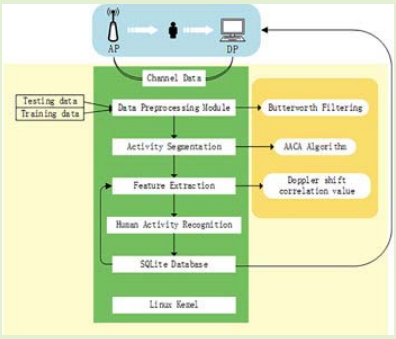
WiAct: A Passive WiFi-Based Human Activity Recognition System

Huan Yan, Yong Zhang, Yujie Wang, and Kangle Xu

***Abstract*—Nowadays, human behavior recognition research plays a pivotal role in the field of human-computer interaction. However, comprehensive approaches mainly rely on video camera, ambient sensors or wearable devices, which either require arduous deployment or arouse privacy concerns. In this paper, we propose WiAct, a passive WiFi-based human activity recognition system, which explores the correlations between body movement and the amplitude information in Channel State Information (CSI) to classify different activities. The system designs a novel Adaptive Activity Cutting Algorithm (AACA) based on the difference in signal variance between the action and non-action parts, which adjusts the threshold adaptively to achieve the best trade-off between performance and robustness. The Doppler shift correlation value is used as classification features, which is extracted by using the correlation of the WiFi device's antennas. Extreme Learning Machine (ELM) is utilized for activity data classification because of its strong generalization ability and fast learning speed. We implement the WiAct prototype using commercial WiFi equipment and evaluate its performance in real-world environments. In the evaluation, WiAct achieves an average accuracy of 94.2% for distinguishing ten actions. We compare different experimental conditions and classification methods, and the results demonstrate its robustness.**



***Index Terms*— Human motion recognition, channel state information, human-computer interaction, activity data cutting, extreme learning machine.**

I. Introduction

WITH the rapid development of information technology, the application of human-computer interaction in life has become a normalization. Many researchers have devoted themselves to the research of human-computer interaction applications, and the human activity recognition system has gradually become an important application in human-computer interaction [1], [2].

There are three main research methods of traditional human behavior recognition: computer vision technology [3]–[5], dedicated sensor sensing technology [6]–[8] and infrared sensing technology. Computer vision sensing technology may not work properly in low-light areas while undermining people's privacy. The method based on the wearable sensor requires users to carry sensing devices, which is inconvenient and causes a certain cost. The infrared sensing technology is primarily implemented using specific equipment. The costs for extensive equipment deployment limit the applicability of the method in certain situations.

In recent years, research on WiFi signals has developed rapidly. Its main research directions include daily behavioral awareness [9]–[11], indoor positioning [12], [13], and physiological signal perception [14]–[16]. Based on the observation that the activities of users impact the WiFi signals to some extent, researchers manage to use Received Signal Strength Indicator (RSSI) [17] and CSI to depict the human activity, such as PAWS [17], CARM [18], E-Eyes[19], WiFall [20], WiHACS[33] and others. PAWS [17] uses the RSSI signal to design a fusion algorithm based on the tree to identify common human actions. CARM [18] is a CSI-based human activity monitoring system, which quantifies the correlation between CSI values and human activity speed. E-Eyes [19] collects CSI information on commercial WiFi devices to recognize 11 fixed activities such as cooking and washing dishes. WiFall [20] applies a Support Vector Machine (SVM) classifier and Random Forest algorithm to classify different human activities and realize fall detection. WiHACS [33] uses OFDM Subcarriers' correlation as a distinctive feature for detecting human activity in different environments.

There are several common problems in activity recognition based on CSI, such as activity data cutting, feature extraction factors, which will influence the final results of

Manuscript received July 22, 2019; accepted August 25, 2019.Date of publication August 29, 2019; date of current version December 20, 2019. This work was supported by the National Natural Science Foundation of China under Grant 61801162. The associate editor coordinating the review of this article and approving it for publication was Dr. Ying Zhang. *(Corresponding author: Yong Zhang.)*

The authors are with the School of Computer and Information, Hefei University of Technology, Hefei 230001, China (e-mail: 1020021278@qq.com; yongzhang@hfut.edu.cn; wjiejie@hfut.edu.cn; 1099940239@qq.com).

Digital Object Identifier 10.1109/JSEN.2019.2938245

human behavior judgment. The system in [32] uses moving variance of CSI to detect the start and end of human activity. Since the threshold must be set based on experimental experience, there are still errors. However, WiHACS [33] uses adaptive window cutting activity data compared to fixed windows, which can more effectively judge the starting point and the ending point of the activity. WiFall [20] uses a Local Outlier Factor based anomaly detection algorithm, which can recognize human activity and isolate the corresponding abnormal pattern. WiFinger [21] extracts human activity using the correlation of signals between OFDM subcarriers. When there is no gesture or action, the second feature vector varies randomly between adjacent subcarriers, otherwise, the subcarriers become correlated, and the second feature changes smoothly. TW-See [22] exploits PCA to obtain the CSI correlation between human activity and its resulting changes in CSI values, and a normalized variance sliding windows algorithm is proposed to detect the beginning and end of human activities. Doppler shift feature can not only identify the direction of activity, but also realize the classification and recognition of human activity. WiSee [24] realizes gesture recognition based on WiFi signal by extracting Doppler shift of gesture action, and the active data cutting method based on a fixed threshold is to determine the threshold of the moving window using experimental data or experience. Sometimes it is not possible to cut the activity data completely, it will affect the extraction of activity feature information. After extracting the action data, it is necessary to extract the key action information, which not only improves the recognition efficiency, but also greatly reduces the matching time. CARM [18] uses the wavelet transform to extract features of different frequency bands, thereby obtaining activity speeds corresponding to different parts of the body. Human activities cause the propagation of WiFi signals and cause changes in CSI signals. Therefore, it is necessary to determine the action endpoints in the CSI signals generated by human activities.

In this paper, we demonstrate the potential to recognize human activity by using the feature information available on Commercial Off-The-Shelf (COTS) wireless devices. The contributions of our work are summarized as follows:

1. We propose a ubiquitous WiFi-based human activity system, which leverages the physical layer Channel State Information in widely deployed off-the-shelf WiFi infrastructure as an indicator. Compared with some traditional human behavior recognition systems. WiAct does not require users wear special equipment, which is convenient and causes a low cost.
2. We design an AACA algorithm based on the difference between the activity and the stationary parts of the variance features extracted from CSI signal data, which adjusts the threshold adaptively to achieve the best trade-off between performance and scalability.
3. We use the correlation of multiple antennas of the WiFi device to extract activity-related Doppler shift correlation value, which is used as input data for ELM. We compare with the traditional classifier, such as Hidden Markov Model (HMM) [27], Sparse AutoEncoder (SAE) [28], [29], Back Propagation Neural Network (BP Neural Network) [22], Long Short-Term Memory (LSTM) [27], the experimental settings identify ten kinds of actions, such as kicking, running, experimental results show that ELM has high accuracy.

The rest of the paper is organized as follows. Section 2 provides an overview of related work. Section 3 describes the framework of our WiAct system in detail. Then the experiments and evaluations conducted in terms of accuracy, robustness and scalability are illustrated in Section 4. Finally, we conclude our work and show some potential future works in Section 5.

II. Related Work

Previous work on human activity recognition can be generally divided into wearable sensors, computer vision and WiFi. Each method has advantages and disadvantages. In this section, we summarize the related work on human activity recognition using these methods.

Human activity recognition systems based on wearable sensors have emerged with the development of mobile computing, but their performance is strongly dependent on the sensor devices worn, such as the type of sensing signal of the device, the wearing position and quantity of the device, and so on. Fortin-Simard *et al.* [6] using RFID technology to extract relevant information in order to detect the activity of daily living. Yatani and Truong [7] proposed BodyScope, which records the sounds through acoustic sensors for classifying activities, such as eating and coughing. Some wearable sensors are used in mobile devices with high frequency and convenience. Chen *et al.* [10] used a smartphone with an accelerometer sensor for human activity recognition.

Video-based detection of human activity requires maintaining the human body in front of an unobstructed camera. The detection of human activity is mainly based on the posture and behavior of the captured human body. Shotton *et al.* [3] proposed a new method to quickly and accurately predict 3D positions of body joints from a single depth image. Rautaray and Agrawal [4] provided a gesture based interface for controlling applications like media player using computer vision techniques. Wang *et al.* [5] proposed a high-precision activity recognition system using deep learning and convolution networks. However, the computer vision sensing technology may not work properly in low-light areas while undermining people's privacy.

Compared with these technologies, wireless signals are widely used and have strong penetrability. It is highly sensitive to the environment and can be used as an important means of environmental awareness. In this context, CSI has been widely used. WiFall [20] applies a Support Vector Machine (SVM) classifier and Random Forest algorithm to classify different human activities and realize fall detection. Wu *et al.* [22] proposed a device-free passive human activity recognition system with Wi-Fi signals does not require any special equipment to meet the signal scene through the wall. Yang *et al.* [23] proposed an IoT platform that supports WiFi for occupancy sensing. Device-free human

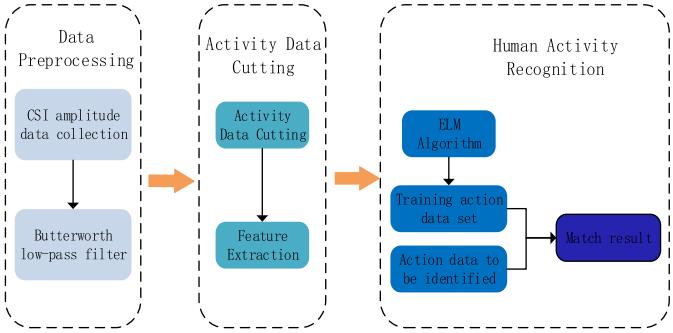


Fig. 1. System architecture of WiAct.

activity recognition has its inherent benefits in terms of cost-effectiveness and compatibility. Therefore, we propose a passive WiFi-based human activity recognition system, which explores the correlations between body movement and the amplitude information in CSI to classify different activities.

III. System Design

A. System Overview

In this section, we introduce the WiAct system for human activity recognition, which consists of data preprocessing, activity data cutting, feature extraction and human activity recognition, as shown in Fig. 1.

In the data preprocessing design, we acquire precise data to ensure the accuracy and robustness of the system. As far as we know, due to the complex environment and signal propagation effects, the original CSI signal is composed of useful signals and some useless noise and outliers. Thus, it is essential to use appropriate methods in the data preprocessing to remove useless noise from the original CSI signal. In the case of CSI signal, Butterworth low-pass filter is a simple and useful algorithm, which can effectively eliminates high-frequency noise caused by hardware devices and sharp environment variation. Therefore, In order to achieve this goal, we adopt Butterworth low-pass filter for denoising, which has stable amplitude-frequency characteristics both inside and outside the passband so that low-frequency signals can be retained to a greater extent. In the activity data cutting and feature extraction design, we need to detect the start and the end of an activity. Given that the activity affects signal paths, the variance of the CSI amplitude during an activity is generally much higher than that with no activity. The fusion of the antenna is used to extract the Doppler shift correlation value produced by the activity. In the human activity recognition design, ELM algorithm trains the activity feature data set, and identifies the test feature data to obtain the recognition result.

B. CSI Overview

The emergence of Orthogonal Frequency Division Multiplexing (OFDM) technology assists us to extract CSI from the transmitting end to the receiving end of the wireless signal. In the field of wireless communication, CSI is the physical layer information with a fine-grained attribute value, which describes the amplitude and phase information of each subcarrier to express channel characteristics. During the propagation of wireless signals, they are affected by the physical environment, which leads to reflections, diffractions and scattering. CSI data can be obtained from the Intel 5300 or Atheros 9390 network interface cards. The IEEE 802.11n physical layer protocol indicates that under the condition of 20 MHz bandwidth, the CSI is divided into 30 subcarriers: $H = [H_1, H_2, \ldots, H_m, \ldots, H_{30}]$, where H_i represents the CSI on each subcarrier, expressed as follows:

$$H_i = \|H(i)\|^{e^{-j\angle H(i)}} \tag{1}$$

where $H(i)$ represents CSI data of the i th subcarrier, $\|H(i)\|$ and $\angle H(i)$ represent the amplitude and phase of the i th subcarrier, respectively. CSI describes the link state information of the wireless signal at the transmitting end and the receiving end, and characterizes the frequency response of the wireless channel. Let $R(f,t)$ and $T(f,t)$ be the frequency representations of the received signal and the transmitted signal, respectively. The relationship between the transmitting end and the receiving end is as follows:

$$R(f,t) = H(f,t)T(f,t) + N(f,t) \tag{2}$$

where $H(f,t)$ represents the frequency response of the channel state information, and $N(f,t)$ represents the noise.

C. Data Preprocessing

Due to the complex wireless propagation and the surrounding environment, there is a large amount of noise in the collected CSI amplitude values for different human actions. Fig. 2(a) and (c) display the raw CSI amplitude values of one subcarrier. It can be observed that there are some abrupt fluctuations. Obviously, these noise are not induced by human actions, the existence of noise will cause the misclassification of the classifier and affect the recognition accuracy of the system. Therefore the purpose of preprocessing is to eliminate the noise in the CSI amplitude values.

WiAct leverages a Butterworth low-pass filter for CSI denoising, the CSI amplitude changes caused by common human activities are mostly lower than 100Hz [22]. Hence, we set the cut-off frequency of the filter to 100Hz based on experimental data and empirical values. Fig. 2(b) and (d) show the results after processing by the Butterworth low-pass filter, which shows that interference noise can be effectively removed.

D. Activity Data Cutting

1) AACA Algorithm: Since the collected CSI information contains static state information, training it into the neural network will result in low classification accuracy. Given that the activity affects signal paths, the variance of the CSI amplitude during an activity is generally much higher than that with no activity. The complete activity information data is an important guarantee for improving the recognition rate of the neural network.

In order to solve the above problem, We design an AACA algorithm by using the difference between the activity and

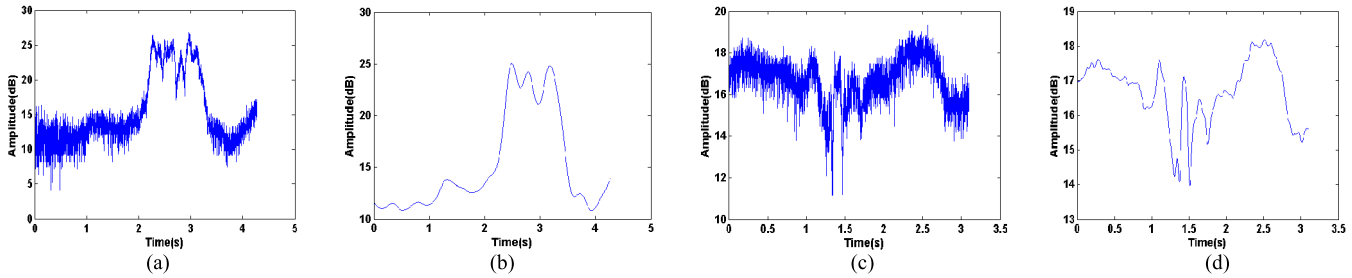


Fig. 2. Raw CSI and its denoising results. (a) Raw CSI of one subcarrier (kicking). (b) Butterworth low-pass filter waveform (kicking). (c) Raw CSI of one subcarrier (waving). (d) Butterworth low-pass filter waveform (waving).

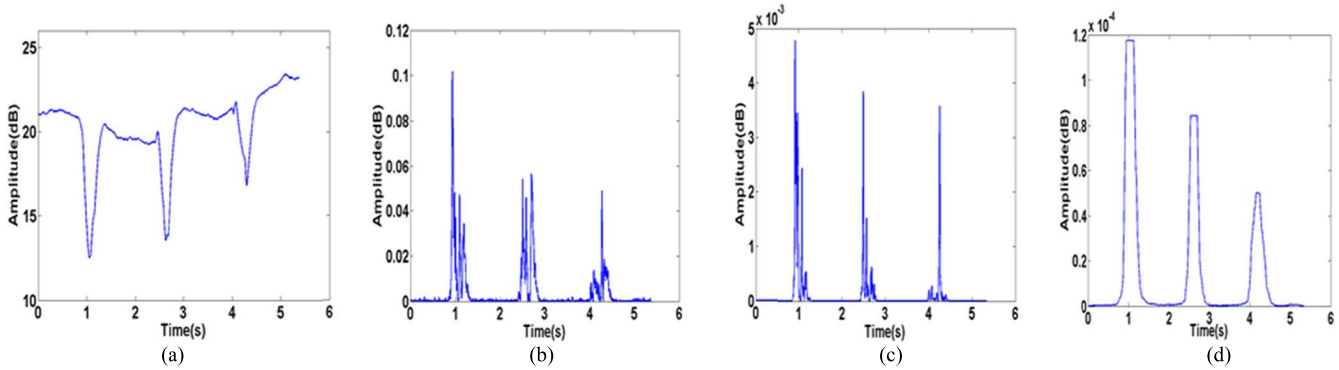


Fig. 3. (a) Original waveform. (b) Tem1. (c) Tem2. (d) Tem3.

the stationary parts in the signal variance feature. The AACA algorithm adjusts the threshold adaptively to achieve the best trade-off between performance and scalability. However, some algorithms [23] for cutting actions are based on thresholds that do not change. This may cause the cutting result to contain some calm parts, maybe some important features are missing from the cutting results.

Our AACA algorithm steps are as follows:

Part1: Variance processing and data smoothing

a. Fig. 3(a) is a waveform of the push action after data preprocessing, which illustrates the AACA algorithm steps. For the denoised signal y, the window size win is set to $f_s/20$, and f_s is the packet transmission rate, *offset* is the step length between the adjacent sliding windows, which is set to be 1. The variances in every slide window consist tem1 and the variance in the t th window is defined as:

$$y_{var}[t] = \frac{1}{win} \sum_{i=t^*offset}^{win+t^*offset} (y[i] - \frac{\sum_{i=offset}^{win+t^*offset} y[i]}{win})^2 \tag{3}$$

The variance represents the fluctuation of signal CSI, as shown in Fig. 3(b). It can be found that the larger the variance value is the part of the human activity. The variance is very small, which means the flat part of the original waveform. However, there are still some small fluctuations in the non-action part of the Fig. 3(b), therefore we need to reduce the fluctuations caused by non-actions.

b. In order to reduce the fluctuations caused by the non-action part of tem1. Sum the sliding window, the result is expressed as:

$$y_{sum}[t] = \sum_{i=t^*offset}^{win+t^*offset} y_{var}[i] \tag{4}$$

where y_{sum} represents the sum of each sliding window, and perform a front-to-back differential operation on each sliding window to obtain the waveform tem2.

$$y_{diff}[t] = y_{sum}[t] - y_{sum}[t-1] \tag{5}$$

where y_{diff} ndicates the result of the difference between each sliding window, then use the sliding window to calculate the sliding window variance, and perform a front-to-back differential operation, and sum the sliding window. After these operations, since there are multiple peaks, as shown in Fig. 3(c), it is possible to divide motion signal into multiple independent parts. To solve this problem, the median filtering method with a window size of $f_s/20$ is used to obtain $tem3$ as shown in Fig. 3(d).

Part2: Automatic threshold setting and activity detection

a. By observing the results of Fig. 3(d), it can be found that the larger the tem3 value is the part of the human activity. The portion where the tem3 value is closer to 0 is the portion where the original waveform is flat. Set the initial threshold and move the window continuously:

$$thret = \sum_{i=1}^{win^*2} tem3[i] \tag{6}$$

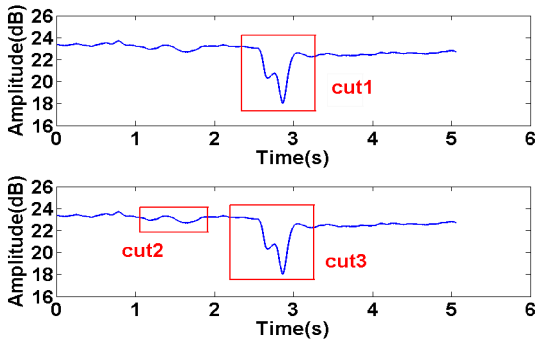


Fig. 4. Segment result.

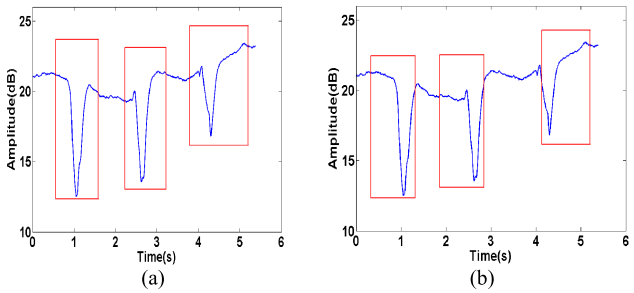


Fig. 5. Segment result.

We need to multiply the initial threshold by a judgment coefficient. In this paper, with the processing of part1, we smoothed out the variance of the non-action part and enlarged the variance of the action part. In the normal case, the amplitude is less than 0.015dB when there is no action, and the amplitude is greater than 0.2dB when there is action, so the ratio of the two is 13.33. In order to ensure as much as possible that the information of the action part is not lost, we set the judgment coefficient to 12. Based on these analyses, the initial threshold is multiplied by the set judgment coefficient during the cutting of activity data, so that more action signals can be retained. If the sum in a window is greater than $12^{*}thret$, set the corresponding time point as the starting point. Otherwise change the threshold:

$$thret_{new} = \frac{thret_{old} + \sum_{i=t}^{win+t} tem3[i]}{2} \tag{7}$$

The updated threshold consists of two parts, one is the previously generated $thret_{old}$ and the other is $\sum_{i=t}^{win+t} tem3[i]$, the updated threshold is not only related to the previously generated threshold, but also related to changes in the current window. Since the generation of the new threshold is closely related to the current window and the old threshold, when we detect the start and end points, the old threshold and the value generated by the current window are equally important, so we set the weight of the two parts to 0.5.

b. If the starting point is found, move the window after the starting point until the sum within a window is less than or equal to $12^{*}thret$, then the threshold is automatically updated:

$$thret_{new} = \frac{thret_{old} + \sum_{i=t}^{2win+t} tem3[i]}{2} \tag{8}$$

When judging the end point, we need to consider the starting point of the action that has been determined at this time. Since the updated threshold is related to changes in the current window, we set the window size to $2^{*}win$ at this time. In Fig. 3(a), the same push action was performed in the vicinity of 1 second, 2.5 seconds, 4.5 seconds, respectively. There is a certain data fluctuation in the calm state, if the variance of the non-action part is large, the data smoothing process without steps 2 and 3 may not cut out the complete activity data.

2) Comparison of Experimental Results of AACA Algorithm: We used a fixed threshold activity data cutting method to compare the cutting results of the data processing after steps 2 and 3. The fixed threshold for the fixed threshold activity data cutting method is set to $12^{*}thret$, the cutting result is shown in Fig. 4. It uses the same CSI raw data for cutting, and performs a push action during 4 to 5 seconds, it shows the experimental method, respectively. The $tem1$ as the cut data segments the non-action part $cut2$, which caused by the non-action portion of $tem1$ has small fluctuations, and does not contain the complete action information. We can see that using tem3 as the cutting data is better than using tem1 as the cutting data. Therefore, we chose to use tem3 as the cutting data for adaptive activity data cutting.

The AACA algorithm also shows good cutting results for continuous motion signals, we test continuous push action, and ask the volunteers to start each action at about 1s, 2.5s, 4s, each action lasts about 0.5s. Fig. 5(a) shows the results of AACA algorithm cutting. Fig. 5(b) shows the result of fixed threshold activity data cutting, the constant threshold at this time is determined according to the experimental scene and experience. In this paper, the constant threshold is set to $12^{*}thret$.

It can be seen from Fig. 5(b) that the result of fixed threshold activity data cutting has a large error, sometimes the result of the cut contains a lot of calm parts, and sometimes it loses some important activity feature data. AACA algorithm can cut out complete activity data, which facilitates key feature extraction.

E. Feature Extraction

The Doppler effect is mainly caused by changes in the relative positions of the wave source and the observer, resulting in changes in the wavelength of the object radiation [25]. Waves in front of the moving wave source are compressed, resulting in higher frequencies and shorter wavelengths; Waves behind the moving wave source are stretched, resulting in lower frequencies and longer wavelengths. In this paper, the wavelength and frequency variation of the reflected signal is generated by the relative position change between the target

TABLE I
DETAILED STEPS OF THE ELM ALGORITHM

ELM Algorithm detailed steps:
Set training samples $\overline{N}=\left\{(z_i,t_i)\middle\vert x_i\in R^p,t_i\in R^q\right\}_{i=1}^{N}$, The activation function is $G(x)$, The number of hidden layer neural units is L.
Step1 Specify the connection weights of the input layer and the hidden layer arbitrarily w_i, b_i, $i=1,2,\ldots,L$.
Step2 Calculate the output matrix of the hidden layer of the neural network H.
Step3 Calculate the connection weight matrix of the hidden layer and the output layer $\beta=H^{\dagger}T$.

and the transmitter end and the receiver end. The Doppler frequency shift of the signal reflected off the object is:

$$\Delta D\propto\frac{2v\cos\alpha}{c}f \tag{9}$$

where ΔD represents the generated Doppler shift, $v$represents the target activity velocity, α represents the target activity direction, c represents the speed of light, and f represents the signal frequency.

When the CSI signal is actually extracted, the surrounding objects will reflect the wireless signal, and the signal generated by the transmitter reaches the receiver through multiple paths, which is called multipath propagation, assuming that the wireless signal arrives at the receiver via N different paths, the frequency channel response $H(f,t)$ of frequency f and time t is:

$$H(f,t)=H_s(f)+\sum_{k\in P_d}\alpha_k(t)e^{-j2\pi f\tau_k(t)} \tag{10}$$

where $H_s(f)$ is the sum of responses of all static path and is the set of dynamic path, P_d is the attenuation factor of the k th path, and $\tau_k(t)$ is the propagation time of the k th path. According to the Doppler effect mentioned above, $H(f,t)$ is expressed as:

$$H(f,t)=H_s(f)+\sum_{k\in P_d}\alpha_k(t)e^{j2\pi\int_{-\infty}^{t}\Delta D_k(v)dv} \tag{11}$$

As can be seen from [25], the link with the largest variance and maximum amplitude can be selected to obtain the most obvious part of the static response and dynamic response. In this paper, the frequency channel response of two links is used for conjugate multiplication, and the amplitude value of the result is used as the feature of activity recognition.

F. Human Activity Recognition

For a single-hidden layer feedforward neural networks, ELM does not need to manually set a large number of network training parameters, which can generate a local optimal solution, and also has a fast training speed and a strong generalization ability for data sets [30], [31].

We use a three-layer ELM in our system, including an input layer with 200 neurons, an output layer with 10 neurons and a hidden layer with 400 neurons. It is assumed that N is the size of the training set, the input is N arbitrary samples (z_i, t_i), z_i is the result of feature extraction of Doppler shift correlation value, and $z_i = [z_{i1}, z_{i2}, \ldots, z_{in}]^T$, $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$, where n and m represent the number of input layer neurons and output layer, respectively. The model of the feedforward neural network is:

$$\sum_{i=1}^{L}G(w_iz_j+b_j)\beta_i=o_j,\quad j=1,2,\ldots,N \tag{12}$$

where $G(x)$ is the activation function and o_j is the actual output vector, the sigmoid function is selected as the activation function, the actual output vector o_j contains the action tag, the label corresponding to each action is as shown in 4.1.

Approximate N samples with zero error of the above formula to obtain:

$$\sum_{j=1}^{N}\left\|o_j-t_j\right\|=o \tag{13}$$

where t_j corresponds to the true output vector of the label o_j, and the above formula can be used to have w, β, b, so that:

$$\sum_{i=1}^{L}G(w_iz_j+b_j)\beta_i=t_j \tag{14}$$

Use the matrix to simplify:

$$H\beta=T \tag{15}$$

$H$is called the output matrix of the neural network's hidden layer. For the single hidden layer feedforward neural network, the connection weights of the input layer, the hidden layer and the threshold of the hidden layer neurons need to be continuously resized by iteration, and the ELM algorithm shows that w and b do not need to be adjusted, they can be arbitrarily specified. By calculating the hidden layer output matrix H, the weight matrix β can be obtained, as follows:

$$\beta=H^{\dagger}T \tag{16}$$

The specific algorithm steps are as below:

The weight matrix is obtained, and the classification model is trained, then the feature value of the action to be identified is used as the input of the ELM. The training label is $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$, $i = 1, 2, \ldots, N$. The test data set $p_{z_i} = [z_{i1}, z_{i2}, \ldots, z_{im}]^T$, $i = 1, 2, \ldots, M$, where $\overline{N} = \{(z_i, t_i) \,|x_i \in R^p, t_i \in R^q\}_{i=1}^{N}$ is the total number of test set samples, the test label is $p_{t_i} = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$, $i = 1, 2, \ldots, M$.

According to the discussion of the related work, different human body activities exhibit different Doppler shift correlation value. ELM supports input fine-grained data. In order to construct the required network model, the Doppler shift correlation value corresponding to different actions are used as training set. In order to ensure the training effect, 25%

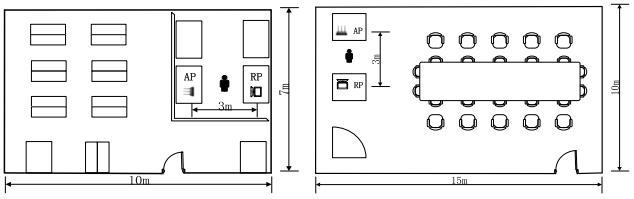


Fig. 6. Floor plans of two experimental environments. (a) Lab. (b) Meeting room.

TABLE II
THE TRAINING SAMPLES AND TEST SAMPLES

Activity	Training Samples	Test Samples
Pushing(Pu)	330	110
Waving(Wav)	330	110
Kicking(Ki)	330	110
Running(Ru)	330	110
Falling(Fa)	330	110
Boxing(Bo)	330	110
Sitting(Si)	330	110
Picking(Pi)	330	110
Walking(Wal)	330	110
Empty(Em)	330	110

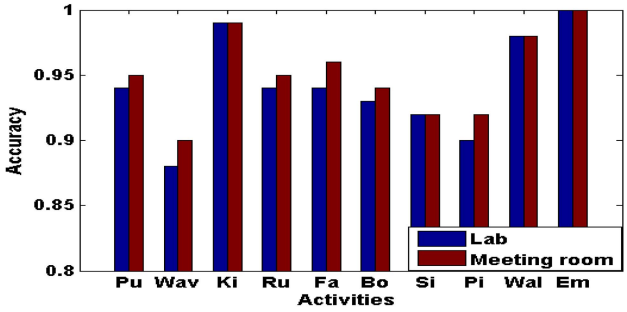


Fig. 7. Experiment results of different environments.

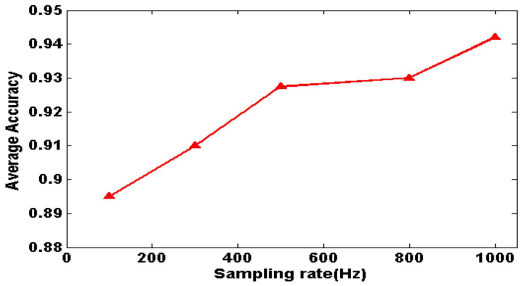


Fig. 8. The impact of different transmission rates.

of the training set is set as the test set, then the connection weight of the input layer, the hidden layer and the threshold of the hidden layer neuron are initialized, and the connection weights of the corresponding hidden layer and the output layer are calculated by the ELM algorithm. Finally, the classification results are obtained.

IV. EXPERIMENTAL EVALUATION

A. Experiment Setup

The proposed system is implemented by using off-the-shelf hardware devices. The WiAct consists of one transmitter and one receiver equips with wireless cards. The desktop computer is equipped with the Ubuntu 12.04 system and the TL-WDR6500 is used as the Access Port (AP). The Intel 5300 NIC is used as the Receive Port (RP). The AP has two 5GHz transmit antennas and the RP has three receive antennas. For a desktop computer, we install the CSI tools developed by Halperin *et al.* [26] to collect the CSI reported by the wireless cards. The received data format is $2 \times 3 \times 30$, where 2 and 3 are the number of antennas of the transmitter and receiver, respectively, and 30 is the number of subcarriers in the OFDM channel between the transmitter and the receiver. In WiAct, since CSI with higher amplitude may have a larger static response, CSI with higher variance may have a larger dynamic response. Therefore, the two links with the largest variance and the largest amplitude are selected from the six links, then the active data is segmented for 60 subcarriers. The frequency channel response of two links is used for conjugate multiplication, and the amplitude value of the result is used as the feature of activity recognition. The sending speed of our transmitting equipment is 1000 packets/second. We also harvested datasets with varying conditions for comprehensive evaluations. Two different environments, i.e., a lab room of size $10m \times 7m$ and a meeting room of size $15m \times 10m$, are considered. The two experimental environments are shown in Fig. 6.

In two experimental scenarios, the distance between the receiver and the transmitter is 3 meters, and the tester performs actions at the midpoint of the transceiver link. We recruited six volunteers with three males and three females to participate in the CSI data collection. We asked each tester to perform each action 440 times, a total of 4400 times in the two experimental environment to evaluate our system. The system randomly selected 25% of the data as the test data set and 75% of the data as the training data set, then use the average of 10 accuracy rates as the final recognition accuracy. The names and sample numbers for each activity is described in Tab. II, the movement of "Empty" which means that no user has done the action near the receiving link.

B. Impact of Different Environments

We evaluate the robustness of WiAct when we collect the same number of samples for identification in different experimental environments. The recognition accuracies for all the ten activities under the two situations are shown in Fig. 7.

In two environments, the overall performance in the meeting room is better than in the lab room. This is because there are additional subjects in the lab room, which will cause some interference. No subjects walked around when we collected the data in the meeting room. In all actions, whether in the lab room or the meeting room, "Empty", "Walking" and "Kicking" maintain high recognition accuracy, which shows that these three actions are easy to distinguish from other seven actions. "Sitting" and the "Picking" may be easily misclassified into

TABLE III
CLASSIFICATION RESULTS OF DIFFERENT CLASSIFICATION METHODS

Classification Algorithm	Average Accuracy
HMM	75.52%
SAE	85.79%
BP	90.64%
LSTM	91.23%
ELM	94.20%

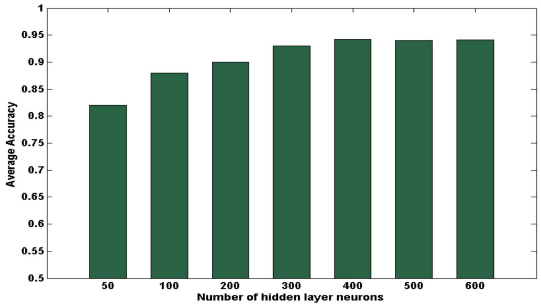


Fig. 9. The impact of the number of hidden layer neurons.

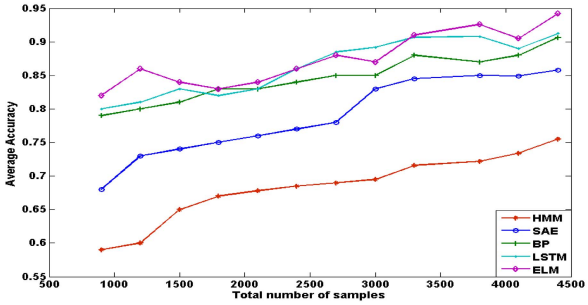


Fig. 10. The impact of the total number of samples.

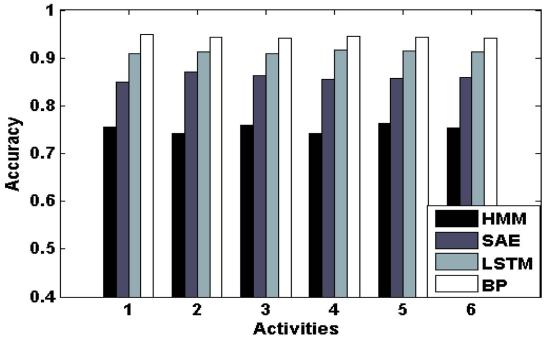


Fig. 11. The impact of different users.

each other as they are indeed similar in nature, and "Waving" is small, resulting in their recognition accuracy is not high. To carry out various experiments in the case of interference, the next comparative experimental scene is in the lab room.

C. Impact of Different Transmission Rates

We investigate the relation between WiAct accuracy and transmission rates. Fig. 8 shows the effect of different transmission rates on average recognition accuracy in the lab room.

We can find that with decreasing of sampling rate, the performance of WiAct slightly degrades. This is because high-frequency noise is related to Doppler shift. However, WiAct still shows high performance at a sampling rate of 800Hz, and speed up the system to process data. We consider that this article improves recognition accuracy. Therefore, under the condition of increasing processing time cost, this paper chooses the sampling rate of 1000HZ.

D. Comparison of Different Classification Methods

To better show the performance of the system. We compare the activity recognition methods of the previous work in the lab room. HMM[27], SAE[28], [29] and LSTM[27] in terms of empty, jumping, picking, running, sitting, waving, walking, BP[22] in terms of empty, walking, hand swing, falling, sitting, boxing, standing. The accuracy of HMM, SAE and LSTM is 77.5%, 85.9%, and 92.2%, the accuracy of BP is 94.46%. We keep the same setting as these methods and record the recognition accuracy of ten actions in 4.1. The comparison results are shown in Table III.

It can be found that when the number of actions is increased to 10, the recognition accuracy of HMM, SAE, LSTM, and BP classification methods is significantly reduced. The ELM algorithm has the highest average accuracy in identifying the ten actions. This is because ELM is selected for activity classification. It only needs to set the number of hidden layer nodes of the network, and does not need to adjust the input weight of the network and the bias of the hidden element during the execution of the algorithm, and generate a unique optimal solution, it has strong generalization ability and fast learning speed, which can identify actions faster and more accurately.

E. Impact of the Number of Hidden Layer Neurons

The number of hidden layer neurons is a tremendously critical parameter in the ELM classifier. Therefore, we choose different hidden layer neurons for comparison experiments in the lab room. As shown in Fig. 9, when the number of neurons in the hidden layer is 50, the average recognition accuracy of WiAct is extremely low in the lab room, as the number of hidden layer neurons increases, the recognition accuracy of the system increases. However, when the number of hidden layer neurons exceeds 400, the recognition accuracy of WiAct becomes stable. Therefore, we set the number of hidden layer neurons to 400.

F. Impact of the Total Number of Samples

The total number of samples is an essential factor for the proposed WiAct. Therefore, we perform an additional experiment to investigate the impact of this factor on the performance of human activity recognition. The experimental results are shown in Fig. 10.

It can be found that ELM has a high average recognition accuracy on small samples. As the total number of samples increases, ELM still has a high average recognition accuracy. Although LSTM average recognition accuracy is also high, as the total number of samples increases, the LSTM recognition is not stable. Therefore the ELM is selected on the recognition method.

G. Impact of Different Users

To evaluate the robustness of WiAct for various users, we asked six volunteers to perform ten actions in the lab room, which were used as training samples for the five classification methods. The results are shown in Fig. 11, it shows the average recognition accuracy of four actions, the x-axis represents the users,and the y-axis represents accuracy. Due to differences in people's behavior, The average recognition accuracy of each person is different. We can find that HMM has the lowest average accuracy rate for each user. For the ELM method, the average recognition accuracy of each user is almost the same. It shows the overall recognition accuracy is very good for each user. This result indicates that WiAct is robust to different people and can achieve high and stable recognition accuracy.

V. Conclusion and Future Work

In this paper, we propose a CSI-based WiFi human activity recognition system to analyze common human behaviors. WiAct has been prototyped on the low-cost and ubiquitous WiFi infrastructures and evaluated in extensive real-world experiments, where its performance has been verified. Our system has made some crucial contributions. First, we extract the variance feature from the CSI signal data, an AACA algorithm for action data is designed, accurate activity data is the guarantee of feature extraction. We use the correlation of multiple antennas of the WiFi device to extract activity-related Doppler shift correlation value. WiAct uses Doppler shift correlation value as input data for ELM, and the experimental results show that the ELM has high accuracy, which achieves an overall recognition accuracy of 94.20%.

For future work, we mainly consider using a smaller sample to get a higher activity recognition rate. Not only reduces the time it takes for the system to process data, but it also improves recognition accuracy. And we plan to identify common continuous actions. We want the computer to cut out different actions and then build a time series model for common continuous actions.

References

[1] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures," in *Proc. INFOCOM*, San Francisco, CA, USA, Apr. 2016, pp. 1–9.

[2] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Poster: WiFi-based device-free human activity recognition via automatic representation learning," in *Proc. Mobicom*, 2017, pp. 606–608.

[3] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 1297–1304.

[4] S. S. Rautaray and A. Agrawal, "A novel human computer interface based on hand gesture recognition using computer vision techniques," in *Proc. IITM*, Allahabad, India, 2010, pp. 292–296.

[5] L. Wang, Y. Qiao, and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 4305–4314.

[6] D. Fortin-Simard, J. S. Bilodeau, K. Bouchard, S. Gaboury, B. Bouchard, and A. Bouzouane, "Exploiting passive RFID technology for activity recognition in smart homes," *IEEE Intell. Syst.*, vol. 30, no. 4, pp. 7–15, Jul. 2015.

[7] K. Yatani and K. N. Truong, "Bodyscope: A wearable acoustic sensor for activity recognition," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 341–350.

[8] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[9] D. Man, W. Yang, X. Wang, J. Lv, X. Du, and M. Yu, "PWiG: A phase-based wireless gesture recognition system," in *Proc. ICNC*, Maui, HI, USA, Mar. 2018, pp. 837–842.

[10] Z. Chen, Q. Zhu, S. Y. Chai, and L. Zhang, "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 3070–3080, Dec. 2017.

[11] C. Wang *et al.*, "Low-power fall detector using triaxial accelerometry and barometric pressure sensing," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2302–2311, Dec. 2016.

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

[13] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016.

[14] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-sleep: Contactless sleep monitoring via WiFi signals," in *Proc. RTSS*, Rome, Italy, Dec. 2014, pp. 346–355.

[15] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using WiFi signals," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2071–2084, Jun. 2018.

[16] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2466–2479, Oct. 2016.

[17] Y. Gu, F. Ren, and J. Li, "PAWS: Passive human activity recognition based on WiFi ambient signals," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 796–805, Oct. 2016.

[18] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 65–76.

[19] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 617–628.

[20] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[21] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger: Talk to your smart devices with finger-grained gesture," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 250–261.

[22] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, "TW-See: Human activity recognition through the wall with commodity Wi-Fi devices," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 306–319, Jan. 2019.

[23] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3991–4002, Oct. 2018.

[24] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Int. Conf. Mobile Comput. Netw.*, 2013, pp. 27–38.

[25] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity Wi-Fi for interactive exergames," in *Proc. CHI*, 2017, pp. 1961–1972.

[26] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," in *Proc. SIGCOMM*, 2010, pp. 159–170.

[27] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[28] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6258–6267, Jul. 2017.

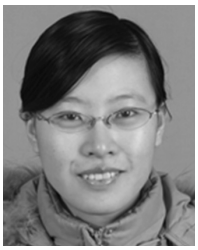
[29] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10346–10356, Nov. 2017.
[30] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.
[31] F. Benoît, M. van Heeswijk, Y. Miche, M. Verleysen, and A. Lendasse, "Feature selection for nonlinear models with extreme learning machines," *Neurocomputing*, vol. 102, pp. 111–124, Feb. 2013.
[32] L. Guo, L. Wang, J. Liu, W. Zhou, and B. Lu, "HuAc: Human activity recognition using crowdsourced WiFi signals and skeleton data," *Wireless Commun. Mobile Comput.*, vol. 2018, Jan. 2018, Art. no. 6163475.
[33] T. Z. Chowdhury, C. Leung, and C. Y. Miao, "WiHACS: Leveraging WiFi for human activity classification using OFDM subcarriers' correlation," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Montreal, QC, Canada, Nov. 2017, pp. 338–342.

Huan Yan was born in Guizhou, China, in 1995. He received the B.Eng. degree from the Hefei University of Technology, where he is currently pursuing the M.Sc. degree.

His research interests include intelligent information processing and wireless sensing and machine learning.

Yong Zhang was born in Anhui, China, in 1973. He received the Doctor's degree in signal and information processing from the University of Science and Technology of China in 2007. He is currently an Associate Professor with the Hefei University of Technology of China.

His research interests include intelligent information processing, gesture recognition, and indoor positioning.



Yujie Wang was born in Shandong, China, in 1980. She received the Ph.D. degree in circuits and systems from the University of Science and Technology of China in 2011.

She is currently a Lecturer with the Hefei University of Technology. Her research interests include intelligent information processing, indoor positioning, and audio signal processing.

Kangle Xu was born in Anhui, China, in 1994. She received the B.Eng. degree from Anhui Polytechnic University. She is currently pursuing the M.Sc. degree with the Hefei University of Technology.

Her research interests include intelligent information processing and gesture recognition.