IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 14, 2019, accepted October 2, 2019, date of publication October 17, 2019, date of current version October 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2948102

Wi-Motion : A Robust Human Activity Recognition Using WiFi Signals

HEJU LI[1], XIN HE[1,2], (Member, IEEE), XUKAI CHEN[1], YINYIN FANG[1], AND QUN FANG[1]
[1]School of Computer and Information, Anhui Normal University, Wuhu 241002, China
[2]School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China

Corresponding author: Xin He (xin.he@ahnu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61702011, in part by the Anhui Provincial Science Foundation under Grant 1808085QF191, and in part by the CERNET Innovation Project under Grant NGII20180118.

ABSTRACT Recent research has shown that human motions and positions can be recognized through WiFi signals. The key intuition is that different motions and positions introduce different multipath distortions in WiFi signals and generate different patterns in the time-series of channel state information (CSI). In this paper, we propose Wi-Motion, a WiFi-based human activities recognition system. Unlike existing systems, Wi-Motion jointly leverages the amplitude and phase information extracted from the CSI sequence. We first construct the classifiers using amplitude and phase, respectively. The output of classifiers is then combined by a posterior probability-based combination strategy. As the simulation results show, Wi-Motion can recognize predefined 5 typical human activities with the mean accuracy of 96.6% in line-of-sight (LOS) environment, and 92% in not line-of-sight (NLOS) environment. Furthermore, Wi-Motion evaluates the effect of the age of the experimental subjects and relatively complex environments.

INDEX TERMS WiFi signals, human activity recognition, posterior probability combination.

I. INTRODUCTION

The machine-centric computing model is shifting toward a people-centric computing model [1], [2], where it is critical to precisely sense and recognize human activities for supporting the people-centric model. In the framework of people-centric computing model, a fruitful of recognizing techniques has been proposed to recognize human activities. These conventional methods can be categorized into three groups: vision-based, low-cost radar-based, and wearable sensor-based approaches. However, there are several limitations in deploying conventional techniques to sense human activities. Vision-based approaches can only operate within a certain range of line-of-sight (LOS) environments. They are susceptible to lighting conditions, obstacles, and suffering from the problem of dead angles. This also arises many concerns on human privacy issues. Low-cost radar-based systems have limited operation distances, which is usually in the order of ten centimeters. Wearable sensor-based solutions although achieve fine-grained behavioral awareness, its high cost and restriction on real-time nature make it not practical in some applications (e.g. rescue applications).

In recent years, with the wide deployment of WiFi access points, the WiFi signals become ubiquitous, particularly, in indoor scenarios. Due to the unguided property of radio signal propagation, radio signals are traveling freely in the atmosphere which may reflect by the wall and/or other objects. At the receiver, antennas thus receive the signals from two or more paths, which is a so-called multipath phenomenon. This phenomenon also applies to WiFi signals. The key insight of human activity recognition through the WiFi signal is that the moving body affects the multipath propagation and different moves have dissimilar effects. We can then analyze the received signals, in particular, amplitude and phase, to recognize human activities. Amplitude and phase information is already contained in the channel state information (CSI), which is a complex value representing the amplitude attenuation and phase shift of multiple paths. Therefore, the question is how to obtain the CSI values and extract the amplitude and phase information to generate a unique pattern of human activities.

Some commercial WiFi devices (e.g. IWL5300 wireless network card) where signals are transmitted and received over

The associate editor coordinating the review of this manuscript and approving it for publication was Hayder Al-Hraishawi.

(a) Bend (b) Halve Squat (c) Step (d) Stretch Leg (e) Jump

FIGURE 1. Five human activities recognized by Wi-Motion.

multiple subcarriers using multiple antennas at transmitter and receiver sides can provide us a fine-grained[1] CSI in time-series. Because of the high data rate provided by these modern commercial WiFi devices, we can get enough samples of CSI measurements within the duration of human activities. Thanks to these systems, researchers are able to focus on analyzing CSI to recognize various types of human activities.

Existing work utilizes either the amplitude change in time domain or phase shift in spatial and frequency domains of the CSI values. The amplitude change usually generates different patterns for different humans, activities, gestures, etc. It thereby can be used for human presence detection [3], [4], fall detection [5], [6], motion detection [7], [8], activity recognition [9], [10], and human identification/ authentication [11], [12]. On the other hand, CSI phase shift highly associated with the signal transmission delay and direction, which can be usually adopted for human localization [13] and tracking [14], [15]. These systems follow the general architecture of machine learning-based classification systems and have four stages: data collection, noise removal, feature extraction, and classification.

In this paper, we propose a WiFi-based human activity recognition system, namely Wi-Motion, which can sensitively recognize predefined 5 different human activities, as shown in Fig. 1. To summarize, the contributions of this paper are shown as follows:

- Unlike most human body recognition systems, which only process amplitude information or phase information, we jointly utilize the amplitude and the phase information in the CSI sequence collected from commercial WiFi devices with three antennas, to improve the recognition accuracy.
- To split the static environment and the waveform caused by human activity, we propose an activity extraction method combining a sliding window with a threshold-based image segmentation. For phase and amplitude information, we adopt different learning methods to extract features and design classifiers.
- After getting the recognition results at each classifier, we generate prediction results based on the output posterior probability of two classifiers. To verify the effectiveness of the combining algorithm, we conduct a large number of comparative experiments. According to the experiment results, the solutions we proposed increase recognition accuracy indeed.

The rest of this paper is organized as follows. We present the related work in Section II, followed by a preliminary in Section III. Then we elaborate on the system overview and data preprocessing methods of Wi-Motion in Section IV and V, respectively. We present the classification algorithms and implementation in Section VI and VII, and finally conclude our work in Section VIII.

II. RELATED WORK

We review several related existing works in this section.

A. DEDICATED HARDWARE-ENABLED APPROACHES

WiSee [16] used the dedicated USRP devices as WiFi transceivers on a 10 MHz channel at center frequency 5 GHz. The system can recognize 9 actions by extracting the Doppler shift from the WiFi signal caused by human motion as a feature, with an accuracy rate of 94%. Adib *et al*. designed WiTrack and WiTrack2.0 that applied designed special carrier wave radio to track human movements behind a wall [17].

B. RSSI-BASED APPROACHES

Bahl *et. al.* proposed Radar [18], which is a system for indoor localization based on received signal strength indicator (RSSI). Sigg *et. al.* used USRPs as specialized hardware devices to capture RSSI values from WiFi signals [19], [20]. They utilized RSSI values of WiFi signals to recognize 4 activities including lying down, crawling, standing and walking and achieved over 80% recognition accuracy for these 4 activities. Abdelnasser *et al*. proposed WiGest [21], a gesture recognition system based on the RSSI obtained by commercial WiFi access points. More specifically, WiGest performed gesture recognition by analyzing the rising and falling edges of RSSI signal variations. The accuracy can reach about 87.5% in the case of a single access point, and it increases to 96% when there are three access points. However, since RSSI only provides coarse-grained information about channel variations, it is often affected by multipath effects and noise.

[1]The fine-grained CSI means that CSI provides much more information than received signal strength indicator.

C. CSI-BASED APPROACHES

Compared to RSSI, CSI is a fine-grained value from the physical layer [22], [23], which provides a channel estimation of each subcarrier for each transmission link, i.e., it describes the amplitude and phase on each subcarrier in the frequency domain. Moreover, CSI can reflect the effect of small scale fading and multipath effects caused by micro-movement [22]. WiHear [24] interpreted the transmitted signal gathered on the mouth via beamforming technique and got the changes in the mouth shape by analyzing reflected signals. The pronunciation was represented by the mouth type, thereby it is possible to implement WiFi-based lip recognition system. WiFinger [23] extracted the fixed pattern of gesture signals through principal component analysis (PCA), and used it as a feature to identify gestures with an accuracy of 93%. Ali *et al*. proposed WiKey that uses CSI values obtained from commercial off-the-shelf (COTS) devices to recognize keystrokes [25]. Zheng *et al*. built a novel non-intrusive smoking detection system, namely Smokey [26], that is able to accurately detect the smoking activities by exploiting the impact of smoking on the CSI of WiFi signals. Shang *et al*. proposed a WiFi signal-based sign language recognition system WiSign [22]. Different from other systems, WiSign used three WiFi devices to improve the recognition performance. SpotFi [27] incorporated super-resolution algorithms to compute the angle of arrival (AoA) of multipath components and to identify AoA of a direct path between the localization target and access point by incorporating novel filtering and estimation techniques. It can achieve a median accuracy of 40 cm in a rich multipath indoor environment. In addition, a breathing detection was investigated using passive WiFi radar in [28], which propose a real-time phase extraction method to detect slow moving and small body movement. It shows good performance in both LOS and not line-of-sight (NLOS) scenarios using WiFi beacon and data packets. With the inspiration of the above-mentioned works which splendid research specialized in specific application scenarios, we propose a framework of WiFi-based activity recognition to improve the human activity robustness.

III. PRELIMINARY

In general, different human activities may cause different multipath distortions in WiFi signals. We can then leverage the multipath property in the received signal to recognize human activities.

In communications, the multipath propagation of wireless channel can be described by its channel impulse response (CIR), as [29]

$$h(\lambda;t)=\underbrace{\sum_{n\in p^s}a_n\delta(\lambda-\lambda_n)}_{h_{static(\lambda)}}+\underbrace{\sum_{m\in p^d}a_m(t)\zeta(f)\delta(\lambda-\lambda_m(t))}_{h_{dynamic(\lambda;t)}} \tag{1}$$

where $h(\lambda;t)$ represents the reaction of the channel in response to an ideal Dirac pulse $\delta(\lambda)$. Due to the presence of multiple propagation paths, the receiver could receive more than one pulse, and each of them may have different transmission delay. For easily understanding how the human activity affect the multipath propagation, we divide the multiple paths into two groups, static p^s and dynamic p^d. The static paths labeling by n are not affected by the human activity and hence have constant delay λ_n and attenuation a_n. Therefore, they are not attracted by us. However, dynamic paths $\forall m \in p^d$ will experience a time-varying change if a person is moving, representing by the variations on both time-varying propagation delay $\lambda_m(t)$ and signal attenuation $a_m(t)$ of each path $m \in p^d$. Moreover, $\zeta(f)$ represents the frequency dependent absorption cross section coefficient, which depends on the body specific signal absorption [30].

For the time-varying impulse response $h(\lambda;t)$, we define a time-varying channel frequency response (CFR) $H(f;t)=\int_{-\infty}^{\infty}h(\lambda;t)e^{-2j\pi f\lambda}d\lambda$ in frequency domain by make a Fourier transform of $h(\lambda;t)$. The CFR is also known as the CSI and has static and dynamic components, as

$$H(f;t)=\underbrace{\sum_{n\in p^s}a_ne^{-2j\pi f\lambda_n}}_{H_{static(f)}}+\underbrace{\sum_{m\in p^d}a_m(t)\varsigma(f)e^{-2j\pi f\lambda_m(t)}}_{H_{dynamic(f;t)}} \tag{2}$$

where the dynamic component $H_{dynamic(f;t)}$ is our mainly focus which contains the wireless channel variations caused by human activities. Therefore, $H_{dynamic(f;t)}$ can be regarded as a unique indicator of different human activities. In particular, we are able to collect $H_{dynamic(f;t)}$ from commercial WiFi devices between a pair of transmitting antenna T_a and receiving antenna R_a at the subcarrier i with a central frequency f_i, which is defined by

$$H_{T_a,R_a}(f_i;t)=\|H_{T_a,R_a}(f_i;t)\|e^{j\angle H_{T_a,R_a}(f_i;t)} \tag{3}$$

where $\|H_{T_a,R_a}(f_i;t)\|$ and $\angle H_{T_a,R_a}(f_i;t)$ denote its amplitude and phase, respectively.

IV. SYSTEM OVERVIEW

In this section, we elaborate on the design of Wi-Motion. Wi-Motion is a recognition system that enables commercial WiFi devices to identify user's activities using CSI measurements collected from commercial WiFi devices. The flows of Wi-Motion are illustrated in Fig. 2. The whole system has two stages: data processing and classification.

A. DATA PROCESSING

Firstly, a regular WiFi signal affecting by human activities is acquired from a commercial WiFi device. Secondly, the collected signal, which can be separated into amplitude and phase information, are respectively preprocessed using signal processing methods, such as filtering and linear transformation. The purpose of preprocessing is to reduce the effect of noise. After that, we would like to reduce the dimension of the processed data in terms of the number of subcarriers. The reason is twofold; (1) Different subcarrier

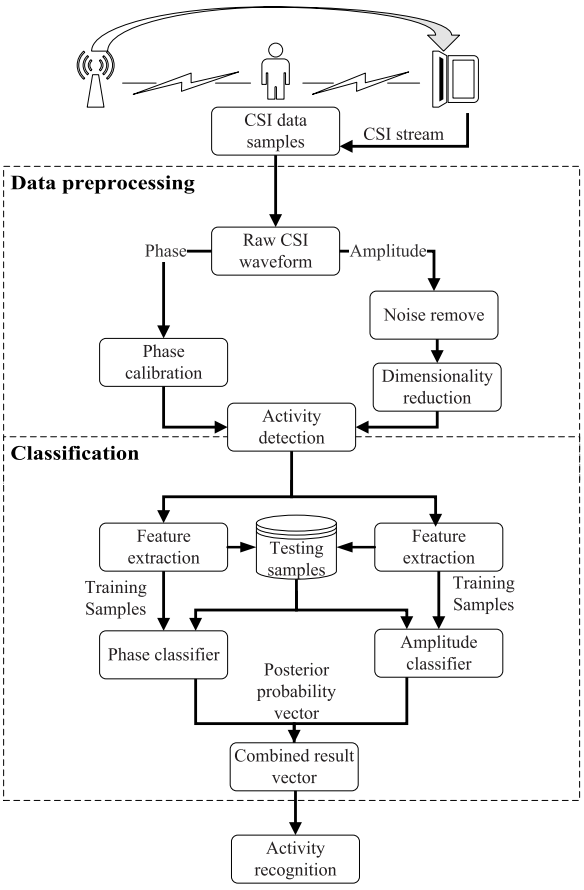


FIGURE 2. System structure of Wi-Motion.

has different sensitivity to the human activities, i.e., some subcarriers are very sensitive to the motion and have relatively evident variations. It is better to only use the CSI data from these sensitive subcarriers. (2) Using all the data from the entire subcarriers also increases the computational complexity of the system. To this end, we analyze the obtained CSI sequence which contains the effect by not only the valid activity but also the environment, and masterly extract the signal segments mainly corresponding to the human activity.

B. CLASSIFICATION

In the classification stage, useful features which can represent the relationship between the time-series of CSI and different human activities are extracted from the processed amplitude and phase information respectively, as a basis for classification. We randomly select parts of feature vectors and utilize the support vector machine (SVM) algorithm to build two classifiers according to amplitude and phase information. When an unknown activity sample enters, consulting the prediction results of both two classifiers, Wi-Motion performs a merge method based on the posterior probability to produce the final recognition.

V. DATA PREPROCESSING

A. PHASE INFORMATION PROCESSING

1) PHASE ANALYSIS

As discussed in Section III, CSI measurements provide the phase information of each subcarrier. The separated phase information $\hat{\phi}_i$ (this paper replaces the phase symbol $\angle H_{T_a,R_a}$ in (3) with $\hat{\phi}$ in order to facilitate the expression.) for the i^{th} subcarrier can be expressed as

$$\hat{\phi}_i = \phi_i - 2\pi \frac{k_i}{N}\vartheta + \beta + Z. \tag{4}$$

where ϕ_i denotes the true phase, ϑ is the timing offset at the receiver, which causes phase error expressed as the middle term, β means an unknown phase offset, and Z indicates some measurement noise. k_i signifies the subcarrier index (ranging from -28 to 28 in IEEE 802.11n) of the i^{th} subcarrier and N represents the fast fourier transformation (FFT) size (which is 64 in IEEE 802.11 a/g/n). Due to the unknowns listed above, it is impracticable to obtain the true phase shifts with solely commercial WiFi devices.

2) PHASE CALIBRATION

To mitigate the effects of random noise, we execute a linear transformation on the raw phases, as recommended in [31]. The key thought is to remove ϑ and β by considering phase across the entire frequency band. Firstly, we define two intermediate variables a and b as

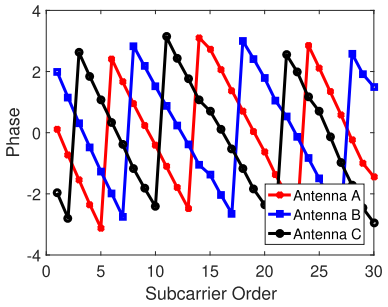
$$a = \frac{\hat{\phi}_n - \hat{\phi}_1}{k_n - k_1} = \frac{\phi_n - \phi_1}{k_n - k_1} - 2\pi\frac{\vartheta}{N}. \tag{5}$$

$$b = \frac{1}{n}\sum_{i=1}^{n}\hat{\phi}_i = \frac{1}{n}\sum_{i=1}^{n}\phi_i - \frac{2\pi\vartheta}{nN}\sum_{i=1}^{n}k_i + \beta. \tag{6}$$

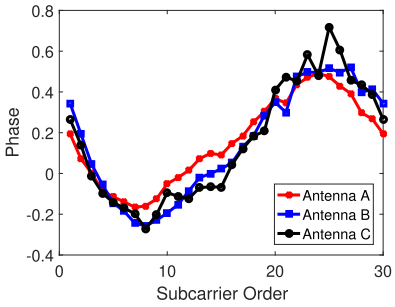
Considering that subcarrier frequency is symmetric, it means that $\sum_{j=i}^{n} k_i$ is vanished and b is simplified as $b = \frac{1}{n}\sum_{i=1}^{n}\phi_i + \beta$. Furthermore, by subtracting the linear term $ak_i + b$ from the raw phase $\hat{\phi}_i - ak_i$ in (4), we can get a linear combination of true phases, denoted as $\tilde{\phi}_i$, where the random phase offsets have been eliminated (omitting the small measurement noise Z).

$$\tilde{\phi}_i = \hat{\phi}_i - ak_i - b = \phi_i - \frac{\phi_n - \phi_1}{k_n - k_1}k_i - \frac{1}{n}\sum_{j=1}^{n}\phi_j. \tag{7}$$

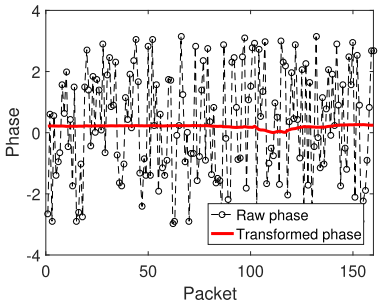
Although the above (7) can be used for calibrating phase information, the raw phase is folded due to the recurrence characteristic of phase, which requires us to map the raw phase into the true value. Fig. 3(a) shows the raw phase values of CSI for the three antennas at the receiver. What we can clearly see is that the raw phase of each of the three antennas is folded with the increase of subcarrier order and the range of the phase is $[-\pi, \pi]$. To obtain the true phase, the folded phase can be recovered by subtracting multiple 2π. Thus, we perform a phase calibration algorithm proposed in [32]. Fig. 3(b) shows the transformed phase values for three different antennas. It is noticed that the range of the transformed phase becomes much smaller than the raw phase for



(a) Raw phase values for three different antennas.



(b) Transformed phase values for three different antennas.



(c) Random noise are removed in sanitized phase information of a stretch leg sample.

FIGURE 3. Processing of raw phase information.

three antennas. Fig. 3(c) makes a comparison of unprocessed raw phase and transformed phase information of the first subcarrier of a stretch leg sample. As can be seen, the phase without further calibration distributes extremely randomly. But after calibration, it is relatively stable as expected.

B. AMPLITUDE INFORMATION PREPROCESSING

1) NOISE REMOVAL

Similarly, the raw amplitude measurement obtained from raw CSI sequence is usually not reliable enough to be used for feature extraction, because of the effect of noise, originating from environmental changes, radio signal interference, etc. In our system, when the outliers are removed by the Hampel [33], we utilize low-pass filtering to remove high-frequency noise and further introduce weighted moving average (WMA) method to process the raw amplitude waveform. The raw amplitude sequence of the first subcarrier at time t is denoted by $\{\hat{\alpha}_{1,1}, ..., \hat{\alpha}_{t,1}\}$ (this paper replaces the amplitude symbol $\|H_{T_a,R_a}\|$ in (3) with $\hat{\alpha}$ in order to facilitate the expression.). According to the WMA algorithm, we process the raw amplitude data as

$$\tilde{\alpha}_{t,1} = \frac{(m \times \hat{\alpha}_{t,1} + ... + 1 \times \hat{\alpha}_{t-m-1,1})}{m + (m-1) + ... + 1} \tag{8}$$

with m being a window size and the largest weighting factor. The filter output is denoted by $\tilde{\alpha}$. It can be easily found that the most recent amplitude is assigned the highest weight. In this paper, we empirically set $m = 10$. Fig. 4 shows the original waveform of a stretch leg sample and the waveform after WMA filtering. The comparison shows that WMA filtering can remove most of the noise, which makes waveform smoother.

The processing of the amplitude sequence from other subcarrier is the same, which is omitted.

2) DIMENSIONALITY REDUCTION

The IWL5300 provides 802.11n CSI in a matrix format including the channel estimation for a group of 30 subcarriers. At each subcarrier, the CSI describes how a signal propagates from the transmitter to the receiver with the combined influence of, for example, scattering, fading, and power decay

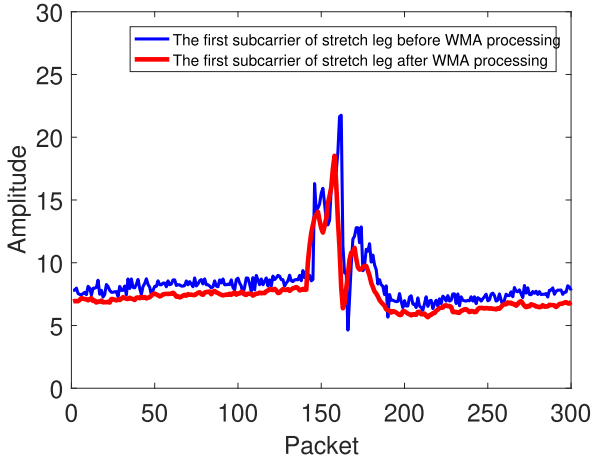


FIGURE 4. The original waveform of a stretch leg sample and the waveform after WMA filtering.

with distance. After filtering noise, we can get a relatively accurate amplitude matrix of each activity sample. However, if all subcarriers are used to perform the following operations, it will cause a high computational complexity of the system. On the other hand, in some existing works, e.g. [23], they already stated that different subcarriers had different sensitivity to the same activity. Hence, the CSI information from different subcarriers will experience different changes. In some subcarriers, the changes are relatively small because these subcarriers are not sensitive to human activities. In contrast, some subcarriers are largely changing caused by human activities. The changes on the CSI value can be represented by the variance of the CSI sequence, where high variance corresponds to the large variations. Therefore, we use the variance as an indicator of the sensitivity of the subcarrier with respect to human activities. Fig. 5 illustrates an example of time series CSI changes for 30 subcarriers when a user performs a stretch leg activity, we notice clearly that the variance of different subcarriers shows a big difference when performing the same activity. Specifically, including these subcarriers with high sensitivity to noise but very low sensitivity to human activity may bring unpredictable effects (usually, negative effect) to the following classification processes.

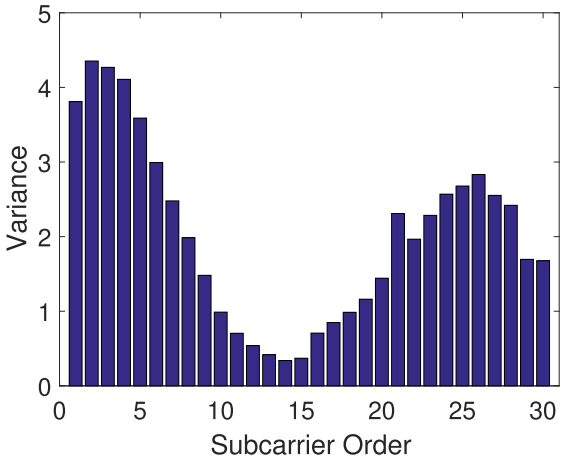


FIGURE 5. Different subcarriers have different sensitivities for the same activity.

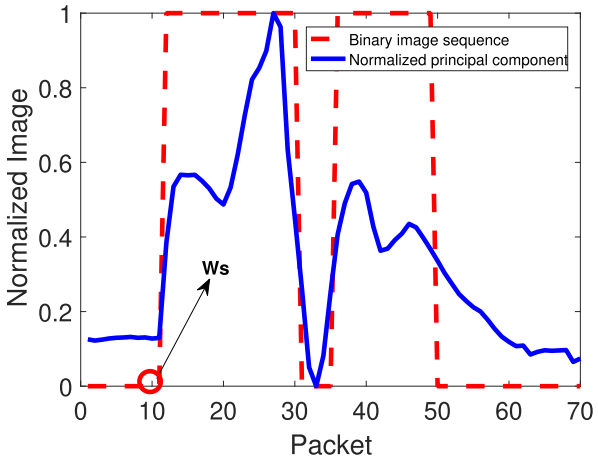


FIGURE 7. The generating binary image inside the approximate range of the first principal component sequence of a stretch leg sample.

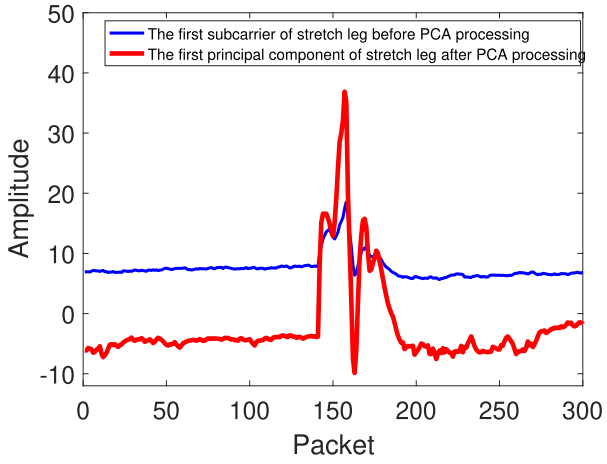


FIGURE 6. The first principal component after PCA processing.

Therefore, reducing the data dimension and eliminating these non-significant subcarriers are very important. In this paper, we leverage PCA algorithm to reduce the dimensions of the CSI sequence and eliminate redundant information remaining in the data sequence. Based on our experiment results, we finally choose the first principal component waveform for subsequent operations, which is shown in Fig. 6.

C. ACTIVITY DETECTION

1) FIRST PRINCIPAL ANALYSIS

After previous processing steps, we can obtain high-quality CSI data, which contains not only valid activity information, but also static environment information when there is no activity occurs. So what we need to do next is to analyze the extracted principal component and pick up the signal segments corresponding to human activities. According to the results of previous research [23], the main effects of human motions on the CSI sequence are either rising edges, falling edges, or pausing. Wi-Motion exploits an activity extraction method which combines sliding window and threshold image segmentation, and it can be used to detect the waveform segment corresponding to human activity in both amplitude and phase information waveforms.

2) ACTIVITY DETECTION

For amplitude information, we extract the activity from the first principal component after dimension reduction. Firstly, setting the sliding window length to 10, we get a window matrix and immediately calculate the variance of each sliding window as $[V_a(1), ..., V_a(l)]$, where l denotes the number of sliding windows. Then, by using a threshold T, an approximate position range (A_s, A_e) of motion waveform can be estimated, where A_s and A_e represent the starting and ending points of the range. What we can imagine is that the obtained sequence can be regarded as a "two-dimensional image". Based on this hypothesis, we transform and convert normalizing sample image into a binary image utilizing optimal threshold computed using maximal variance between clusters (MVBC) algorithm proposed in [34]. Fig. 7 shows a binary image of the approximate range. Subsequently, we refer the time when the rising or falling edge that first appears into the pseudo starting time W_s, indicating by the red circle. By utilizing W_s, we obtain an intermediate split time S_t, which is computed as $S_t = A_s + W_s$ to separate the above-mentioned approximate range into two parts (A_s, S_t) and (S_t, A_e). For the region within (A_s, S_t), the above processing is similarly performed to obtain another binary image. Synchronously, we consider the rising or falling edge time in the binary image as the preliminary starting time, expressed by P_s. Similarly, the last time point where rising or falling edge occurs in the binary image of another part (i.e., (S_t, A_e)) can be regarded as preliminary ending time P_e. Fig. 8 shows these two preliminary time points. However, if the variance of the pixel values is relatively high, the starting and the ending points determined by MVBC are not the one what we expect for. The reason is that the threshold in MVBC is mainly affected by the pixel values, rather than the variance. For this case, we perform sliding variance processing on the

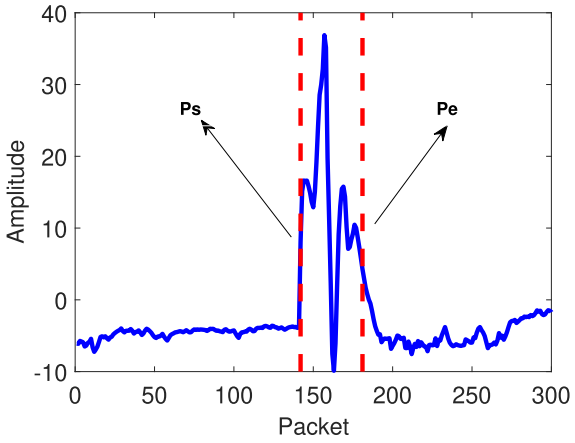


FIGURE 8. The preliminary start time and the preliminary end time.

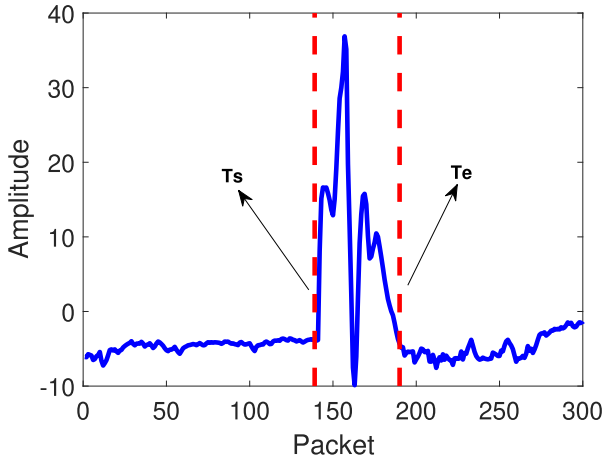


FIGURE 9. Start and end points detected by Wi-Motion.

sequences in (A_s, P_s) and (P_e, A_e) respectively, where the window size L is 3 and the threshold is the average of variance sequence. Fig. 9 gives the final result detected by Wi-Motion using the above method. Two red dotted lines respectively represent the true starting point T_s and true ending point T_e of activity. Experimental results clearly show that our method can accurately extract the activity waveform.

Similarly, for phase information, we can also extract the motion waveform using the above method by setting suitable parameters, such as window length, threshold, etc.

VI. CLASSIFICATION

A. FEATURE EXTRACTION

1) AMPLITUDE FEATURE EXTRACTION

Discrete wavelet transform (DWT) can analyze signals on multiple frequency scales and has a better extraction ability for local features. Considering the speed on different body parts may be different, direct extraction in time domain will lose a lot of detail related to human activity. Through the wavelet transform, the wavelet coefficients of each frequency band can be obtained. Specifically, we perform DWT on

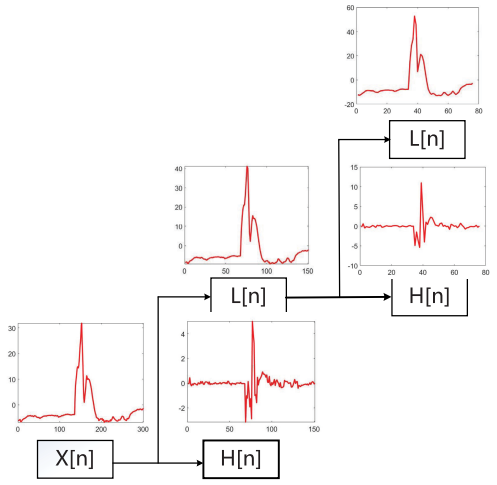


FIGURE 10. Complete process of discrete wavelet transformation.

the extracted amplitude waveform based on the first-order Daubechies wavelet, where the decomposition layer number is 2. Then, the approximate coefficients of the last layer are taken out and the normalized coefficient sequence can be used as the feature vector. By the DWT process, the contour information of the amplitude waveform preserves completely in the feature vector, and the remaining noise is suppressed as the detail coefficients are discarded. The complete binary tree of the DWT process can be shown in Fig. 10.

2) PHASE FEATURE EXTRACTION

For the phase information matrix, which includes phase information of 30 subcarriers, is also necessary for us to remove residual noise. In this paper, Wi-Motion further introduces WMA method into the transformed phase information to obtain more accurate data that can be used for feature extraction. Subsequently, we sort the subcarriers according to the variance of phase sequence in each subcarrier. After selecting the top 20 (which is an empirical value) subcarriers for PCA process, the obtained first component is chosen for constructing the classifier.

B. CLASSIFIER TRAINING

We select a high-effective SVM classification to recognize 5 activities according to the performance of existing works. As is known to all, the choice of kernel function plays a key role in the performance of classical SVM. For example, a Gaussian kernel function k_ζ that is simple in form and widely used can be defined as

$$k_\zeta(x, x_c) = \exp[- \| x - x_c\|^2/2\sigma^2]. \tag{9}$$

where vector x_c represents the center of the kernel function, σ is a width parameter which controls the radial extent of the function, and $\| x - x_c\|^2$ denotes the Euclidean distance of any vector x to the center of the kernel function.

For our extracted features, we find that the feature vectors of different human activities may not share the same

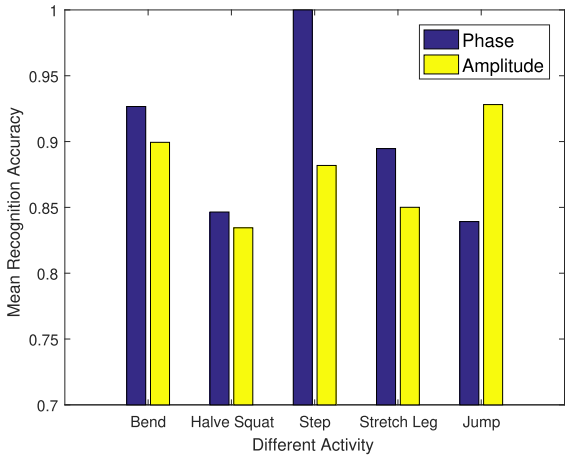


FIGURE 11. Classification performance of two classifiers.

length, so the traditional SVM algorithm, which requires the dimension of the feature vector to be consistent cannot be nicely applied in classifying our features. Under this situation, Wi-Motion further exploits dynamic time warping (DTW) to calculate the distances among feature vectors. In contrast to Euclidean distance, DTW offers intuitive distance between two waveforms and can be resilient to signal distortion or shift. DTW distance is the Euclidean distance of the optimal warping path between two waveforms calculated under boundary conditions and local path constraints [35]. DTW aims to compare two time-dependent series which can be discrete signals (time series) or, more generally, feature sequences sampled at equidistant points in time. In this paper, Wi-Motion replaces the Euclidean distance in the Gaussian kernel function by using DTW distance to construct a new kernel function

$$k_\ell(x, x_c) = \exp[-D_\ell(x, x_c)^2/2\sigma^2]. \tag{10}$$

where k_ℓ represents the DTW kernel function, and D_ℓ denotes the DTW distance. Finally, we classify our feature vectors using the SVM model with the kernel function defined in (10).

C. PREDICTION

In our experiments, we collect the CSI samples of 5 different activities (''bend'', ''halve squat'', ''step'', ''stretch leg'' and ''jump'') to test our two classifiers. Results are shown in Fig. 11. Obviously, for an activity like ''step'', we can use phase features individually to perfectly classify it. However, for an activity like ''jump'', the classification performance of the phase classifier is not satisfactory. Conversely, the amplitude classifier performs well for ''jump'' activity, but bad for recognizing ''step'' motion. To tackle this, we propose a joint prediction algorithm based on the output from two classifiers, which outperforms the case of only using a single classifier.

Traditional combination algorithms, such as boosting algorithm, multiple decision method, etc., facing higher complexity and requiring at least three classifiers. However, in our experiment, there are only two classifiers, which leads to that the traditional combination algorithms are not suitable for Wi-Motion system. In WiSign [22], Shang *et al.* proposed weighted voting on two laptops and got the final prediction result, where they combined two prediction vectors of classifiers on two laptops instead of choosing the result with the highest confidence on one laptop. Inspired by their prediction mechanism, we propose a combination strategy based on the output of two SVM classifiers. The output is given in the form of posterior probability, representing the probability of each activity. More specifically, given training samples ξ_j where $j = 1..., s$, labeled by $y_j \in (+1, -1)$, the binary SVM computes a decision function $f(\xi)$ such that sign$(f(\xi))$ can be used to predict the label of any test sample ξ. According to the method proposed by Platt *et al.* [36], the SVM standard output value can be mapped to [0, 1] using the Sigmoid function to obtain the SVM posterior probability, as

$$P(y = 1|\xi) \approx P_{\psi,\omega}(f(\xi)) = \frac{1}{1 + \exp(\psi f(\xi) + \omega)}. \tag{11}$$

where $P(y = 1|f(\xi))$ indicates the posterior probability on each class, ψ and ω are parameters that need to be optimized and can obtain using the training set for maximum likelihood estimation. Let each τ_j be an estimate of $f(\xi_j)$. The target model function can be expressed as following equation

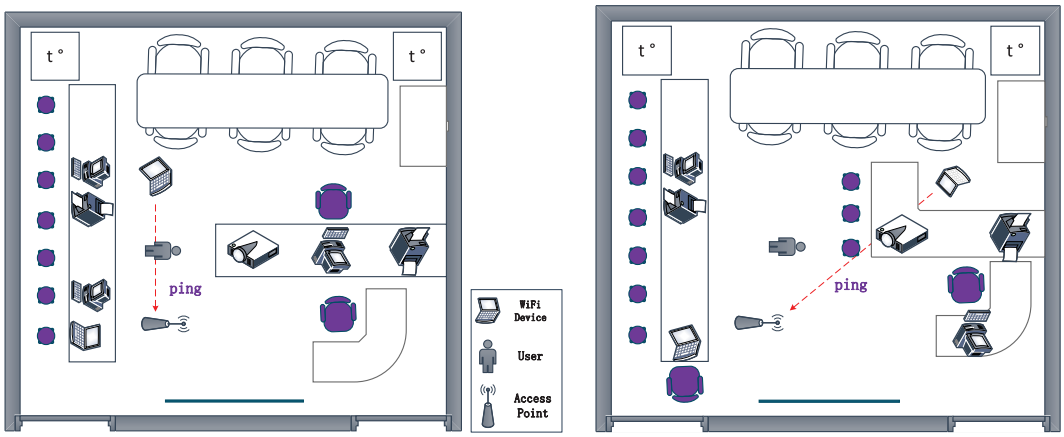
$$\min_{z=(\psi,\omega)} F(z) = -\sum_{j=1}^{s} [t_j \log(p_j) + (1 - t_j) \log(1 - p_j)]. \tag{12}$$

with

$$t_j = \begin{cases} \dfrac{s_+ + 1}{s_+ + 2}, & \text{if } y_j = +1 \\ \dfrac{1}{s_- + 2}, & \text{if } y_j = -1 \end{cases} \tag{13}$$

where (ξ_j, y_j) represents the training sample (with s_+ of the y_j's positive and s_- negative), $p_j = P_{\psi,\omega}(f(\xi_j))$. In our experiments, we extend the two-class SVM probability-based to the multi-class in a one-to-one manner using the method completed in [37].

After an unknown sample enters, two classifiers predict it and generate a posterior probability vector respectively. Immediately after, Wi-Motion adds these two prediction vectors and gives the final prediction. For example, assuming the prediction vector reported by the two classifiers are (0.1, 0.2, 0.78, 0.9, 0.27) and (0.12, 0.2, 0.87, 0.14, 0.24), we can clearly see that the first classifier cannot distinguish the third and the fourth activity (since the difference between 0.78 and 0.9 is too small). If we always choose the result with the highest confidence on one classifier, we possibly make a wrong decision. But if we combine these two prediction vectors, we can get (0.22, 0.4, 1.65, 1.04, 0.51). Based on the final combined prediction vector, the correct prediction (third activity) can be obtained.



(a) Environment A of data collection (LOS). (b) Environment B of data collection (NLOS).

FIGURE 12. Environment set of data collection.

TABLE 1. Characteristics of test users.

ID	Gender	Age	Height [cm]	Weight [kg]	ID	Gender	Age	Height [cm]	Weight [kg]
1	Male	23	180	65	6	Female	24	169	53
2	Male	23	178	70	7	Male	22	175	62
3	Male	24	180	69	8	Male	29	177	70
4	Male	23	176	64	9	Female	42	162	57
5	Female	23	165	47	10	Female	51	161	55

VII. IMPLEMENTATION AND EVALUATION

A. ACTIVITY DATASET GENERATION

In order to ensure that our system can operate in any environment, two complex office environments expressed in Fig. 12 are chosen. In our paper, we firstly build a dataset for 6 users, and each of them provides five human activities, as shown in Fig. 1. The basic characteristics of test users are presented in Table 1. We use a commercial TP-Link wireless router as the transmitter operating in the IEEE 802.11n AP mode at 2.4GHz. An Acer Aspire EC laptop running Ubuntu 14.04 is used as a receiver, which is equipped with off-the-shelf Intel 5300 card (three antennas) and the modified firmware. During the process of receiving WiFi signals, the receiver pings the router 33 pkts/s and records the CSI of each packet. For each activity in different environments, every user provides 30 instances to evaluate the performance of our system. We randomly select some activity samples from the dataset as the training set and the rest as the testing set. Specifically, a random sequence generator is adopted to determine the serial number of selected samples for each selection process. The impact of the number of training samples on overall recognition will be discussed in Section VII-D.

B. ACTIVITY RECOGNITION ACCURACY

Figure 13 shows the mean prediction accuracy of our prediction combination model and classification model on each classifier of user 1. We can clearly see that our system

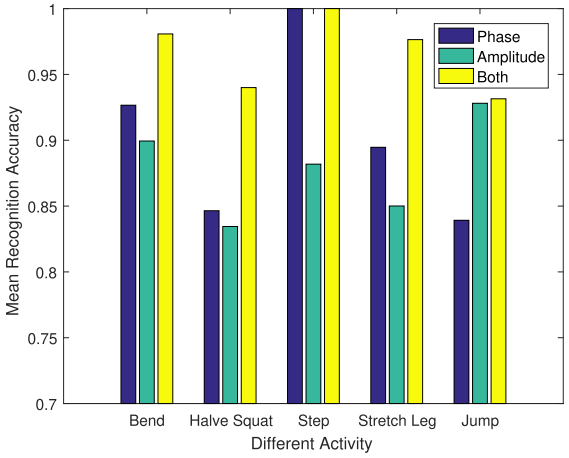


FIGURE 13. Prediction accuracy of different activities.

improves recognition performance for all supported activities. For example, "halve squat" is unsatisfactory to classify on both amplitude and phase classifier, however, after merging the prediction results of two classifiers, our system has a great prediction accuracy of 94.7%. For "step", both the phase classifier and our system achieve an accuracy of 100%, while the amplitude classifier only has a prediction accuracy of 87.9%. For "stretch leg", the prediction accuracy is respectively 85% and 89.5% for the amplitude and phase classifier, while our system has a better result of 97.8%. In short,

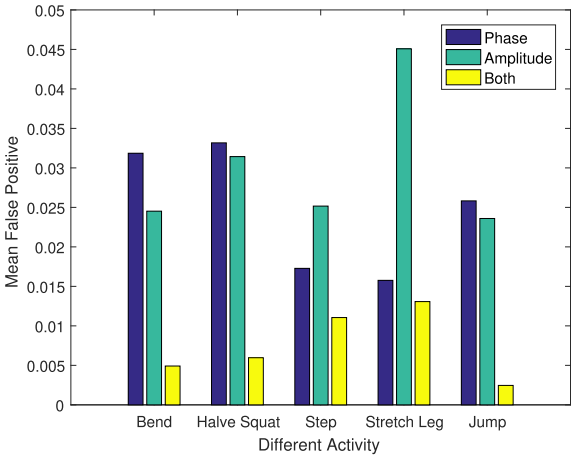


FIGURE 14. False positives of different activities.

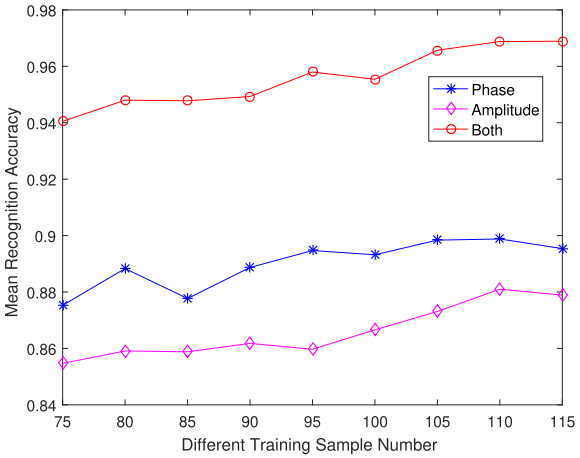


FIGURE 16. The influence of the different number of training samples.

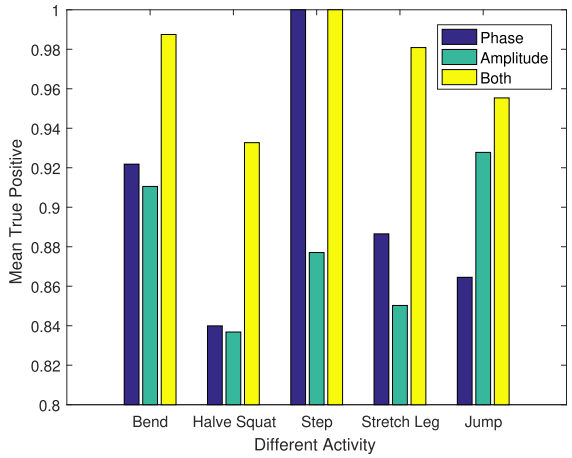


FIGURE 15. True positives of different activities.

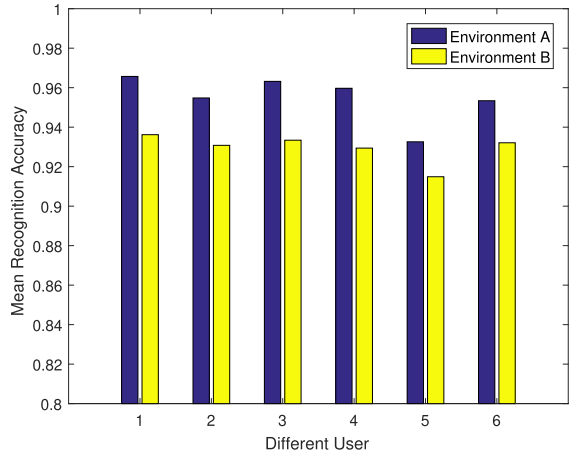


FIGURE 17. The influence of different users on different environments.

experiment results show that we can get more accurate activity estimation by combining output posterior probability of two classifiers.

C. FALSE POSITIVE AND TRUE POSITIVE

In order to test the recognition performance of Wi-Motion, we further explore the false positive and true positive of each activity supported. We use the dataset collected from user 1, and the evaluation results are illustrated in Figs. 14 and 15. It is clear that the false positive of prediction can be improved to about 0.75% and the true positive rate generated by the combined model reaches 97.1% by combining two classifiers, which is significantly higher than that produced by the single classifier. To summarize, the combination algorithm we proposed reduces the recognition errors and facilitates the recognition of activity.

D. DIFFERENT NUMBER OF TRAINING SAMPLES

In spite, the activity samples Wi-Motion requires are not on a very high magnitude, according to our experiment results, we notice that the number of training samples has a certain impact on the recognition accuracy. Fig. 16 shows our experimental results which generate based on the dataset collected from user 1. Obviously, with the increase of training samples, the classification accuracy of the amplitude and phase classifier has a certain degree of increase. The reason is that the more training samples the richer the information, i.e., the hyperplane position of the SVM is more accurate. However, the gain of our combination system is relatively small because the average recognition accuracy of our system is already at a very high level. What we all know is that the more training samples, the longer training time the SVM system takes, inevitably leading to an increase in system complexity. Therefore, we set the number of samples used for training to 105 in our experiments, which can achieve a good balance between the training time and accuracy.

E. DIFFERENT USERS

Even for the same activity, the operation range and speed may not be the same since different people tend to have different habits. Thus, to make sure our system can work for different users, we perform an experiment to evaluate the influence

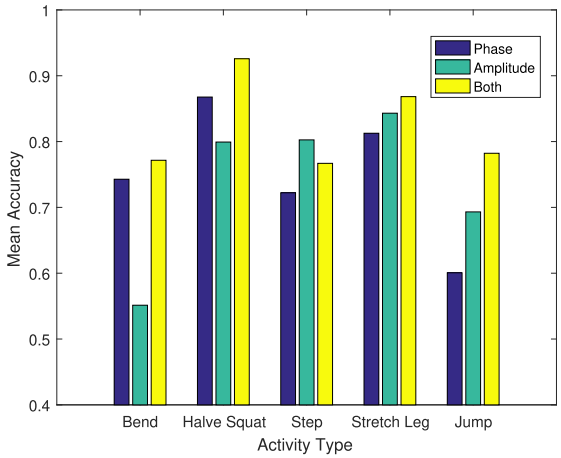


FIGURE 18. Performance for two users.

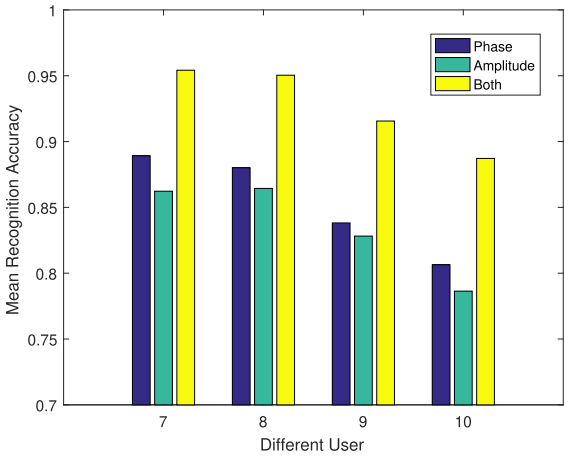


FIGURE 19. Performance for users with different ages.

generated by different users. In each environment, we further evaluate the data collected from users 2 to 6 in Table 1, and the results are shown in Fig. 17. In environment A, we can find that the prediction performances are still good for these users. Although the mean prediction accuracy decrease by about 3% respectively in environment B, the recognition accuracy is always at an acceptable level.

F. PERFORMANCE FOR MULTIPLE USERS

In order to ensure that our system can work properly in a multi-user environment, we do an experiment where not only the target user is moving, but also there exists another user who generates interference by doing some activities, such as walking, standing up, sitting down, etc. Moreover, the generation time of the interferential activity is not fixed, which means it may occur during the time period where the target user generates any activities. Here we present Wi-Motion's performance for two users as shown in Fig. 18. As we can expect, the overall performance decreases with the number of users increasing compared with a single user. More specifically, We can clearly see that in the case of two users, the average accuracy of phase and amplitude classifiers reduced to 70% because the probability of confusion among the users increases in such situation. Nonetheless, the classification accuracy still remains above 80% by combining two classifiers. Furthermore, we also do experiments under the situation where there are two interferential users, but the accuracy achieved was less than 68%. Obviously, it performs not as good as our expectation and is left as our future study.

G. DIFFERENT AGE OF USERS

The motivation behind studying the impact of user age is to see whether Wi-Motion's accuracy is impacted by any physiological properties caused by human's age. We do more data collection sessions in environment A for users 7 to 10 whose basic characteristics are presented in Table 1. Similarly, each user also provides 30 samples for each of the 5 activities in each collection session. Finally, we evaluated

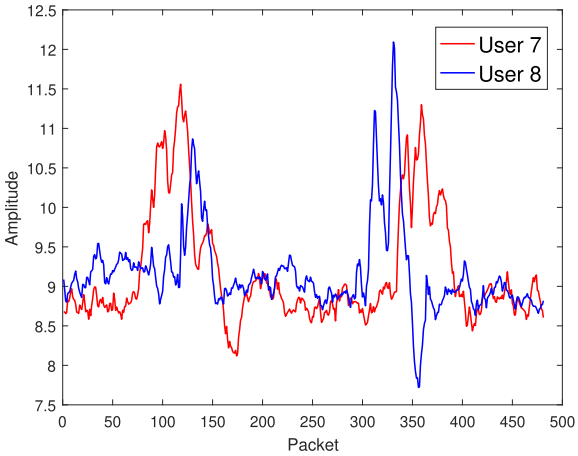


FIGURE 20. Waveform comparison when users with different ages do the same activity continuously.

the performance of Wi-Motion and the final results are as follows. Fig. 19 plots the mean recognition accuracy for each user. We observe that our combination strategy still has good performance in recognition accuracy, while the achieved performance gradually decreases with a certain degree as the age increases. We further analyzed the potential reason for this phenomenon. Fig. 20 plots two consequent amplitude waveforms generated by the same activity "stretch leg " for user 7 and user 8. Obviously, the waveforms of user 7 are basically the same while the two waveforms of user 8 still have a large difference although the waveforms are generated by the same activity. This shows that during the process of data collection, the physiological function of the person decays as the age increases, which makes the movement slower and difficult to control in a stable situation.

VIII. CONCLUSION

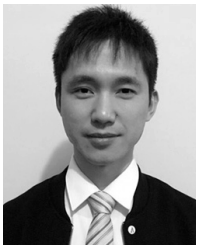
In this paper, we proposed a WiFi-based indoor activity recognition system called Wi-Motion. Compared to existing related systems, we adopt both amplitude and phase information to construct classifiers in our system in order to further

improve the recognition performance. Moreover, to enhance the robustness of Wi-Motion, the final recognition result of our system is determined by combining prediction results on all classifiers based on output posterior probability rather than simply obtaining from the single classifier. We intensively evaluated our system performance in various aspects. Experimental results verified that our system can get better mean false positive of 0.75% and mean true positive of 97.1%. Besides, Wi-Motion improves the recognition accuracy to 96.6% in the LOS environment compared with the original implementation that uses only one classifier constructed with amplitude or phase information. More meaningfully, in NLOS and the environments with multi-users interference, Wi-Motion also shows certain robustness in recognition performance.

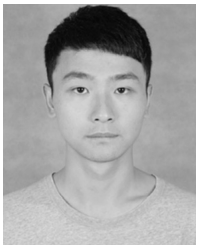
REFERENCES

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *Proc. Workshop World-Sensor-Web (WSW)*, Oct. 2006, pp. 117–134.
[2] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: An end-to-end participatory urban noise mapping system," in *Proc. 9th ACM/IEEE IPSN*, Apr. 2010, pp. 105–116.
[3] F. M. Adib and D. Katabi, "See through walls with WiFi!" in *Proc. ACM SIGCOMM*, Aug. 2013, pp. 75–86.
[4] L. Gong, Y. Wu, Z. Zhou, D. Man, H. Cai, X. Zhou, and Y. Zheng, "An adaptive wireless passive human detection via fine-grained physical layer information," *Ad Hoc Netw.*, vol. 38, pp. 38–50, Mar. 2016.
[5] 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. 2016.
[6] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, Dec. 2018, Art. no. 155.
[7] L. Gong, W. Yang, D. Man, G. Dong, M. Yu, and J. Lv, "WiFi-based real-time calibration-free passive human motion detection," *Sensors*, vol. 15, no. 12, pp. 32213–32229, 2015.
[8] Y. Gu, J. Zhan, Y. Ji, J. Li, F. Ren, and S. Gao, "MoSense: An RF-based motion detection system via off-the-shelf WiFi devices," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2326–2341, Dec. 2017.
[9] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li, "Wi-chase: A WiFi based human activity recognition system for sensorless environments," in *Proc. 18th IEEE WoWMoM*, Jun. 2017, pp. 1–6.
[10] M. De Sanctis, E. Cianca, S. Di Domenico, D. Provenziani, and M. Ruggieri, "WIBECAM: Device free human activity recognition through WiFi beacon-enabled camera," in *Proc. 2nd Workshop Workshop Phys. Anal.*, May 2015, pp. 7–12.
[11] L. Cheng and J. Wang, "How can i guard my AP?: Non-intrusive user identification for mobile devices using WiFi signals," in *Proc. 17th ACM MobiHoc*, Jul. 2016, pp. 91–100.
[12] Y. Chen, W. Dong, Y. Gao, X. Liu, and T. Gu, "Rapid: A multimodal and device-free approach using noise estimation for robust person identification," in *Proc. ACM on Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, Sep. 201, Art. no. 41.
[13] R. Zhou, X. Lu, P. Zhao, and J. Chen, "Device-free presence detection and localization with SVM and CSI fingerprinting," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7990–7999, Dec. 2017.
[14] K. Joshi, D. Bharadia, M. Kotaru, and S. Katti, "WiDeo: Fine-grained device-free motion tracing using RF backscatter," in *Proc. 12th USENIX NSDI*, May 2015, pp. 189–204.
[15] N. Husted and S. Myers, "Mobile location tracking in metro areas: Malnets and others," in *Proc. 17th ACM CCS*, Oct. 2010, pp. 85–96.
[16] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th ACM MobiCom*, Sep. 2013, pp. 27–38.
[17] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX NSDI*, Jan. 2014, pp. 317–329.
[18] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, Mar. 2000, pp. 775–784.
[19] S. Sigg, S. Shi, F. Buesching, Y. Ji, and L. Wolf, "Leveraging RF-channel fluctuation for activity recognition: Active and passive systems, continuous and RSSI-based signal features," in *Proc. MoMM*, Dec. 2013, p. 43.
[20] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 907–920, Apr. 2014.
[21] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE INFOCOM*, Apr. 2015, pp. 1472–1480.
[22] J. Shang and J. Wu, "A robust sign language recognition system with multiple Wi-Fi devices," in *Proc. Workshop Mobility Evolving Internet Archit.*, Aug. 2017, pp. 19–24.
[23] S. Tan and J. Yang, "WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition," in *Proc. 17th ACM MobiHoc*, Jul. 2016, pp. 201–210.
[24] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!" *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, Nov. 2016.
[25] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st ACM MobiCom*, Sep. 2015, pp. 90–102.
[26] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures," in *Proc. IEEE INFOCOM*, Apr. 201, pp. 1–9.
[27] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proc. ACM Conf. Special Interest Group Data Commu*, Aug. 2015, pp. 269–282.
[28] Q. Chen, K. Chetty, K. Woodbridge, and B. Tan, "Signs of life detection using wireless passive radar," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2016, pp. 1–5.
[29] A. Pokkunuru, K. Jakkala, A. Bhuyan, P. Wang, and Z. Sun, "NeuralWave: Gait-based user identification through commodity WiFi and deep learning," in *Proc. 44th IEEE IECON*, Oct. 2018, pp. 758–765.
[30] G. Melia, "Electromagnetic absorption by the human body from 1-15 GHz," Ph.D. dissertation, Dept. Phys., Univ. York, York, U.K., 2013.
[31] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the Mona Lisa: Spot localization using PHY layer information," in *Proc. 10th ACM MobiSys*, Jun. 2012, pp. 183–196.
[32] 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.
[33] L. Davies and U. Gather, "The identification of multiple outliers," *J. Amer. Stat. Assoc.*, vol. 88, no. 423, pp. 782–792, Sep. 1993.
[34] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[35] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007, pp. 69–84.
[36] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
[37] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011. [Online]. Available:http://www.csie.ntu.edu.tw/~cjlin/libsvm

HEJU LI received the B.Eng. degree in computer science and technology, in 2017. He is currently pursuing the master's degree with the School of Computer and Information, Anhui Normal University, Wuhu, China. His current research interests mainly include wireless communication and joint source-channel coding.



XIN HE (S'16–M'19) received the M.Sc. degree in information science from the School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), Ishikawa, Japan, in 2013, and the Ph.D. degree from JAIST and the University of Oulu, Oulu, Finland, in 2016. He is currently an Associate Professor with the School of Computer and Information, Anhui Normal University. He is also a Postdoctoral Researcher with the School of Computer Science and Technology, University of Science and Technology of China (USTC), Hefei, China. His research interests include joint source-channel coding, cooperative wireless communications, network information theory, energy harvesting, and backscatter communications.



XUKAI CHEN received the B.Eng. degree in computer science and technology, in 2013. He is currently pursuing the M.Sc. degree in computer science and technology with the School of Computer and Information, Anhui Normal University, Wuhu, Anhui. His research interests include cooperative wireless communications, energy harvesting, and backscatter communications.

YINYIN FANG received the B.Eng. degree in computer science and technology, in 2017. She is currently pursuing the M.Sc. degree in computer science and technology with the School of Computer and Information, Anhui Normal University, Wuhu, Anhui. Her main research interests include data analysis and data mining.

QUN FANG received the M.E. and Ph.D. degrees in computer science and engineering from Southeast University, in 2006 and 2011, respectively. He is currently a Professor with the School of Computer and Information, Anhui Normal University, Wuhu, Anhui, China. His current research interests include trusted computing, cloud computing, and the IoT security.