Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8546237, 14 pages
https://doi.org/10.1155/2020/8546237

Research Article

Air Gesture Recognition Using WLAN Physical Layer Information

Xiaochao Dang,[1,2] **Yang Liu**,[1] **Zhanjun Hao**,[1,2] **Xuhao Tang**,[1] **and Chenguang Shao**[1]

[1]*College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China*
[2]*Gansu Internet of Things Engineering Research Center, Lanzhou 730070, China*

Correspondence should be addressed to Zhanjun Hao; haozhj@nwnu.edu.cn

Received 7 November 2019; Revised 16 May 2020; Accepted 29 June 2020; Published 13 August 2020

Academic Editor: Luca Reggiani

In recent years, the researchers have witnessed the important role of air gesture recognition in human-computer interactive (HCI), smart home, and virtual reality (VR). The traditional air gesture recognition method mainly depends on external equipment (such as special sensors and cameras) whose costs are high and also with a limited application scene. In this paper, we attempt to utilize channel state information (CSI) derived from a WLAN physical layer, a Wi-Fibased air gesture recognition system, namely, WiNum, which solves the problems of users' privacy and energy consumption compared with the approaches using wearable sensors and depth cameras. In the process of recognizing the WiNum method, the collected raw data of CSI should be screened, among which can reflect the gesture motion. Meanwhile, the screened data should be preprocessed by noise reduction and linear transformation. After preprocessing, the joint of amplitude information and phase information is extracted, to match and recognize different air gestures by using the S-DTW algorithm which combines dynamic time warping algorithm (DTW) and support vector machine (SVM) properties. Comprehensive experiments demonstrate that under two different indoor scenes, WiNum can achieve higher recognition accuracy for air number gestures; the average recognition accuracy of each motion reached more than 93%, in order to achieve effective recognition of air gestures.

1. Introduction

With the continuous progress of science and technology, human-computer interaction technology has been developing rapidly. The flourish on the Internet of Things (IoT) and Artificial Intelligence (AI) has boosted human-machine interaction technology based on human gestures to become a hot topic in academia and industry. Through the close integration of the emerging technology, human-computer interaction technology has gradually changed from the earliest mouse and keyboard to now touch screen, voice, and gesture control to be intelligent, friendly, and highly adaptable [1]. Gesture recognition has strong flexibility and environmental adaptability and can be widely used in a variety of scenarios [2]. Traditionally, the primary function of the Wi-Fi signal is to support high-throughput data communication between the terminal equipment and the Internet. However, through the continuous exploration of the Wi-Fi signal, it is found that a new technology based on the Wi-Fi signal is attracting more and more attention in academic circles. The CSI [3], in Wi-Fi signal, can be used as a wireless channel measurement index which describes how the signal propagates from the transmitter to the receiver [4], reflecting the signal scattering, environmental change, power attenuation, and other influencing factors on each transmission path [5]. As the WLAN physical layer (PHY) information, CSI can be used in the fields of indoor localization [6], trajectory tracking [7, 8], gesture recognition [9–11], keystroke detection [12], driver activity [13], and lip-reading service [14], which is easy to implement and very sensitive to the changes of indoor environment, as well as in low cost.

Behavior awareness technology based on the Wi-Fi signal has become an essential research direction in the field of HCI [15]. The existing gesture recognition technologies are mainly divided into device gesture recognition and device-free gesture recognition [16]. The advantage of having device gesture recognition technology is that it can directly obtain the movement information of gesture motion track and hand

joint, and its recognition accuracy is high [17], while the user must carry the measuring equipment [18, 19], which will influence the experience of the user. It neither can make the user carry on the human-computer interaction more natural nor can it embody the human-computer interaction design. In order to recognize gestures accurately and make users get a better experience, a large number of researchers began to pay attention to the device-free gesture recognition technology without any measuring equipment. The advantages of Wi-Fi wireless sensing technology in passive gesture recognition are obvious, especially for its low cost, easy to obtain, suitable for the user habits, etc., which has become the focus of research at home and abroad [20, 21].

The process of air gesture recognition can be classified into three steps: the extraction of motion signal, the preprocessing of effective signal, and motion classification and matching, among which the stage of classification and matching is particularly important [22, 23]. The traditional gesture classification algorithm uses a single matching algorithm to seek the best matching for the processed gesture data, which not only leads to high calculation complexity but also has limited recognition precision. However, in this paper, we propose a new algorithm to recognize gesture data, which includes the classification stage and recognition stage. The improved algorithm enhances the recognition of precision and shortens the matching time. The S-DTW algorithm [24, 25], in accordance with the properties of SVM and DTW algorithm, the requirements of gesture recognition in the classification stage, after extracting the effective gesture data, combined the DTW algorithm into the kernel function of SVM so as to realize the classification and matching of gesture data.

In this paper, the main contribution of our work is that we build a gesture recognition system called WiNum. Details are as follows:

(1) The phase information about CSI, as the auxiliary information, combined with the amplitude information of CSI, which improves the utilization rate of CSI information and also finds a slight change of air gesture

(2) It can be found that the gesture motion has influenced CSI signal, and the differences of subcarrier sensitivity of the CSI signal is also proved

(3) With the help of an effective noise reduction filtering algorithm, the amplitude and phase of CSI signal can be processed, and the detailed fine-granularity CSI signal can be used to represent the different meanings expressed by each gesture motion

(4) Combined with the properties of SVM with DTW algorithms, the S-DTW gesture matching algorithm is obtained, which can recognize air gesture motion quickly and effectively

The rest of the paper is as follows: In Section 2, we introduce the properties of CSI signal and compare the advantages and disadvantages of the existing gesture recognition methods. Then, we propose a gesture recognition method and discuss the key parts of the system specified in Section 3. In addition, we describe the experiment and performance analysis that results in Section 4 and finally summarize the proposed method results in Section 5.

2. Related Work

In this paper, the used Wi-Fi signal data are obtained from the Intel 5300 network card. Orthogonal Frequency Division Multiplexing (OFDM) technology is used for modulating the signal [26]. The transmission channel response can be extracted in the format of CSI, and CSI is WLAN physical layer information [27], which is used to estimate the channel characteristics in the communication link and is helpful to analyze the signal propagation in the process of gesture recognition [28]. Using T_x, R_x, and N_s to represent the number of antennas transmitted and received and the number of OFDM subcarriers transmitted in a basic model of channel transmission, the OFDM system can be modeled in the frequency domain as follows:

$$Y_j = H_j X_j + N_j, \quad j \in [1, N_s], \tag{1}$$

where Y_j and X_j denote the signal vectors of the receiver and the transmitter, respectively, H_j denotes the channel information matrix, and N_j denotes the white Gaussian noise. According to the formula (1), $N_S = 30$, which $\widehat{H}$ is expressed:

$$\widehat{H} = \frac{Y}{X}, \tag{2}$$

where $\widehat{H}$ is the channel frequency response (CFR) of a wireless channel; it can express the variations of the Wi-Fi channel. Each packet at the receiving end using the Intel 5300 wireless card contains a set of measurements for the CSI:

$$H(k) = |H(k)|e^{j\angle H(k)}, \tag{3}$$

where $|H(k)|$ and $j\angle H(k)$ denote the amplitude and phase, respectively [29, 30]. In the indoor environment, CSI can still maintain the overall structural stability by using its own characteristics, which is more conducive to the subsequent analysis and extraction of air gesture features.

The gesture is a visual body language with a strong visual effect, which is easy to understand and contains abundant information. The gesture motions are aimed at conveying relevant information, which is the "second language" in people's daily life [31, 32]. Nowadays, perfecting Wi-Fi infrastructure makes Wi-Fi signals almost everywhere. Through the study on developing conditions of the motion-sensing system at home and abroad, it is found that the majority of the existing Wi-Fi signal sensing systems only use the change of CSI amplitude to distinguish different behavioral activities [33, 34]. Using amplitude as a system to measure information not only wastes the phase information that CSI can provide but also the use of a single measurement signal may limit the improvement of system recognition accuracy. Considering phase as an auxiliary signal not makes full use of CSI information and improves the utilization rate of information, but the phase information is more sensitive to motion

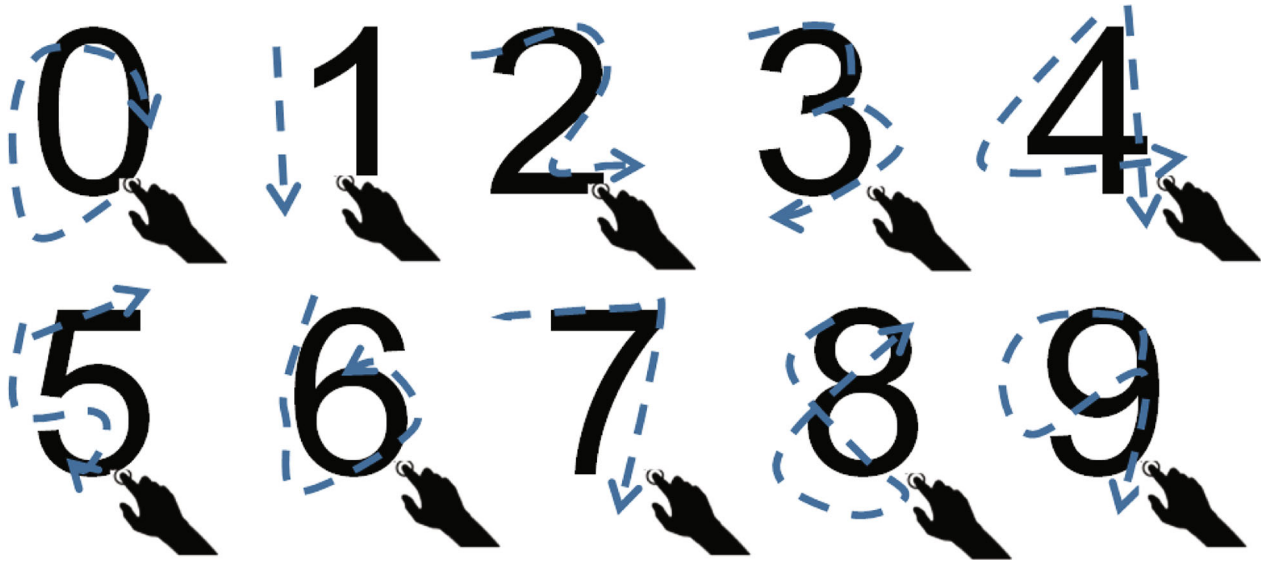

Figure 1: Air gesture motion.

changes in different directions; it can capture small motion changes, which is very vital to deal with the applications related to the recognition of dynamic motion. In 2013, [35] introduced WiSee, a Doppler frequency shift method by using wireless signals to realize the perception and recognition of gestures. In 2015, the author proposed WiGest in [36]. The method analyzes the change of Received Signal Strength Indication (RSSI) received by the Wi-Fi signal to sense the user's air gesture. For a single access point and three access points, the recognition accuracy of WiGest is 0.87 and 0.96. Compared with RSSI, fine-granularity CSI is more suitable for gesture recognition. By analyzing the CSI fluctuations caused by the gesture, in [37], the author proposed WiGeR, a method with an average recognition accuracy of 0.92 among different gestures in five scenarios. In [38], WiFinger is proposed to extract gesture patterns by principal component analysis, and the model is used as a feature to recognize gestures; its accuracy is 0.92.

Although there are many kinds of research on human behavior perception through CSI in the field of wireless sensor networks, many problems are still needed to be solved among detection methods at present. In view of the above cases, this paper uses fine-granularity CSI to recognize the air gesture, which combined the amplitude information with phase information about CSI as a new sensing method of sending information, which can recognize air gestures efficiently and quickly without wearing any sensors, and work in line-of-sight (LOS) and non-line-of-sight (NLOS) environments. In this paper, a motion gesture is used to recognize any number in aerial handwritten 0-9, each gesture represents a different meaning, and the specific gesture motion is shown in Figure 1.

3. WiNum Design

3.1. System Overview. By using the WiNum method, the recognition process first collects the raw data packets of CSI, then selects the collected data; the preprocessing is applied to the data which can reflect the gesture motion, extracting the feature information. To establish the joint information fingerprint database of amplitude and phase, classifying different gesture motions and establishing the air gesture model finally come to the recognition of the dynamic gesture.

The preprocessing is divided into the amplitude processing and phase processing of the CSI, the amplitude is denoised by a wavelet, and the phase is unwrapped and corrected, as well as linearly changed. To extract the effective gesture image by gesture signal preprocessing, then the effective data is obtained and the data features are extracted to find the corresponding data of different gestures, finally, with the help of the SVM algorithm of machine learning in the offline so as to train the gesture clustering model of the same number stage, while in the online stage, by using the DTW algorithm and taking out the trained model to recognize the gesture. In this paper, we have proposed the recognition flow chart of an aerial handwritten number, as is shown in Figure 2; it is mainly divided into the following four main stages to study: (1) the selection of dynamic gesture data, (2) the preprocessing of the selected data, (3) feature matching of dynamic gesture data, and (4) acquisition of the recognition results of gesture motions.

Because the CSI signal is vulnerable to the multipath effect, signal attenuation, and other interfering factors, the directly obtained gesture data contains various interference data, which cannot be directly used for extracting signal features, so firstly, it is necessary to preprocess the data; its aim is to remove the noise data from gesture data. Amplitude and phase information of CSI are preprocessed to obtain the data that is needed for the extraction of gesture data feature in the later stage.

3.2. Amplitude Sanitization. The process of preprocessing for the amplitude information, which firstly includes the selection of the subcarriers and the raw data of the selected subcarrier CSI amplitude, should be processed; the raw data are subjected to noise reduction and smoothing by using

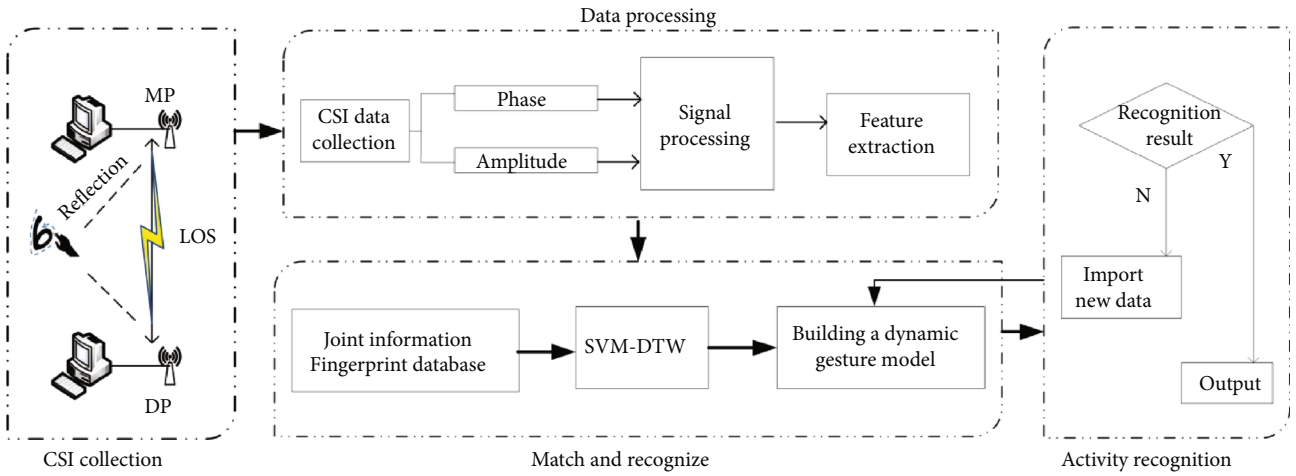

Figure 2: Structure diagram of system WiNum.

the wavelet threshold value in the course of processing, so as to show the local features of the change of each gesture corresponding to the subcarrier more clearly. Through theoretical analysis, it is found that the larger the variance of subcarrier amplitude in the same group data stream, the more sensitive it is to the change of environment. The simple CSI amplitude information cannot well reflect the characteristics of each gesture motion and can only represent the changes of the whole sequence caused by different gesture motion, and the changes and CSI values of each subcarrier are also different. Therefore, the variance is selected as the eigenvalue. For the CSI sequence corresponding to any motion, the distribution of the variance in each selected time slot can reflect the degree of dispersion of the CSI amplitude, and at the same time, it can also reflect the change of the corresponding gesture motion; hence, it is reasonable and effective to extract the features of each motion by calculating the variance of the CSI value [39]. As is shown in Figure 3, data stream 2 is more sensitive to the change of the environment than that of data stream 1; meanwhile, the variance of the former is larger than the latter. Therefore, the variance is selected as eigenvalue, and to the CSI sequences that are corresponding to any motion, the distribution of the variance can reflect the dispersion degree of the CSI amplitude in each of the selected time slots and the change of the corresponding gesture motion, serving as the features of each motion.

Therefore, it is reasonable and effective to extract the features of each motion by calculating the variance of the CSI value. For each gesture, CSI amplitude streams are obtained from the 30 subcarriers of the same data stream. Assuming that the number of samples is m, the matrix used for storing the CSI amplitude stream includes rows and 30 columns, calculating the variance of each column of the matrix, filtering out the subcarriers of the small variance, and selecting the subcarriers of the maximum variance as the selected raw data. Selecting the No. 27 subcarrier with the biggest variance, this is shown in Figure 4.

After selecting the subcarrier signal which can depict the change of gesture motion, due to the features of the CSI signal, the noise of the subcarrier signal will cause serious interference to the quality of the signal, which will directly affect the process of motion detection, feature extraction, and so on, leading to the deviation or even error for the later gesture recognition results, so it is requisite to further reduce the noise and smooth the selected signal. The amplitude of the selected subcarrier data should be denoised by the wavelet threshold, $s(i) = f(i) + e(i)\ (i = 1, 2, \cdots, n-1)$, in which the raw signal is represented by $f(i)$, the noise signal is represented by $e(i)$, and the raw signal with noises is represented by $s(i)$ and transformed by a discrete wavelet transform.

$$\int S(i)\Psi_{j,k}(t)dt = \int f(i)\Psi_{j,k}(t)dt + \int \sigma e(i)\Psi_{j,k}(t)dt, \quad (4)$$

where $\Psi_{j,k}(t)$ is the discrete wavelet primary function; the raw signal can be represented as

$$S_{j,k} = F_{j,k} + E_{j,k}, \quad (5)$$

where $S_{j,k}$ is the signal with noise, $s(i)$ is the wavelet coefficients of each layer after wavelet transforms, $F_{j,k}$ is the wavelet transform coefficients of the raw signal $f(i)$, $E_{j,k}$ is the noise signal, and $e(i)$ is the wavelet transform coefficients of the noise signal $E_{j,k}$. According to the statistical characteristics of the useful signal and noise wavelet coefficient, the soft threshold function is used as follows:

$$S'_{j,k} = \begin{cases} \operatorname{sgn}\left(S_{j,k}\right)\left(\left|S_{j,k}\right| - \lambda\right), & \left|S_{j,k}\right| \geq \lambda, \\ 0, & \left|S_{j,k}\right| < \lambda, \end{cases} \quad (6)$$

where λ is the general threshold $\lambda = \sigma\sqrt{2InN}$, N is the signal length, and σ is the noise standard deviation after the wavelet threshold function is processed, compared with the raw data effect; the effect of the mutation data is eliminated, a relatively smooth data curve is fitted, and after the CSI data processing, the complete CSI gesture graph shown in Figure 5 is obtained, which lays a foundation for the later-period WiNum feature extraction of amplitude information.

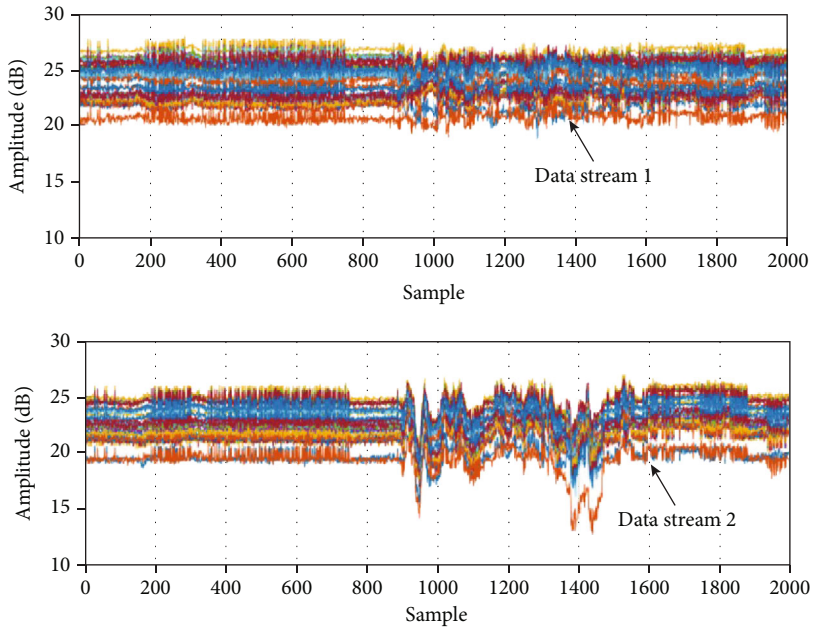

Figure 3: Selection of the data.

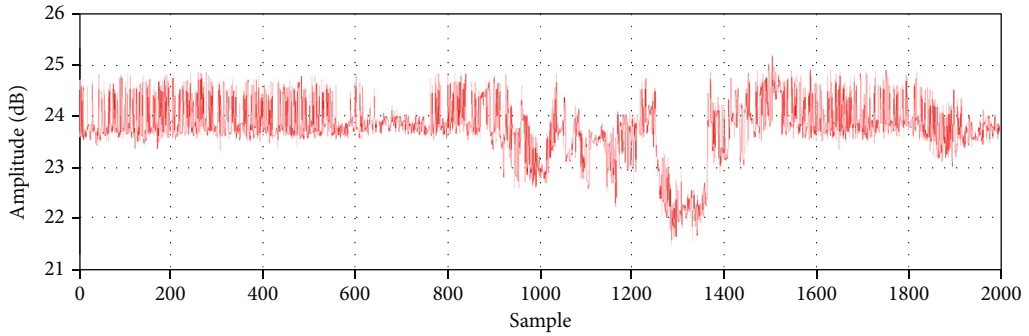

Figure 4: Raw CSI amplitude.

3.3. Phase Sanitization. Combined features of CSI phase information with the analysis of the properties of the 802.11n protocol [40], it can be seen that Figure 6 shows the collected raw CSI phase information; the existence of clock synchronization and random noise in the raw CSI phase information is not available for any kind of detection. By using the symmetry of the center frequency of the 802.11n communication protocol, it is clear that the phase information can be used to realize the perception of human motion after the linear change of the raw phase information [41, 42].

The phase of the i subcarrier of the measured CSI signal is ϕ_i, then

$$\widehat{\Phi_i} = \phi_i - 2\pi \frac{k_i}{N}\delta + \beta + Z. \tag{7}$$

Among them, ϕ_i is the real phase, δ is the time offset between the receiver and the transmitter, which is the main factor causing the phase error, β is the unknown phase offset, Z is the noise introduced in the measurement process, k_i is the subcarrier index of the i subcarrier, respectively, the subcarrier index of the 30 subcarriers is -28, -26, -24, •••, -4, -2, -1, 1, 3, 5, •••, 25, 27, 28, and N is the number of FFT points.

In order to eliminate the influence of δ and β, two variables Δ and ∇ are defined.

$$\begin{aligned} \Delta &= \frac{\widehat{\phi_n} - \widehat{\phi_i}}{k_n - k_1} = \frac{\phi_n - \phi_1}{k_n - k_1} - \frac{2\pi}{N}\delta, \\ \nabla &= \frac{1}{n}\sum_{j=1}^{n}\widehat{\phi_j} = \frac{1}{n}\sum_{j=1}^{n}\phi_j - \frac{2\pi}{Nn}\delta\sum_{j=1}^{n}k_j + \beta. \end{aligned} \tag{8}$$

Assuming that the frequency of the subcarrier is completely symmetric, that is, if there is $\sum_{j=1}^{n}k_j = 0$, then ∇ can be expressed as

$$\nabla = \frac{1}{n}\sum_{j=1}^{n}\phi_j + \beta. \tag{9}$$

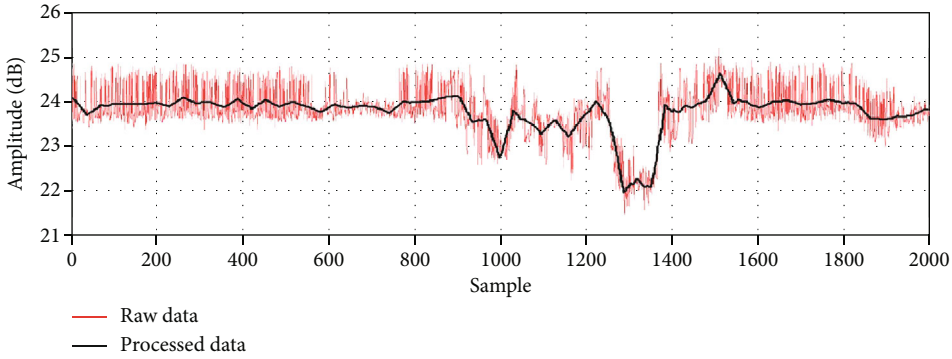

Figure 5: Processed CSI amplitude.

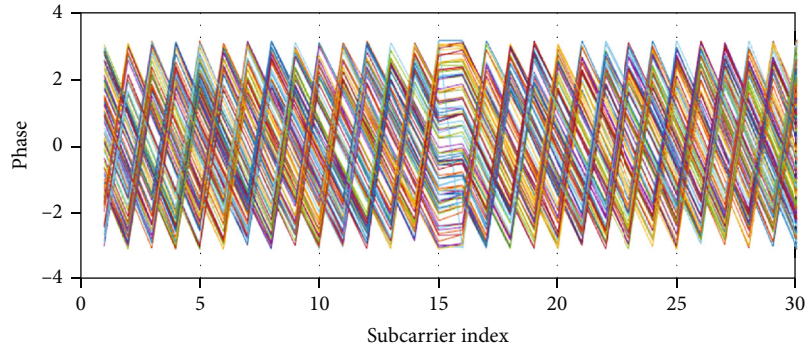

Figure 6: Original CSI phase.

The linear variable $\Delta k_i - \nabla$ is subtracted from the measured phase $\widehat{\phi_i}$, and the linear combination of the real phase to remove the random phase offset is obtained.

$$\widehat{\Phi_i} = \widehat{\phi_i} - \Delta k_i - \nabla = \phi_i - \frac{\widehat{\phi_n} - \widehat{\phi_i}}{k_n - k_1} k_i - \frac{1}{n}\sum_{j=1}^{n} \phi_j. \tag{10}$$

It can be found that the phase signal no longer contains the error term of random noise. Although the phase obtained after linear calibration is not the real CSI phase, the linear transformation value of its real phase, it is clear that the variance of the phase before and after calibration satisfies a certain mathematical relationship. Assuming that ϕ_I about frequency is independent and the same distribution, it should be

$$\Sigma_{\phi_i \wedge}{}^2 = c_i \sigma_{\phi_i}{}^2, \quad c_i = 1 + \frac{k_i^2}{(k_n - k_1)^2} + \frac{1}{n}. \tag{11}$$

There is only one constant multiplier of frequency c_I between the calibrated phase variance and the real phase variance. That is to say, the changing trend of the calibrated phase signal can be used to reflect the fluctuation of the real phase, which theoretically solves the problem that the phase cannot be used because of the random distribution of the real phase. In order to prove the effect of actual phase processing, the CSI phase of the first time of the hand potential signal is processed by using the linear calibration algorithm, and the results of the CSI phase corresponding to different subcarriers after processing is shown in Figure 7, the phase distribution of the calibrated CSI has a strong regularity, and the distribution of the phase processing value, which is removed from random noise and other factors, is no longer too random, and it is possible to distinguish different motions. Similar conclusions can be obtained from repeated testing of data packets at other times.

3.4. Feature Extraction. In the process of studying air gesture recognition, feature extraction of the gesture data is a particularly important link. The amplitude and phase of the signal are calculated after the gesture data are collected, but a unified feature measure is still needed. The information of the raw data cannot be directly recognized by the classifier, so it is necessary to extract and select the features that can best represent the gesture from the raw gesture data. While gesture motion exists, the difference between amplitude and phase is significantly larger than that in the static case. Figure 8 further illustrates the difference of influence on a subcarrier in the presence of gesture motion and stillness. The difference between amplitude and phase will be two good indicators of gesture motion. Variance cannot be directly used as a feature of detection, because it is related to signal power, so it cannot be extended to different scenarios in different link states. The preprocessed amplitude and phase are used as the input of gesture recognition, and the features are

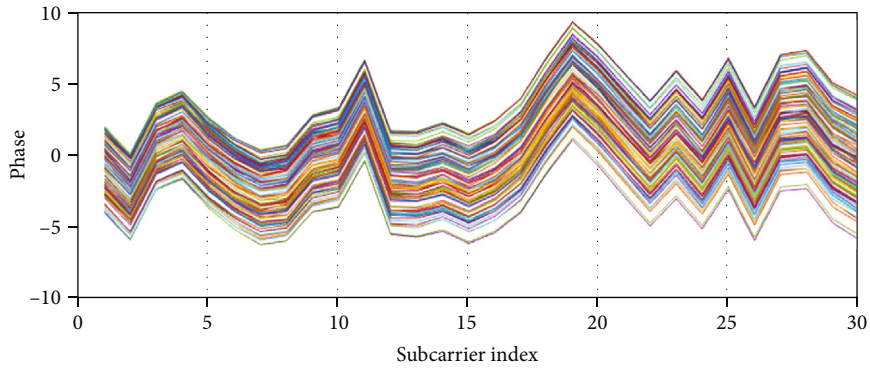

Figure 7: Processed CSI phase.

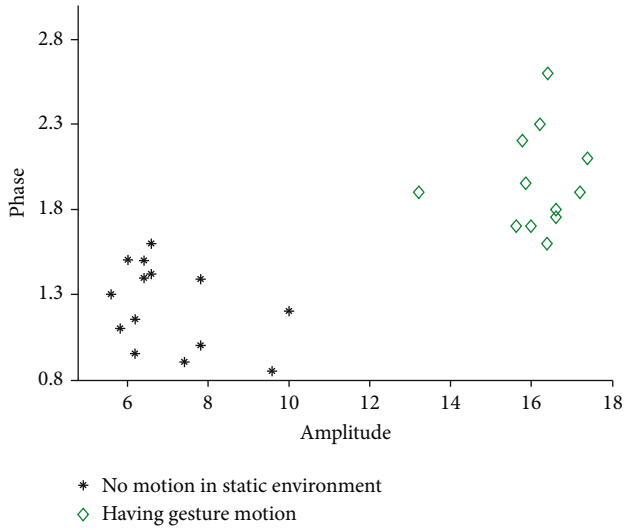

Figure 8: Comparison of feature extraction under different states.

extracted from the covariance matrix. The $|\overline{\mathbb{H}}|$ and $\overline{\phi}$ are represented as amplitude and phase sequences, and then, their covariance matrices are corresponding to each other.

$$\begin{aligned}\sum|\overline{\mathbb{H}}| &= \left[\mathrm{COV}\left(\bar{H}_I, \bar{H}_j\right)\right]_{k\times k},\\ \sum\overline{\phi} &= \left[\mathrm{COV}\left(\overline{\phi}_I, \overline{\phi}_j\right)\right]_{k\times k}.\end{aligned} \tag{12}$$

In the above two matrices, the lower covariance represents the stillness or the case that no motion and the higher covariance can indicate the occurrence of gesture motion, and the covariance of different gesture motion is different. In order to extract the features for further detection, the eigenvalues of the two matrices are calculated, and the maximum eigenvalues of each matrix are selected. Finally, a binary group $F = [\alpha, \beta]$ of joint information features are formed.

$$\begin{aligned}\alpha &= \max\left(\mathrm{eigen}\left(\sum|\overline{\mathbb{H}}|\right)\right),\\ \beta &= \max\left(\mathrm{eigen}\left(\sum\overline{\phi}\right)\right).\end{aligned} \tag{13}$$

3.5. Motion Detection. As for human gesture recognition, great attentions should be attached to the classification of gesture data features. The algorithm of SVM was originally designed to solve the binary classification problem, but when it comes to a multiclassification problem, it can be constructed into multiclass classifiers by direct and indirect methods. In this paper, air gestures are recognized as multiclassification problems. According to the properties of the SVM algorithm and the various types of gestures, the indirect one-versus-one method is used. Since the CSI signal is a time-varying signal, the duration of gesture work is variable and unpredictable, which leads to the scale of the extracted feature matrix being variable, while the traditional SVM kernel function can only deal with a vector of equal length, and the inner product of the CSI signal feature matrix of two gesture motions cannot be directly calculated. The matching process is to extend the measured data evenly until it is consistent with the length of the reference template. The similarity of different length data can be obtained by using the idea of DTW and algorithm dynamic programming. And the time difference of the data can be adjusted to make the matching closest.

Taking the $[\alpha, \beta]$ in feature data $F = [\alpha, \beta]$ of the stage of feature extraction as the classification input feature of the support vector machine, after classifying the gesture data, the template receipt $F_m = [\alpha_m, \beta_m]$, $m = 1, 2, \cdots, n$ for each kind of gesture is calculated. Finally, the characteristic value of the gesture signal data $F_d = [\alpha_d, \beta_d]$, $d = 1, 2, \cdots, n$ to be tested is compared with the template data, and the recognition of the gesture to be tested is realized by using the DTW algorithm. The DTW algorithm uses the idea of dynamic programming to calculate the similarity of the data of different lengths. And the time difference of the data can be adjusted to make the matching closest. The distance matrix from the characteristic value of the gesture to the eigenvalue of the template gestures can be represented as $\mathrm{DS}[F_m, F_d]$, and the shortest matching distance D between each data can be obtained according to different conditions.

$$D = \min\left\{\sum_{n=1}^{N} DS[F_m, F_d(w(n))]\right\}, \tag{14}$$

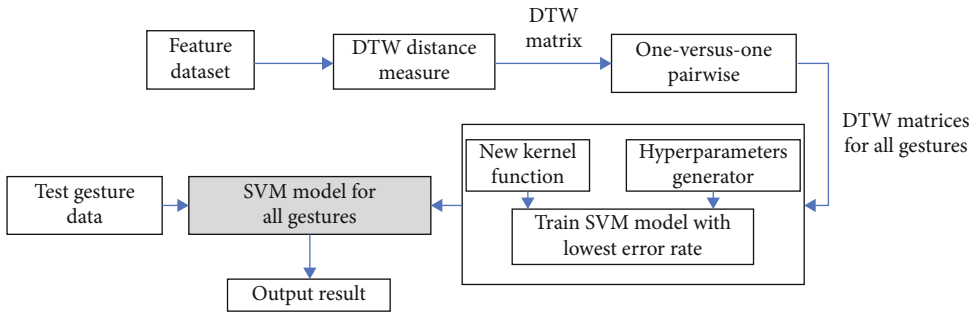

Figure 9: S-DTW algorithm structure diagram.

where $w(n)$ is the time warping function; the time axis of the data to be measured is nonlinearly mapped to the timeline of the template.

Because of the properties of air gestures, the signal values of each gesture motion are not fixed, and the time of occurrence will also be different. While in the SVM algorithm, the kernel function is the classification of the vector of equal appearance. Therefore, according to the time-varying properties of the air gesture signal, the use of the DTW algorithm can effectively process the time-varying properties of the signal. In the combination of DTW and SVM, the kernel function directly affects the accuracy and operation time. Combined with the properties of the gesture signal, the radial basis kernel function is used in this paper.

$$K(x, x_i) = e^{\left\{-|x-x_i|^2/\sigma^2\right\}}, \tag{15}$$

$$K(x, x_i) = e^{\left\{-D^2/\sigma^2\right\}}, \tag{16}$$

where D is obtained by the formula (14), by replacing the value of (15) kernel function $|x - x_i|$ with the shortest matching distance, get the new kernel function (16), which can solve the problem of the time-varying characteristics of air gestures and the equal length signal required by the kernel function. For the whole construction process of the S-DTW algorithm, the existing SVM training methods can be used directly. The whole mechanism is summarized in Figure 9 for a better understanding.

4. Experimental Validation

4.1. Experimental Data Acquisition. The performance of WiNum is verified by a large number of experiments. The PC1 with Intel 5300 NIC wireless network card is used as a receiver (DP) to receive signals, and another PC2 with Intel 5300 NIC wireless network card is used as a transmitter (MP) to transmit signals on channel 149 at 5.74 GHz. The detailed parameter setting is listed as Table 1. The network card device driver is modified to read the CSI value, which is a parsed 802.11n CSI tool that put forward Halperin and others [40, 43], and the data is processed by MATLAB software. The experiment evaluated the method in the multipath laboratory and open conference room. Ten air gestures of the handwritten number 0-9 were selected. The number of samples collected for each gesture was 1000, 70% of which were used as train samples and the other 30% as test samples.

Table 1: Parameter setting.

Parameters	DP	MP
Mode	Injection	Monitor
Channel number	5.74 GHz (channel 149)	
Bandwidth	20 MHz	
Channel sample rate	200 times per second	
Number of subcarriers	30	
Index of subcarriers	[-28, -26, …, -4, -2, -1, 1, 3, ..., 27, 28]	
Transmit power	8 dBm	

The tester did handwritten numeral gestures under two environments. At the same time, the CSI information needed in the experiment was obtained by the CSI tool at the receiving end, and the gesture motion recognition was carried out by the WiNum method.

Figure 10(a) shows the plan of the laboratory. Under this scene, there are a lot of items, such as office desks, chairs, bookcases, computers, flowers, and people's interferences. The size of the laboratory is 7 meters ∗ 8 meters, and Figure 10(b) is the floor plan in a conference room; the relatively empty conference room is 6 meters ∗ 4 meters in size. The gesture database is established in two environments: indoor emptiness and indoor multipath. The database contains the above ten kinds of test gestures. Each gesture has a period of static time before and after the test behavior, and the gesture motion lasts for 4 seconds. The tester experimented repeatedly in two scenarios.

4.2. Analysis of Different Information on Accuracy. In the process of the experiment, the following indexes are used to test and evaluate the performance of WiNum. The performance index of the system is the key to the quality of the air gesture model, and all the indicators are surrounded comparing the differences between the recognition situation and the real situation. The effectiveness of the method can be verified by comparing the recognition accuracy of different gestures, while the robustness of the method can be verified by comparing the performance changes in different environments. There is no absolute standard for the selection of the performance index of the system, which is usually based on the actual function and performance index that needed for the natural selection.

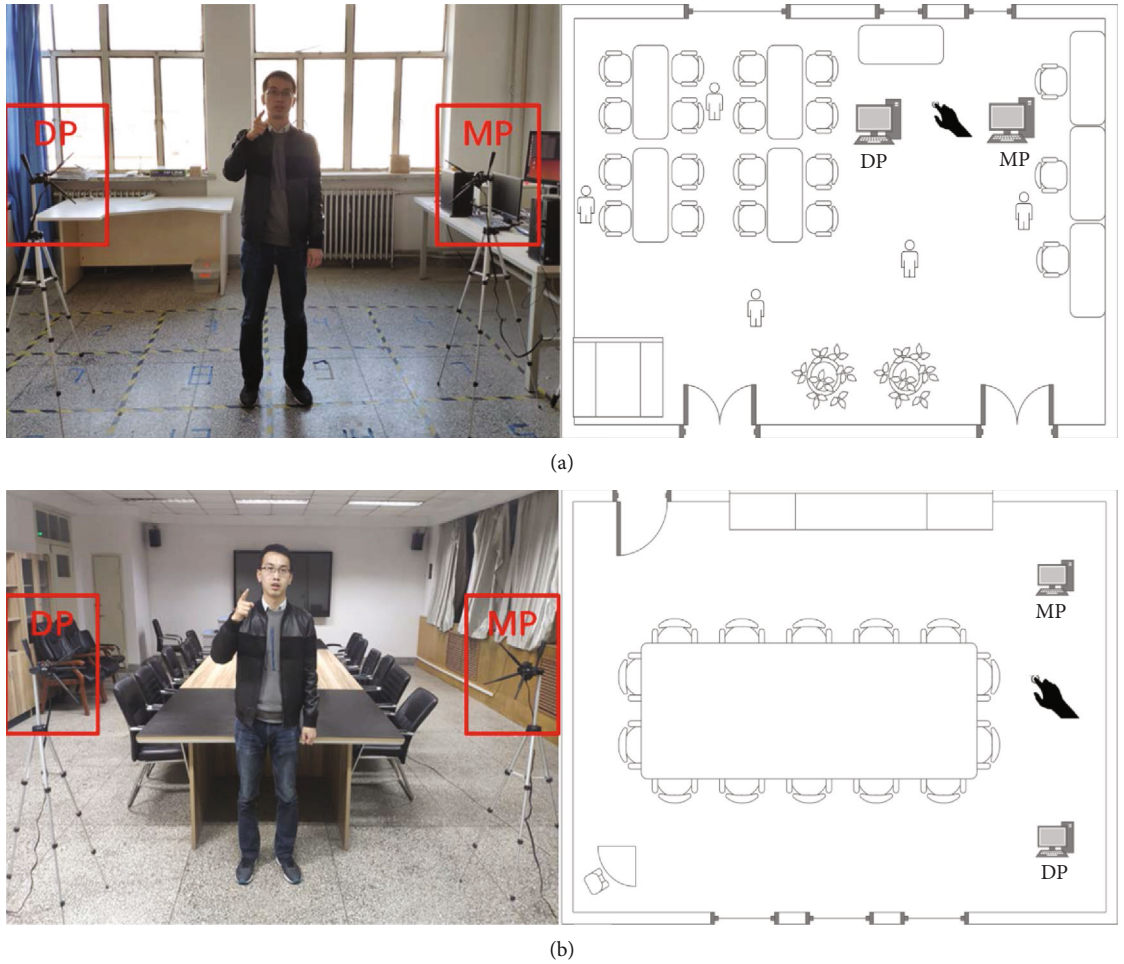

Figure 10: Experimental scenarios. (a) Laboratory scene; (b) conference room.

(i) The true positive rate (TPR): the TPR of gesture A is defined as the percentage of gesture A that is correctly recognized as gesture A

(ii) The false-positive rate (FPR): the FPR of gesture A is defined as the percentage of all test gestures except A that are mistakenly recognized as A

(iii) Recognition accuracy: the percentage of the correct number of gesture motion recognition in the total number of tests

(iv) Average accuracy: the average accuracy of gesture motions experimenting in two indoor environments

In order to test the recognition performance of WiNum, using the single signal data compared with the joint data under two different experimental scenarios, "single data" means that in the feature extraction stage, only the amplitude information that is collected in the CSI signal of gesture motion is used as the gesture motion recognition signal, and the phase information is ignored. "Joint data" represents that in the feature extraction stage, not only the amplitude information in the CSI signal is used but also the amplitude information and the phase information are used as fusion information to prevent the loss of related information. In order to further analyze the FPR and TPR under different conditions, in the course of the experiment, the control variable method is used to ensure that the environmental factors such as testers and hardware parameters remain unchanged. In most cases, the evaluation results are shown in Figures 11(a) and 11(b), the recognition of different indoor environments. 80% of FPR in multipath laboratories is lower than 5%. In the open conference room, the TPR of most gestures is higher than 95%, and the FPR is lower than 3%. Therefore, the joint signal data proposed by WiNum reduces the recognition error and is beneficial to the recognition of the motion.

4.3. Optimization Analysis of Hardware Parameters. We mainly study the recognition of Wi-Fi signals for air gestures in the 5 GHz band. Therefore, in order to verify that the 5 GHz band is more beneficial to the recognition of human perception, we selected the common 10 handwritten number gestures to do experiments repeatedly. In the experiment, the effects of the two signals on the accuracy are compared under the 2.4 GHz and 5 GHz band signals, at three different packet transmission rates. In order to ensure the stability of the

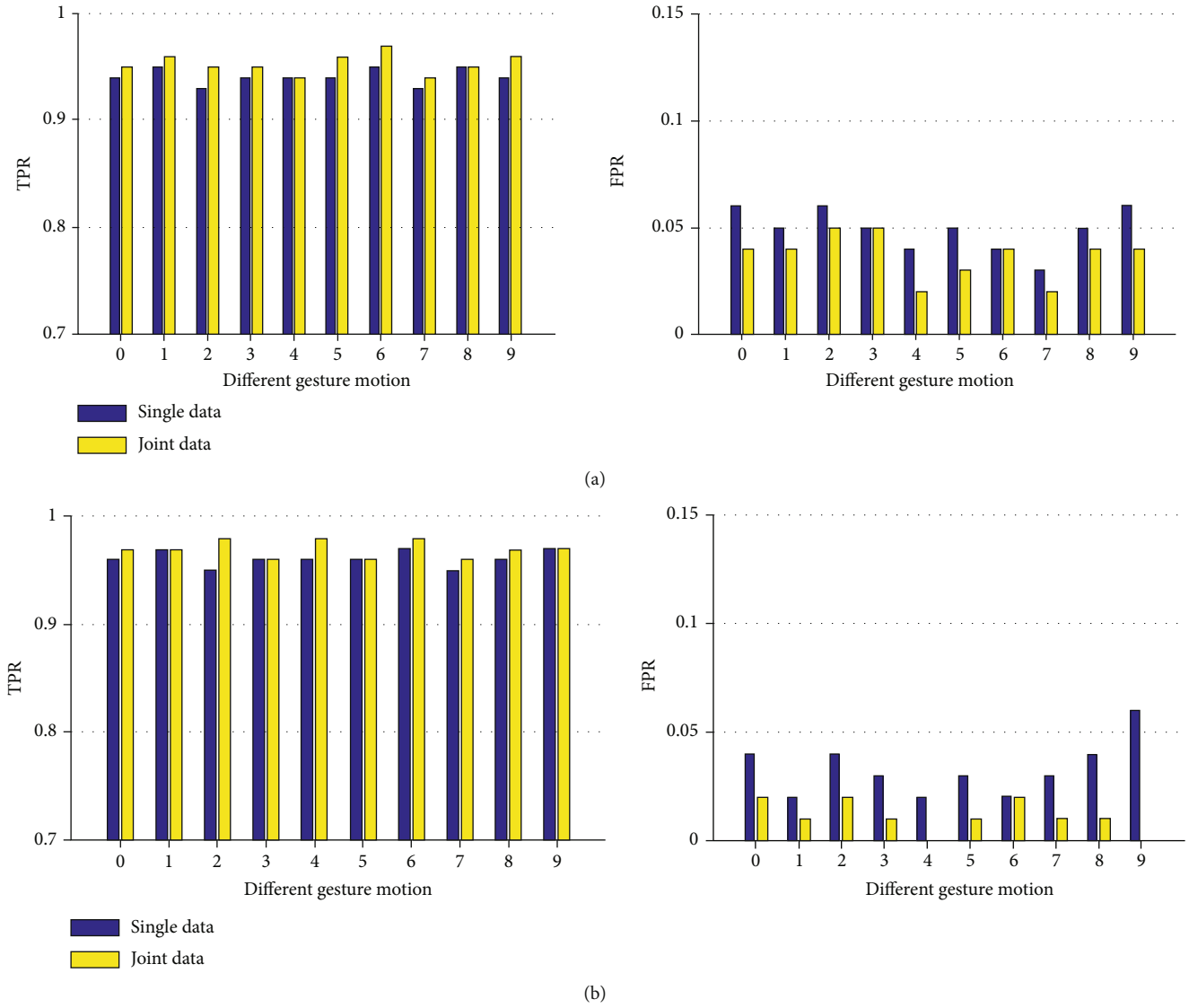

Figure 11: Comparison of TPR and FPR of different reference information. (a) Under multipath laboratory environment; (b) under a spacious conference room environment.

experimental results, in the process of the two scenarios, the tester and the experimental equipment are kept unified, and the test is carried out at the same time. The experimental results are shown in Figure 12(a) that the accuracy of the 5 GHz band is 7% higher than that of the 2.4 GHz band, and the accuracy of the 5 GHz band is 4% higher than that of the 2.4 GHz band. The accuracy of 300 packets/second packet rate is 3% higher than that of the 2.4 GHz band on average, which indicates that the 5 GHz band is more suitable for indoor human behavior, perception, and different packet transmission rates usually having a great influence on experimental recognition. No matter the Wi-Fi signal is 2.4G or 5G, the accuracy of the conference room is higher than the lab area, because the conference room is more spacious than the lab area. In the following experiments, the 5 GHz signal is used uniformly and the transmission rate is set to 200 packets per second so that the experimental results can achieve the best results.

The number of antennas at the receiving end and the transmitting end is different, and the effect of the experiment is greatly influenced. The number of transmit antenna TX (transmitting antennas) and receive antenna RX (receiving antennas) determines the number of communication links and can also more finely characterize the selective channel. In this paper, the numbers of 2-6 antennas are selected for comparison, the result of the recognition is the most preferred, and the same number is identified as the final result. The experimental results are shown in Figure 12(b), the CSI gesture data is acquired in two different scenes, the gesture recognition accuracy is the highest when the number of the antennas is 1TX-3RX, the identification accuracy is reduced when the number of the selected antennas is greater than 4, and the generated data amount is large; it also leads to as follows: the data processing process is complicated, the identification precision is reduced, and the number of subsequent experimental antennas is set to 4 antennas of 1TX-3RX.

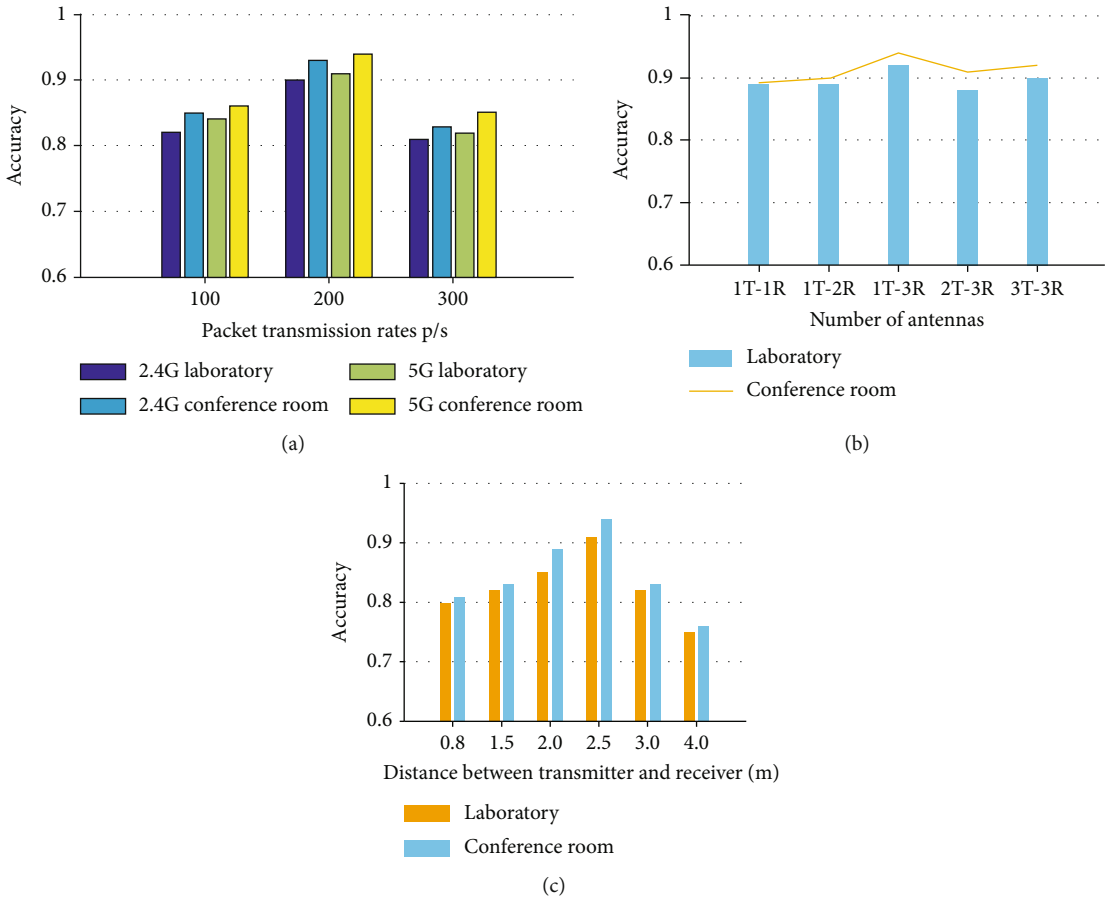

Figure 12: Effect of hardware parameters on accuracy. (a) Influence of packet rate in different frequency bands; (b) effect of the number of antennas; (c) influence of different distances.

Wireless signals propagate in straight lines indoors, reflecting and refracting on the ground, walls, equipment, and so on. When the air gesture motion occurs, the propagation path of the wireless signal will be changed, but when the tester is in the same straight line as the transmitter and the receiver, and the distance between the transmitter and the receiver is different, the influence degree on the CSI signal is different. In this paper, the straight distance between the transmitter and the receiver varies from 0.8 m to 4 m in two different indoor scenes. The experimental results show that the accuracy decreases sharply to about 75% with the increase of distance to 4 meters. Therefore, if the distance between devices is too far, it is likely that it will not be recognized effectively. The tests were carried out at 0.8 m, 1.5 m, 2 m, 2.5 m, 3 m, and 4 m distances, respectively. Figure 12(c) describes the effect of distance variation on WiNum performance. Under initial conditions, due to the unrelated body movement and serious multipath effect, the average accuracy of gesture motion is the lowest, and the recognition effect is the best at 2.5 meters.

4.4. Optimization Analysis of Sample Number and Eigenvalue Quantity. In order to determine the influence of other parameter changes on the accuracy of the handwritten number gesture recognition, this experiment has tested the feature value of different number of tuples in the feature extraction stage, taking the maximum number of eigenvalues of amplitude and phase, 2-tuple characteristic value represents maximum (amplitude, phase), 4-tuple characteristic value represents maximum first two amplitude and phase, and 6-tuple characteristic value represents maximum first three amplitude and phase, and the comparative analysis under different numbers of training samples. Figure 13 shows the recognition average accuracy of using different tuple characteristics under different training sample sets and compares the data processing execution time in different cases. As can be seen from the figure, when the characteristic tuple is 2, the recognition result is poor as the characteristic tuple is too small, and when the characteristic tuple is 6 and the number of samples is increased, the identification result is confused and the level accuracy is not high; when the characteristic tuple is 6 and the number of samples is 500, the identification accuracy is better. The shorter the number of samples and the smaller the characteristic tuples, the shorter the data processing execution time, but the accuracy of the data processing is not high. Therefore, when the number of samples is 500, the characteristic value is 4-tuple, the identification effect is the best, and the number of samples was collected for each gesture, 70% of which were used as train samples and the other 30% as test samples.

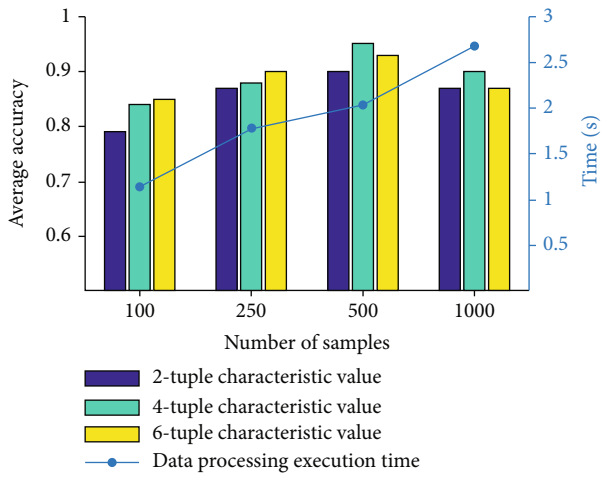

Figure 13: Effect of data sample and characteristic tuple deployment on average accuracy.

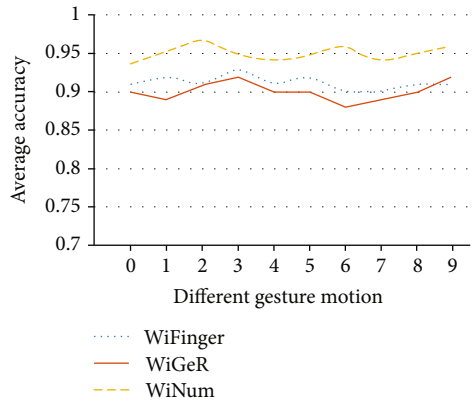

Figure 14: Comparison of accuracy of different systems.

4.5. The Comprehensive Performance Analysis of the System. In addition, in order to further evaluate the performance of the WiNum method, the recognition of different gestures under the optimal conditions is described, that is, using 5 GHz signal, and the sending rate is set to 200 packets per second, the number of antennas is 1TX-3RX, and the distance between the devices is 2.5 m. When processing the data, the quaternion features are selected, and two kinds of indoor scene tests are carried out in this case, and the recognition rate is compared with the recognition rate of the same scheme WiFinger and WiGeR under the same experimental conditions, as shown in Figure 14, which shows not only the recognition rate of all gestures but also the average recognition rate in different environments; its recognition effect is maintained at a high level and better than the same kind of schemes. This also verifies the effectiveness of the WiNum air gesture data processing method and matching algorithm.

5. Conclusions

In this paper, we proposed an air gesture recognition system WiNum that is based on WLAN physical layer information-CSI. The joint information of amplitude information and phase information of CSI is regarded as a new method of perceptual information. Firstly, signal acquisition, data processing, effective gesture extraction, etc. are carried out. Then, the S-DTW matching algorithm, which combines the features of the SVM algorithm with the DTW algorithm is used to identify different gesture motions. This method has great effectiveness and stability. According to the indoor scenes in different environments, the parameters can be adjusted accordingly, and 10 handwritten number gestures in the air can be recognized stably and efficiently, so as to realize the purpose of air gesture recognition. The overall experimental results show that the WiNum system has good performance in sensitivity, robustness, accuracy, and so on. The system should be improved in the future work, which can promote its application prospect so as to make good use in the home environment.

Data Availability

The 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

This work was supported by the National Natural Science Foundation of China under grant nos. 61762079 and 61662070 and the Key Science and Technology Support Program of Gansu Province under grant nos. 1604FKCA097 and 17YF1GA015.

References

[1] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.

[2] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, pp. 27–38, New York, NY, USA, September 2013.

[3] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: indoor localization via channel response," *ACM Computing Surveys*, vol. 46, no. 2, pp. 1–32, 2013.

[4] J. Ma, H. Wang, D. Zhang, Y. Wang, and Y. Wang, "A survey on wi-fi based contactless activity recognition," in *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, pp. 1086–1091, Toulouse, France, July 2016.

[5] Z. Zhou, C. Wu, Z. Yang, and Y. Liu, "Sensorless sensing with WiFi," *Tsinghua Science Technology*, vol. 20, no. 1, pp. 1–6, 2015.

[6] J. Xiong, K. Sundaresan, and K. Jamieson, "Tonetrack: leveraging frequency-agile radios for time-based indoor wireless localization," in *Proceedings of the 21st Annual International*

Conference on Mobile Computing and Networking, pp. 537–549, New York, NY, USA, September 2015.

[7] X. Li, D. Zhang, Q. Lv et al., "IndoTrack: device-free indoor human tracking with commodity wi-fi," *Proceedings of the ACM on Interactive, Mobile, Wearable Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017.

[8] G. Yang, "WiLocus: CSI based human tracking system in indoor environment," in *2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 915–918, Macau, China, March 2016.

[9] L. Shangguan, Z. Zhou, and K. Jamieson, "Enabling gesture-based interactions with objects," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 239–251, New York, NY, USA, June 2017.

[10] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu, "Leveraging directional antenna capabilities for fine- grained gesture recognition," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 541–551, New York, NY, USA, September 2014.

[11] H. F. T. Ahmed, H. Ahmad, and C. V. Aravind, "Device free human gesture recognition using Wi-Fi CSI: a survey," *Engineering Applications of Artificial Intelligence*, vol. 87, article 103281, 2020.

[12] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Recognizing keystrokes using WiFi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1175–1190, 2017.

[13] S. Duan, T. Yu, and J. He, "Widriver: driver activity recognition system based on wifi csi," *International Journal of Wireless Information Networks*, vol. 25, no. 2, pp. 146–156, 2018.

[14] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with wi-fi!," *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907–2920, 2016.

[15] L. A. Nguyen, M. Bualat, L. J. Edwards et al., "Virtual reality interfaces for visualization and control of remote vehicles," *Autonomous Robots*, vol. 11, no. 1, pp. 59–68, 2001.

[16] Q. Wan, Y. Li, C. Li, and R. Pal, "Gesture recognition for smart home applications using portable radar sensors," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6414–6417, Chicago, IL, USA, August 2014.

[17] M. Kavakli, "Gesture recognition in virtual reality," *International Journal of Arts Technology*, vol. 1, no. 2, pp. 215–229, 2008.

[18] H. Ren, Y. X. Zhu, G. Xu, X. Lin, and X. Zhang, "Vision-based recognition of hand gestures: a survey," *Acta Electronica Sinica*, vol. 28, no. 2, pp. 118–121, 2000.

[19] Y. Zou, W. Liu, K. Wu, and L. M. Ni, "Wi-Fi radar: recognizing human behavior with commodity Wi-Fi," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 105–111, 2017.

[20] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint activity recognition and indoor localization with WiFi fingerprints," *IEEE Access*, vol. 7, pp. 80058–80068, 2019.

[21] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Gesture recognition using wireless signals," *GetMobile: Mobile Computing and Communications*, vol. 18, no. 4, pp. 15–18, 2015.

[22] Z. Tian, J. Wang, X. Yang, and M. Zhou, "WiCatch: a Wi-Fi based hand gesture recognition system," *IEEE Access*, vol. 6, pp. 16911–16923, 2018.

[23] P. Ziaie, T. Müller, M. E. Foster, and A. Knoll, "A naïve Bayes classifier with distance weighting for hand- gesture recognition," in *Advances in Computer Science and Engineering. CSICC 2008. Communications in Computer and Information Science, vol 6*, H. Sarbazi-Azad, B. Parhami, S. G. Miremadi, and S. Hessabi, Eds., pp. 308–315, Springer, Berlin, Heidelberg, 2008.

[24] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-free presence detection and localization with SVM and CSI fingerprinting," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7990–7999, 2017.

[25] H. Jianxin and L. Zhenxiang, "Combined SVM/DTW for speech recognition," *Journal of Guizhou University (Natural Science)*, vol. 4, 2002.

[26] R. Van Nee, V. Jones, G. Awater, A. Van Zelst, J. Gardner, and G. Steele, "The 802.11n MIMO-OFDM standard for wireless LAN and beyond," *Wireless Personal Communications*, vol. 37, no. 3-4, pp. 445–453, 2006.

[27] Z. Wang, B. Guo, Z. Yu, and X. Zhou, "Wi-Fi CSI-based behavior recognition: from signals and actions to activities," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 109–115, 2018.

[28] X. Dang, X. Tang, Z. Hao, and Y. Liu, "A device-free indoor localization method using CSI with Wi-Fi signals," *Sensors*, vol. 19, no. 14, article 3233, 2019.

[29] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "Fimd: fine-grained device-free motion detection," in *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, pp. 229–235, Singapore, Singapore, December 2012.

[30] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10346–10356, 2017.

[31] S. Mitra and T. Acharya, "Gesture recognition: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311–324, 2007.

[32] M. S. Aljumaily and G. A. Al-Suhail, "Towards ubiquitous human gestures recognition using wireless networks," *International Journal of Pervasive Computing and Communications*, vol. 13, no. 4, pp. 408–418, 2017.

[33] Z. Jiang, J. Zhao, X.-Y. Li, J. Han, and W. Xi, "Rejecting the attack: source authentication for wi-fi management frames using CSI information," in *2013 Proceedings IEEE INFOCOM*, pp. 2544–2552, Turin, Italy, April 2013.

[34] M. A. A. Al-qaness, "Device-free human micro-activity recognition method using WiFi signals," *Geo-Spatial Information Science*, vol. 22, no. 2, pp. 128–137, 2019.

[35] F. Adib and D. Katabi, "See through walls with WiFi!," in *SIGCOMM '13: Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, New York, NY, USA, August 2013.

[36] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: a ubiquitous wifi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, pp. 1472–1480, Kowloon, Hong Kong, April-May 2015.

[37] M. Al-qaness and F. Li, "WiGeR: WiFi-based gesture recognition system," *ISPRS International Journal of Geo-Information*, vol. 5, no. 6, p. 92, 2016.

[38] S. Tan and J. Yang, "WiFinger: leveraging commodity WiFi for fine-grained finger gesture recognition," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 201–210, New York, NY, USA, July 2016.

[39] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf wifi,"

in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 267–276, New York, NY, USA, 2015.

[40] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11 n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[41] X. Dang, X. Si, Z. Hao, and Y. Huang, "A novel passive indoor localization method by fusion CSI amplitude and phase information," *Sensors*, vol. 19, no. 4, p. 875, 2019.

[42] X. Wang, L. Gao, and S. Mao, "PhaseFi: phase fingerprinting for indoor localization with a deep learning approach," in *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, San Diego, CA, USA, December 2015.

[43] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "802.11 with multiple antennas for dummies," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 1, pp. 19–25, 2010.