	24	0	86	0	0	1	0	60	28	0	0	9	1	0	0	0
2	0	237	0	56	0	0	0	0	0	0	0	0	0	0	0	0
3	36	0	27	0	0	0	0	27	0	0	0	142	0	0	0	0
4	0	0	0	206	0	0	0	0	0	0	0	0	0	0	0	0
5	2	0	0	0	205	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	248	29	0	0	0	0	0	17	0	0	0
7	0	0	0	0	0	29	269	0	0	2	0	0	0	0	0	0
8	3	0	10	0	0	0	0	183	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	159	0	0	0	1	0	0	0
10	0	0	0	0	0	0	2	0	0	290	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	300	0	0	0	0	0
12	13	0	0	0	0	0	4	0	0	0	175	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	274	0	0	0	0
14	0	0	0	0	0	3	0	4	1	0	0	0	284	0	0	0
15	3	0	0	0	0	0	0	0	0	0	0	0	0	277	0	0
16	0	0	0	0	0	0	0	0	0	3	0	0	0	0	296	0

FIGURE 8. Confusion matrix of experimental result in the laboratory.

shown in Fig. 4 and the corresponding floor-plans are shown in Fig. 5 and Fig. 6, respectively. Nowadays, CSI can be obtained from the computer with the Intel 5300 NIC, but this paper uses the smartphone instead of the computer to collect CSI for indoor localization for the first time. We execute some commands on the smartphone whose WiFi firmware is modified to extract the CSI signal. Therefore, the experimental equipment only needs a smartphone and a wireless routing device, which is easy to operate and simple to implement.

We used the AC1200 dual-band wireless router as the transmitter, and the Google Nexus 5 smartphone as the receiver. We chose the 5GHz frequency band with 20MHz bandwidth of WiFi. The experimental area was divided into square small areas at intervals of 1.2m, that is, the area of each small area was 1.2 $m \times 1.2$ m. 16 training points and 16 test points were selected in the laboratory, and 14 training points and 14 test points were selected in the corridor. The original CSI data are collected from the connected AP at each reference position by the smartphone. 500 samples

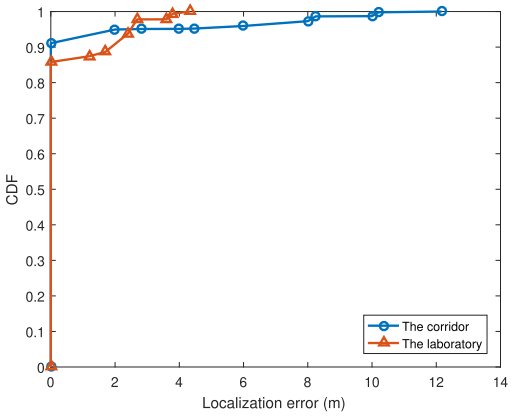


FIGURE 9. CDF of localization errors in the corridor and laboratory.

are continuously sampled at each position, and 100 samples before and after the sample points are respectively removed, so 300 samples are obtained for the experiment. The smartphone can obtain 55 available subcarrier information in total. We first collected the CSI data of the training positions, and then collected the CSI data of the test positions after a period of time, that is, the training samples and the test samples are not acquired at the same time.

B. EXPERIMENTAL RESULTS

In this paper, the localization results of the algorithm designed in this paper are measured by two indexes: localization accuracy and average localization error.

1) Localization accuracy
The localization accuracy is defined as the ratio of the number of correctly classified samples to the total number of test samples.

$$\text{loaccy} = \frac{\text{num}_{\text{corr}}}{\text{mum}_{\text{total}}}, \tag{12}$$

where num_{corr} represents the number of correctly classified test samples and num_{total} represents the total number of test samples.

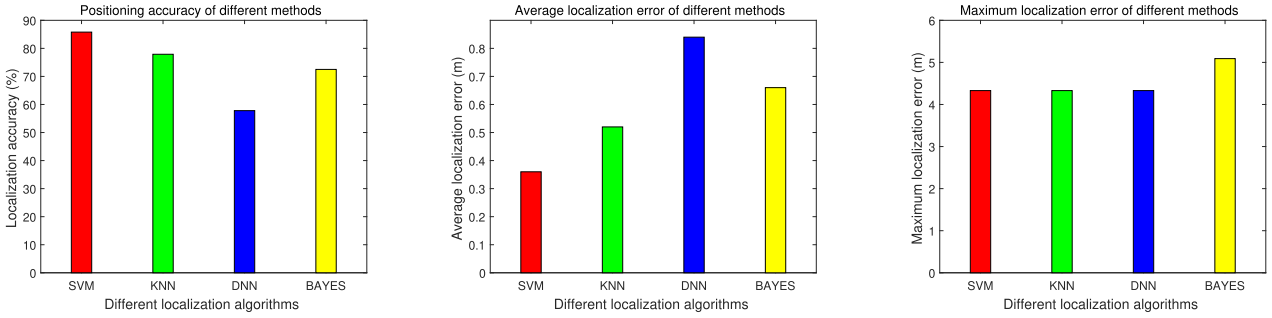
2) Average localization error

$$avgerr = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i - x_i')^2 + (y_i - y_i')^2}, \tag{13}$$

where x_i and y_i represent the actual coordinate of the i-th test sample, x_i' and y_i' represent the position estimate coordinate of the i-th test sample, and n is the total number of test samples.

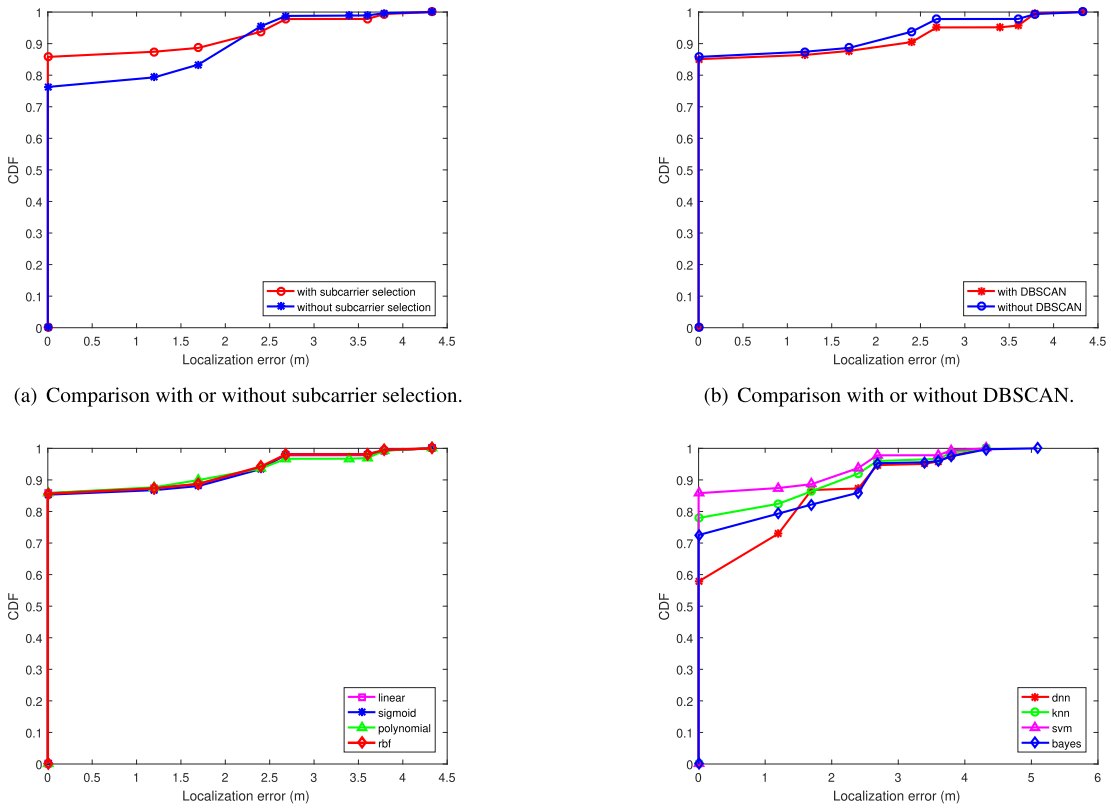
We evaluate the performance of the algorithm designed in this paper in the corridor and laboratory. The confusion matrices of the corresponding test results are shown in Fig. 7 and Fig. 8, respectively.

When experimenting in the corridor, the localization accuracy is about 91% and the average localization error is about 0.30 m. The prediction result of each test position is basically accurate. When experimenting in the laboratory,



(a) Localization accuracy. (b) Average localization error. (c) Maximum localization error.

FIGURE 10. Localization effect of different localization algorithms.



(a) Comparison with or without subcarrier selection. (b) Comparison with or without DBSCAN.

(c) Comparison of different kernel functions. (d) Comparison of different localization algorithms.

FIGURE 11. Comparison of localization effects of different methods.

the localization accuracy is about 86% and the average localization error is about 0.36m. However, the prediction results of position 1 and position 3 are not good, probably because their positions are far from the router and the received signal is weak. The cumulative distribution function(CDF) of localization errors in the corridor and laboratory are shown in Fig. 9.

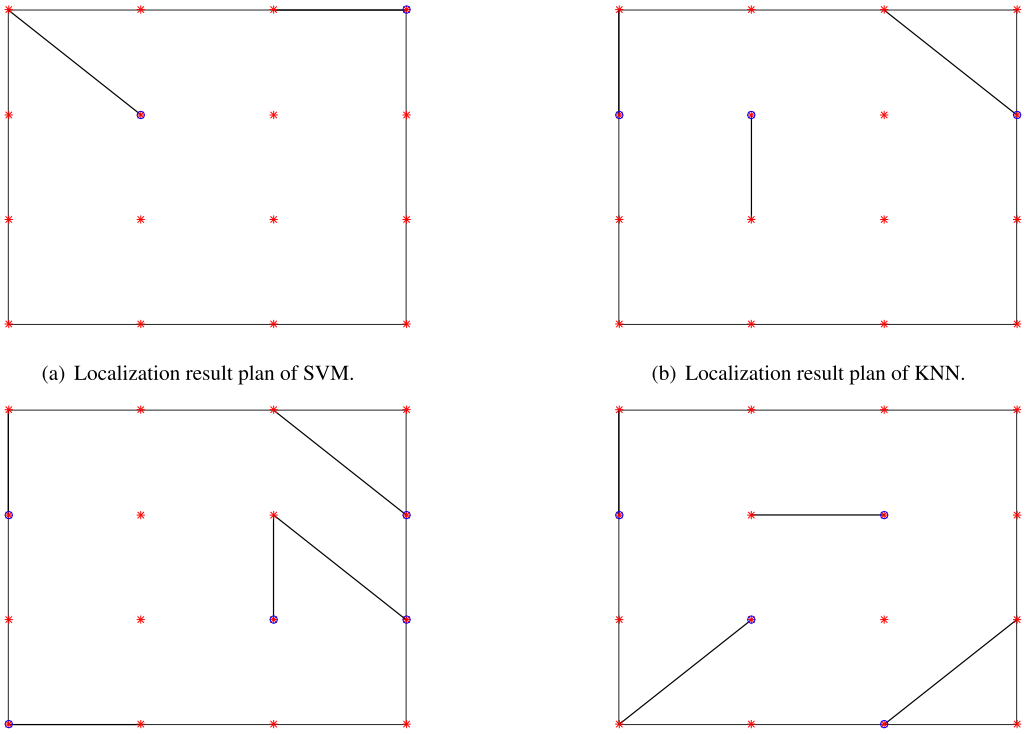
C. COMPARISON AND ANALYSIS

This paper conducts experimental tests in the laboratory to evaluate the effects of the different methods used in this paper.

1) Impact of the optimal subcarrier selection method: The optimal subcarrier selection method is applied to the CSI data, which can remove unstable subcarriers and select optimal subcarrier information, thereby improving localization accuracy. The evaluation results with and without the optimal subcarrier selection method are shown in Table 1. The cumulative error distribution function of localization errors is shown in Fig. 11(a). Compared with not using this method, the localization accuracy after use is improved by 9.5% and the average localization error is reduced by 0.17m. Obviously, the

TABLE 1. Localization effect of different methods.

	Method	Localization accuracy	Average error (m)	Maximum error (m)
Optimal subcarrier selection method	With	85.8	0.36	4.33
	Without	76.3	0.53	4.33
DBSCAN method	With	85.8	0.36	4.33
	Without	85.1	0.42	4.33
Kernel function	RBF	85.8	0.36	4.33
	Linear	85.6	0.36	4.33
	Sigmoid	85.3	0.37	4.33
	Polynomial	85.6	0.36	4.33



(a) Localization result plan of SVM.

(b) Localization result plan of KNN.

(c) Localization result plan of DNN.

(d) Localization result plan of BAYES.

FIGURE 12. Localization result plans of different methods.

effect after the optimal subcarrier selection method outperforms the unprocessed effect.

2) Impact of the abnormal sample point removal method: DBSCAN clustering method can be used for CSI data to remove abnormal sample points. The evaluation results with and without DBSCAN method are shown in Table 1. The cumulative error distribution function of localization errors is shown in Fig. 11(b). It can be found that the localization result using the DBSCAN method is better.
3) Impact of Kernel function selection: This paper compares the influence of some common kernel functions on the localization accuracy. The comparison results are shown in Table 1. The localization accuracy of the kernel function for RBF is slightly higher than other kernel functions. The cumulative distribution function of the localization error is shown in Fig. 11(c). It can be seen that the four kernel functions have little impact on the final localization accuracy.
4) Impact of different localization algorithms: This paper uses SVM multi-classification algorithm to achieve localization and other common localization algorithms include DNN, KNN and Bayesian classification algorithm, and so forth. For comparison, this paper selects the KNN, DNN and BAYES algorithm for the localization effect test. The comparison results are shown in Fig. 10(a), (b) and (c), and the cumulative error distribution function of the localization error is

shown in Fig. 11(d). It can be seen that the SVM multi-classification algorithm is more suitable than other algorithms in terms of localization problems.The localization result plans are shown in Fig. 12, where the line is the connection between the real position and the estimated position. Obviously, the algorithm in this paper performs better than other localization algorithms.

V. CONCLUSION

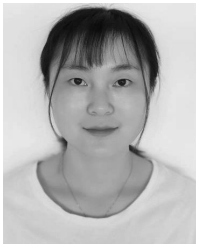
Different from existing indoor localization methods, which can only obtain CSI by the computer equipped with Intel 5300 NIC, this paper uses the smartphone instead of the computer to collect CSI and proposes the smartphone-based indoor fingerprinting localization using channel state information. However, compared with the CSI collected by the computer, the CSI signal collected by the smartphone fluctuates greatly, which is also an inherent problem of the CSI collected by the smartphone. In order to solve this problem, the CSI signal is processed first, and the optimal subcarrier selection and abnormal sample point removal methods are designed to obtain relatively stable and effective data, and the SVM multi-classification method is applied to achieve indoor localization. The high efficiency of the algorithm is proved by a large number of experiments. We conducted experiments in two typical indoor environments, the corridor and laboratory. The localization accuracies in the corridor and laboratory are 91% and 86%, respectively, and both average localization errors are less than $0.5m$. The smartphone that obtains CSI in this paper is a specific smartphone, namely Google Nexus 5 smartphone. At present, it is impossible to obtain CSI through any type of smartphone. In the future work, we can try to study the firmware of any smartphone to achieve CSI collection using any smartphone. Nowadays, the existing technology is better for achieving single-person localization, and the performance of multi-person localization is not good. In order to cope with this challenge, designing more efficient algorithms to solve multi-person localization problems is also the focus of future work.

REFERENCES

[1] X. Liu, M. Zhao, A. Liu, and K. K. L. Wong, "Adjusting forwarder nodes and duty cycle using packet aggregation routing for body sensor networks," *Inf. Fusion*, vol. 53, pp. 183–195, Jan. 2020.
[2] X. Liu, A. Liu, T. Qiu, B. Dai, T. Wang, and L. Yang, "Restoring connectivity of damaged sensor networks for long-term survival in hostile environments," *IEEE Internet Things J.*, to be published.
[3] H. Jue, L. Zhuo, L. Dianjie, and L. Sanglu, "Sleeping schedule-aware local broadcast in wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 2013, pp. 1–10, 2013, Art. no. 451970.
[4] M. Wu, S. He, Y. Zhang, J. Chen, Y. Sun, Y.-Y. Liu, J. Zhang, and H. V. Poor, "A tensor-based framework for studying eigenvector multicentrality in multilayer networks," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 31, pp. 15407–15413, 2019
[5] S. Shi, S. Sigg, and Y. Ji, "Probabilistic fingerprinting based passive device-free localization from channel state information," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.
[6] X. Huang, S. Guo, W. Yan, and Y. Yang, "A fine-grained indoor fingerprinting localization based on magnetic field strength and channel state information," *Pervasive Mobile Comput.*, vol. 41, pp. 150–165, Oct. 2017.
[7] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2210–2218.
[8] K. Whitehouse, C. Karlof, and D. E. Culler, "A practical evaluation of radio signal strength for ranging-based localization," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 11, no. 1, p. 41, 2007.
[9] S. Lanzisera, D. T. Lin, and K. S. J. Pister, "RF time of flight ranging for wireless sensor network localization," in *Proc. 4th Workshop Intell. Solutions Embedded Syst.*, Jun. 2006, pp. 1–12.
[10] S. Tomic, M. Marikj, M. Beko, R. Dinis, and N. Órfão, "Hybrid RSS-AoA technique for 3-D node localization in wireless sensor networks," in *Proc. Int. Wireless Commun. Mobile Comput. Conf.*, Aug. 2015, pp. 1277–1282.
[11] S. L. Yang, J. W. Park, and L. Barolli, "A localization algorithm based on AOA for ad-hoc sensor networks," *Mobile Inf. Syst.*, vol. 8, no. 1, pp. 61–72, 2012.
[12] Q. Li, A. Liu, and T. Wang, "Pipeline slot based fast rerouting scheme for delay optimization in duty cycle based M2M communications," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 6, pp. 1673–1704, 2019.
[13] Y. Liu, A. Liu, X. Liu, and M. Ma, "A trust-based active detection for cyber-physical security in industrial environments," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6593–6603, Dec. 2019.
[14] Y. Hu and G. Leus, "Robust differential received signal strength-based localization," *IEEE Trans. Signal Process.*, vol. 65, no. 12, pp. 3261–3276, Jun. 2017.
[15] C. Liang and F. Wen, "Received signal strength-based robust cooperative localization with dynamic path loss model," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1265–1270, Mar. 2016.
[16] Y. Sun, F. Tong, Z. Zhang, and S. He, "Throughput modeling and analysis of random access in narrowband Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1485–1493, Jun. 2018.
[17] P. Xia, S. Zhou, and G. B. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 202–213, Jan. 2004.
[18] Y. Xu, P. Chen, S. Gao, and N. Qiang, "CSI-based low-duty-cycle wireless multimedia sensor network for security monitoring," *Electron. Lett.*, vol. 54, no. 5, pp. 323–324, 2018.
[19] J. Wang, H. Jiang, J. Xiong, K. Jamieson, and B. Xie, "LIFS: Low human effort, device-free localization with fine-grained subcarrier information," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 243–256.
[20] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based Indoor Localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1300–1309, Jul. 2013.
[21] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.
[22] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *Proc. 21st Int. Conf. Comput. Commun. Netw.*, Jul./Aug. 2012, pp. 1–7.
[23] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2015, pp. 1666–1671.
[24] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: Indoor Wi-Fi fingerprinting system," in *Proc. 39th Annu. IEEE Conf. Local Comput. Netw.*, Sep. 2014, pp. 202–209.
[25] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2017, p. 6.
[26] O. Mugahid and T. G. Yun, "Indoor distance estimation for passive UHF RFID tag based on RSSI and RCS," *Measurement*, vol. 127, pp. 425–430, Oct. 2018.
[27] M. Schulz, D. Wegemer, and M. Hollick, "Nexmon: Build your own Wi-Fi testbeds with low-level MAC and PHY-access using firmware patches on off-the-shelf mobile devices," in *Proc. 11th Workshop Wireless Netw. Testbeds, Exp. Eval. Characterization*, 2017, pp. 59–66.
[28] M. Schulz, J. Link, F. Gringoli, and M. Hollick, "Shadow Wi-Fi: Teaching smartphones to transmit raw signals and to extract channel state information to implement practical covert channels over Wi-Fi," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2018, pp. 256–268.
[29] M. Schulz, D. Wegemer, and M. Hollick, "The nexmon firmware analysis and modification framework: Empowering researchers to enhance Wi-Fi devices," *Comput. Commun.*, vol. 129, pp. 269–285, Sep. 2018.
[30] R. Liu, H. Wang, and X. Yu, "Shared-nearest-neighbor-based clustering by fast search and find of density peaks," *Inf. Sci.*, vol. 450, pp. 200–226, Jun. 2018.

[31] H. Zhang and L. Cao, "A spectral clustering based ensemble pruning approach," *Neurocomputing*, vol. 139, no. 139, pp. 289–297, 2014.
[32] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[33] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-free presence detection and localization with SVM and CSI fingerprinting," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7990–7999, Dec. 2017.
[34] R. Zhou, J. Chen, X. Lu, and J. Wu, "CSI fingerprinting with SVM regression to achieve device-free passive localization," in *Proc. IEEE 18th Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2017, pp. 1–9.
[35] Y. Leng, C. Sun, X. Xu, Q. Yuan, S. Xing, H. Wan, J. Wang, and D. Li, "Employing unlabeled data to improve the classification performance of SVM, and its application in audio event classification," *Knowl.-Based Syst.*, vol. 98, no. C, pp. 117–129, 2016.
[36] Z. Sun, K. Hu, T. Hu, J. Liu, and K. Zhu, "Fast multi-label low-rank linearized SVM classification algorithm based on approximate extreme points," *IEEE Access*, vol. 6, pp. 42319–42326, 2018.

PENGPENG CHEN was born in China, in 1983. He received the Ph.D. degree from the Department of Computer Science and Technology, Ocean University of China, Qingdao, China, in 2011. He is currently a Professor with the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China. He has published more than 20 academic articles in important academic journals at home and abroad. His main research interests are sensor networks, the Internet of Things, and data modeling.



FEN LIU was born in China, in 1995. She is currently pursuing the M.D. degree with the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, with a focus on wireless sensor networks.

SHOUWAN GAO was born in China, in 1982. She received the Ph.D. degree from the Department of Computer Science and Technology, Ocean University of China, Qingdao, China. She is currently an Associate Professor with the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China. Her research interests include networked control systems and networked estimation.

PEIHAO LI was born in China, in 1994. He is currently pursuing the M.D. degree with the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, with a focus on distributed measurement systems and sensor networks.

XU YANG was born in China, in 1995. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, with a focus on distributed measurement systems, big data, and sensor networks.

QIANG NIU was born in China, in 1974. He received the Ph.D. degree from the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China. He is currently a Professor and the Associate Dean with the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou. His main research interests are intelligent information processing, artificial intelligence and pattern recognition, machine learning, and data mining.