IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Device-free Human Activity Recognition Based on GMM-HMM using Channel State Information

XIAOYAN CHENG, BINKE HUANG, AND JING ZONG

School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Binke Huang (bkhuang@mail.xjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471293.

ABSTRACT This paper presents a machine learning method, Gaussian Mixture Hidden Markov Model (GMM-HMM), for device-free activity recognition using WiFi channel state information (CSI). The basic concept of CSI is introduced and signal changes caused by human activity are described, which demonstrates that human activity can be identified using a unique mapping between action and signal variations. The phase difference expanded matrix is built by the mean and standard deviation of phase difference as feature matrix after linear correction and Savitzky-Golay filter is performed on the CSI raw phase information. The GMM-HMM is used for classification as the human activity can be modeled as the Markov process and the complex activity patterns can be fitted by multiple Gaussian density functions, respectively. The proposed system is verified on the self-collected datasets and several factors affecting the recognition accuracy are analyzed. Furthermore, the system has compared with the previous work. High accuracy and robustness in universal scenarios are realized. Experimental results show that the average recognition accuracy of the proposed system is over 97%.

INDEX TERMS Activity recognition, channel state information (CSI), device-free, Gaussian Mixture Hidden Markov Model (GMM-HMM), phase difference.

I. INTRODUCTION

Recent years have witnessed increasing research interest in human activity recognition as it benefits multiple applications, such as intrusion detection[1], [2], smart homes[3], and health care services[4]. With the popularity of WiFi devices and the rich channel characteristics of channel state information (CSI), human activity recognition based on WiFi CSI has attracted widespread attention. Traditional recognition methods, such as sensor-based applications, usually require users to wear or attach smart devices, which increases inconvenience and obstruction for users [5], [6]. Vision-based human activity recognition methods have the problems of privacy security invasion and susceptibility to environmental influences such as light interference[7].

Compared with these approaches, WiFi-based activity recognition is capable of overcoming those disadvantages. There is no need for users to carry any equipment. Moreover, it enjoys the advantages of low cost, easy installation, and privacy protection. The current WiFi-based activity recognition technology typically utilizes two wireless signals, namely Radio Signal Strength Indicator (RSSI) and Channel State Information(CSI)[8]. RSSI is susceptible to narrowband interference and multipath interference, with low identification accuracy and limited performance. In the contrast, CSI can present the amplitude and phase of multipath propagation at different frequencies, thereby providing more abundant and stable channel parameters.

At the phase of activity recognition, the traditional Gaussian Mixture Hidden Markov Model (GMM-HMM) is used for it has strong data modeling capabilities and especially Gaussian Mixture Model (GMM) is known as the universal distribution approximator. Hidden Markov Models (HMM) builds a statistical model for the time series structure of the WiFi signal, and GMM is applied to fit the probability density function to generate the HMM observation sequence. The architecture of the human activity recognition system is shown in Fig. 1. In previous studies, CARM[9] uses the HMM to identify human activities by establishing a corresponding model between CSI changes and human activity speed and a corresponding model between human activity speed and activity type. CARM realizes activity recognition with an accuracy of more than 96%, but as mentioned above, the construction of the model is extremely complicated. WiHACS[10] trains and tests multi-class support vector machine (SVM) to realize human activity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access

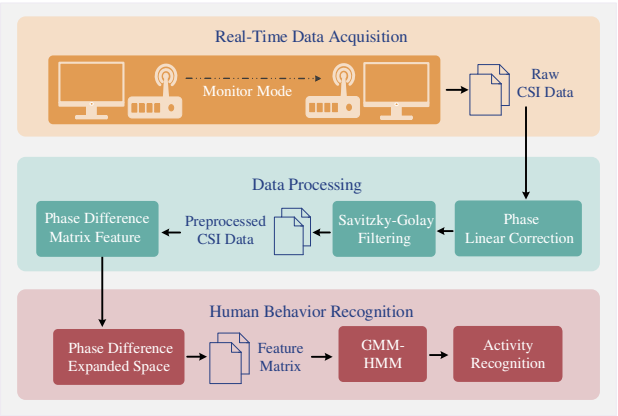


FIGURE 1. Activity recognition system architecture.

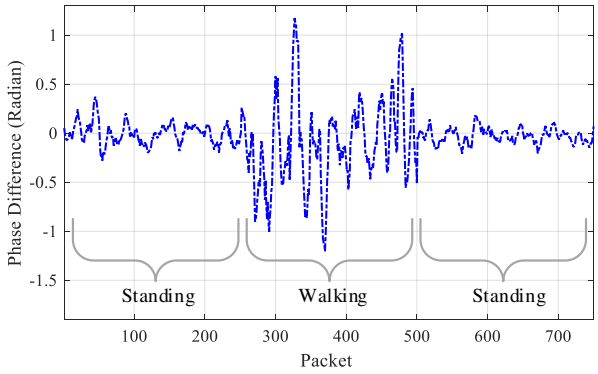


FIGURE 2. The phase difference when performing different actions.

classification based on OFDM Subcarriers' correlation. Even if it has a nice recognition performance, SVM is difficult to realize multi-classification when the sample size is too large.

The main innovativeness and contributions of this paper are summarized as below:

1) Phase difference is used as the characteristic signal of human activity recognition, due to it uses space diversity and frequency diversity technology, and can better perceive the weak changes of the environment than phase. Furthermore, the phase difference, the mean, and variance of phase difference are used as the expanded matrix to make the recognition result more accurate.

2) HMM is selected as a WiFi signal-based recognition methodology to enhance robustness and adaptability in the work. Different from initializing the HMM parameters directly, the HMM parameter B (observation probability matrix), which has a great impact on the system accuracy, is initialized with GMM.

3) A powerful system is established which can recognize human activities with high precision. Experiments are conducted in self-collected datasets to verify the validity of the system for user activity recognition under WiFi signals. The system performance is evaluated and the factors influencing the accuracy are analyzed. Experimental results show that the system has a strong ability to identify different activities and has robustness under different environments.

The remaining of the paper is organized as follows: Section II introduces the basic knowledge of CSI, data preprocessing, and the construction of the feature matrix. Section III describes the classification methodology. Then, the work of data collection and the experiment results are presented in Section IV. Finally, Section V concludes this work.

II. DATA PREPARATION AND FEATURE EXTRACTION

Three basic characteristic quantities are usually obtained from CSI data: amplitude, phase, and phase difference between adjacent antennas. The CSI phase information collected initially will change turbulently, and there are no rules at all, for the time and frequency synchronization between the receiver and the transmitter in WiFi equipment. Therefore, instead of using phase information, CSI amplitude information is used in the initial research work. However, the amplitude information lacks sensitivity to weak actions and is difficult to apply to the recognition of fine-grained behaviors. Thanks to the phase correction algorithm[11], phase information is gradually being used in various studies. In this paper, the phase difference is used as the characteristic signal of human activity recognition, due to it uses space diversity and frequency diversity technology, and can better perceive the weak changes of the environment than phase.

The phase difference changes when performing different activities is illustrated in Fig. 2. It can be observed that the phase difference during walking will fluctuate sharply compared to in the stationary state. Therefore, the phase difference can be used as an evaluation criterion to distinguish different actions, and the characteristic information which could distinguish any two states can be extracted from the phase difference.

A. BASIC CONCEPT OF CSI

CSI is a fine-grained signal feature captured from the physical layer (PHY) of the WiFi communication via Orthogonal Frequency Division Multiplexing (OFDM) technology. It describes channel properties of wireless communication links and amplitude and phase variations caused by path loss and multipath effects, including scattering, diffraction, and distance attenuation. CSI can make the signal transmission of the communication system adapt to the channel conditions of the current system, which enables a MIMO system to achieve reliable system communication with high robustness, stability, and more transmission information.

In the frequency domain, the wireless channel can be expressed as $Y(f,t) = H(f,t) \times X(f,t)$, where $X(f,t)$ and $Y(f,t)$ are the transmitted and received signals at a certain OFDM carrier frequency of f and time of t, respectively. $H(f,t)$ is the complex-valued Channel Frequency Response (CFR) between a pair of antennas and the time-series of CFR values for a given antenna pair and OFDM subcarrier is called a CSI stream. Therefore, CFR can be represented as $H(f,t) = Y(f,t)/X(f,t)$. Leveraging the off-the-shelf Intel 5300 NIC wireless network with modified drivers, CSI

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access

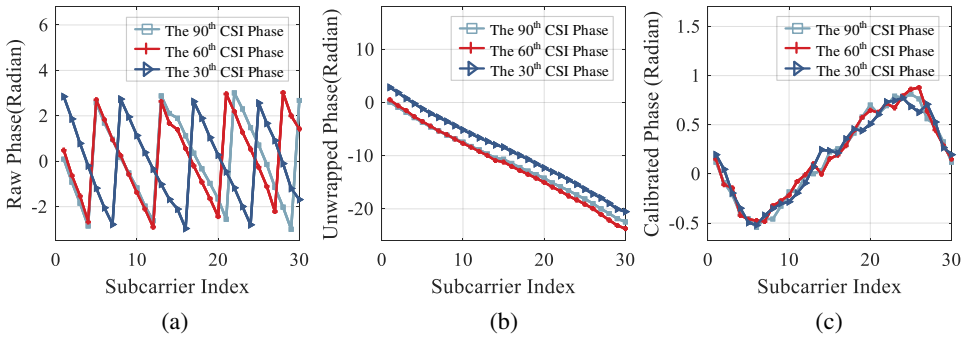


FIGURE 3. Process of raw phase correction. (a) Raw phase, (b) Unwrapped phase, (c) Calibrated phase.

information can be extracted from the received packets. Generally, each CSI measurement contains N matrices with dimensions of $NTx \times NRx$, where N, NTx, and NRx are the number of OFDM subcarriers, the number of transmitting and receiving antennas, respectively. This paper uses 20MHz bandwidth with 30 subcarriers in5.320GHz for more stable and robust signals can be provided compared to 2.4GHz. Consequently, the CSI of three data streams from each data packet can be expressed in the following equation (1) when adopt one transmitting antenna and three receiving antennas.

$$CSI = \begin{bmatrix} H_{1,1} & \cdots & H_{1,30} \\ \vdots & \ddots & \vdots \\ H_{3,1} & \cdots & H_{3,30} \end{bmatrix} \quad (1)$$

B. PHASE DIFFERENCE EXTRACTION

The received CSI data contain a relevant noise in the real communication process for the wireless signal is interfered with by the hardware equipment and the environment. Thus, the wireless channel is represented as:

$$\mathrm{H}(f,t) = \left(\sum_{i=1}^{N} a_i(f,t) e^{-j2\pi f \tau_i(f,t)}\right) e^{-j(2\pi \Delta f t + \theta S + \theta N)} \quad (2)$$

where $a_i(f,t)$ is the complex-valued representation for both the attenuation and the initial phase offset of the i^{th} path, $e^{-j2\pi f \tau_i(f,t)}$ is the phase shift on the i^{th} path caused by a propagation delay of $\tau_i(f,t)$, N is the number of subcarriers, $e^{-j2\pi \Delta f t}$ is the phase shift caused by the Carrier Frequency Offset (CFO), which is mainly induced by the difference in central frequencies (lack of synchronization) between the transmitter and receiver clocks. θS and θN are the phase offset caused by sampling frequency offset (SFO), environment, and hardware noise, respectively.

In general, the measured phase of the channel response of the i^{th} subcarrier can be expressed as:

$$\hat{\phi}_i = \phi_i - 2\pi \frac{k_i}{N} \delta + \beta + Z \quad (3)$$

where $\hat{\phi}_i$ and ϕ_i are the measured and actual phase information of the i^{th} subcarrier, respectively. δ is the time offset of the receiver, β is the constant phase offset, Z is the measured noise, k_i is the index of the i^{th} subcarrier which is from -28 to 28 in IEEE 802.11n, and N is the window size of the Fourier Transform which is 64 in IEEE 802.11a/g/n. The unprocessed phase information received by the receiver cannot be directly used for human behavior analysis for it contains a lot of noise. According to reference[12], a linear correction is applied to the original phase data to eliminate the main noises δ and β.

Define two variables a and b:

$$a = \frac{\hat{\phi}_n - \hat{\phi}_1}{k_n - k_1} = \frac{\phi_n - \phi_1}{k_n - k_1} - 2\pi\delta \quad (4)$$

$$b = \frac{1}{n}\sum_{j=1}^{n} \hat{\phi}_j = \frac{1}{n}\sum_{j=1}^{n} \phi_j - \frac{2\pi\delta}{N}\sum_{j=1}^{n} k_j + \beta \quad (5)$$

when the sub-carrier frequency is symmetrical, $\sum_{j=1}^{n} k_j = 0$, b can be simplified to $b = \frac{1}{n}\sum_{j=1}^{n} \phi_j + \beta$. The calibrated phase of the i^{th} subcarrier can be expressed as:

$$\tilde{\phi}_i = \hat{\phi}_i - ak_i - b = \phi_i - \frac{\phi_n - \phi_1}{k_n - k_1} k_i - \frac{1}{n}\sum_{j=1}^{n} \phi_j \quad (6)$$

The linear noise induced by δ and β is eliminated while the measurement noise Z is small. Fig. 3(a) shows the original phase sequence of 30 subcarriers in the stationary state, which can be observed that there is a 2π jump. The phase is converted to a continuous form by unwrapping, as shown in Fig. 3(b). This process can be referred to as the unwrapping function in MATLAB. Then the phase deviation is removed by a linear transformation, as shown in Fig. 3(c). The calibrated phase $\tilde{\phi}_i$ is used to construct the initial phase difference matrix.

Subsequently, the Savitzky-Golay filter[13] is utilized to eliminate sudden changes and small random variations. Savitzky-Golay filter fits successive subset of adjacent data points with a low degree polynomial by the method of linear least square. The polynomial order and the length of the frame are set to be 3 and 7 in our experiments. Finally, the feature matrix is built by the phase expanded matrix which is expanded using the mean and standard deviation of phase difference, as well as phase difference itself.

III. CLASSIFICATION METHOD

The method of activity recognition is introduced in this section. Inspired by Sheng *et al.* [14], which propose an HMM-based methodology for action recognition using star skeleton, GMM-HMM is used as our method for activity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access

recognition. The basic concept of GMM and HMM is presented firstly. And then, the feasibility of GMM-HMM for activity recognition is illustrated.

A. GMM BASICS

GMM is a parameterized model of the probability distribution, which aims to build the probability distribution $P(x)$ of N-dimensional datasets into a mixture of finite multivariate Gaussian distributions. K-order Gaussian GMM probability density function is as follows:

$$P(x|\mu,\Sigma) = \sum_{k=1}^{K} c_k N(x|\mu_k,\Sigma_k) \tag{7}$$

where $c_k > 0$ are mixture coefficient that sum to 1, and $N(x|\mu_k,\Sigma_k)$ represents a multivariate Gaussian distribution which is parameterized by its mean vector μ, and covariance matrix Σ.

The initialization of GMM parameters is based on the K-means algorithm. Since the observation probability density function is the k^{th} component of GMM, they are completely defined by the parameters (c_k,μ_k,Σ_k). The K-means algorithm can calculate the optimization center of component μ and covariance Σ for each state, and iterate step by step to adapt to the state sequence.

After that, an optimized GMM model will be realized with the Expectation-Maximization (EM) method. The EM algorithm comprises two steps: E-Step and M-Step. E-Step calculates the posterior probability $\gamma(n,k)$ according to the current c_k,μ_k,Σ_k. Then, the calculated value is delivered to M-step, so that c_k,μ_k,Σ_k are updated. This process is repeated until the log-likelihood $lnp(x|c,\mu,\Sigma)$ converges or reaches the maximum number of iterations. Use the following formula to updatec_k,μ_k,Σ_k:

$$\mu_k = \frac{\sum_{n=1}^{N}\gamma(n,k)x_n}{\sum_{n=1}^{N}\gamma(n,k)} \tag{8}$$

$$c_k = \frac{\sum_{n=1}^{N}\gamma(n,k)}{\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(n,k)} \tag{9}$$

$$\Sigma_k = \frac{\sum_{n=1}^{N}\gamma(n,k)(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N}\gamma(n,k)} \tag{10}$$

where,

$$\gamma(\mathrm{n},\mathrm{k}) = \frac{c_k N(x_n|\mu_k,\Sigma_k)}{\sum_{k=1}^{K} c_k N(x_n|\mu_k,\Sigma_k)} \tag{11}$$

B. HMM BASICS

HMM is a statistical framework for modeling time-varying spectral vector sequences and is a potent tool in pattern recognition. HMM is expressed conventionally by five basic elements N, M, π, A, B.

N and M represent the number of model hidden states and observable symbols generated in each state, respectively. The state transition probability distribution between state $q_i$to q_j is $A = [a_{ij}]_{N\times M}$, and the observation probability distribution of emitting any vector o_t at state q_j is given by $B = [b_j(k)]_{N\times M}$. The probability distribution of the initial state is $\pi = [\pi_i]$.

$$\pi_i = P(i_1 = q_i) \tag{12}$$

$$a_{ij} = P(i_{t+1} = q_j | i_t = q_i) \tag{13}$$

$$b_j(k) = P(o_t = \mathrm{v}_k | i_t = q_j) \tag{14}$$

As for HMM, there exist three main problems[15]. First is the evaluation problem. Given an observation sequence $O = \mathrm{o}_1, \mathrm{o}_2, \dots, \mathrm{o}_T$ and a model $\lambda = (A, B, \pi)$, the probability of the observation sequence generated by the given model $P(O|\lambda)$ can be calculated with the Forward-Backward algorithm. After defining the forward probability $\alpha_t(i)$ and the backward probability $\beta_t(i)$ respectively, the forward probability and the backward probability at the next moment can be recursively achieved. The probability of the observation sequence can be calculated by either of the following two formulas:

$$P(O|\lambda) = \sum_{i=1}^{N}\alpha_T(i) \tag{15}$$

$$P(O|\lambda) = \sum_{i=1}^{N}\pi_i\beta_1(i)b_i(o_1) \tag{16}$$

where,

$$\beta_t(i) = P(\mathrm{o}_{t+1}, \mathrm{o}_{t+2}, \dots, \mathrm{o}_T | i_t = q_i, \lambda) \tag{17}$$

$$\alpha_t(i) = P(\mathrm{o}_1, \mathrm{o}_2, \dots, \mathrm{o}_t, i_t = q_i | \lambda) \tag{18}$$

The second problem is the learning problem. Given the observation sequence of the model $O = \mathrm{o}_1, \mathrm{o}_2, \dots, \mathrm{o}_T$, estimate the parameters of the model $\lambda = (A, B, \pi)$ through the observation sequence to maximize the probability of the observation sequence $P(O|\lambda)$ under the model. The essence is the problem of using maximum likelihood estimation to obtain parameters. The Baum-Welch algorithm is widely used to solve learning problems.

The Baum-Welch algorithm uses the principle of the EM algorithm to find the expected $L(\lambda,\bar{\lambda})$ of the joint distribution $P(O,I|\lambda)$ based on the conditional probability $P(I|O,\bar{\lambda})$ at E-step, where $\bar{\lambda}$ is the current model parameter. The expectation is maximized at M-step to get the updated model parameter λ. EM iteration is continued until the values of the model parameters converge. The following parameters are updated to test whether the values converge or not:

$$a_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \tag{19}$$

$$b_j(k) = \frac{\sum_{t=1,o_t=v_k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(i)} \tag{20}$$

$$\pi_i = \gamma_1(i) \tag{21}$$

where $\xi_t(i,j)$ is the probability of being in the state q_i at time t and being in the state q_j at time $t+1$ giving model λ and observation O, denoted as:

$$\begin{aligned}\xi_t(i,j) &= P(i_t = q_i, i_{t+1} = q_j | O, \lambda) \\ &= P(i_t = q_i, i_{t+1} = q_j, O|\lambda)/P(O|\lambda) \\ &= \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}\end{aligned} \tag{22}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access

And $\gamma_t(i)$ represents the probability of being in the state q_i at time t, denoted as:

$$\gamma_t(i) = P(i_t = q_i|O,\lambda) = \frac{P(i_t = q_i, O|\lambda)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)} \quad (23)$$

The last problem is the prediction problem. Also known as the decoding problem. Given the model $\lambda = (A, B, \pi)$ and the observation sequence $O = o_1, o_2, \dots, o_T$, the state sequence I corresponding to the model can be found when the maximum observation sequence conditional probability $P(I|\lambda)$ is obtained. In other words, when the observation sequence is given, find the most likely hidden state sequence corresponding to it. The Viterbi algorithm is usually used to solve the prediction problem.

Define two variables: δ and Ψ.

The maximum probability of all single paths $(i_1, i_2, \dots, i_t)$ in state i at time t is:

$$\delta_t(i) = \max P(i_{t+1} = q_i, i_t, \dots, i_1, i_{t+1}, \dots, i_1), \ i = 1, \dots, N \quad (24)$$

Recursion can be obtained according to the above definition:

$$\delta_{t+1}(i) = \max\left[\delta_t(j)a_{ji}\right] b_i(o_{t+1}), i = 1, \dots, N; \ t = 1, \dots, T-1 \quad (25)$$

At the same time, define all the single paths $(i_1, i_2, \dots, i_{t-1}, i)$ whose state is q at time t, and record the maximum probability of $t-1$ nodes in the path as:

$$\Psi_t(i) = \arg\ max\left[\delta_{t-1}a_{ji}\right], i = 1,2, \dots, N \quad (26)$$

In other words, the backward pointer Ψ is used to record the previous state which leads to the maximum local probability of a certain state, and it is used to backtrack the optimal path (optimal hidden state sequence) in the algorithm.

C. FORMULATION OF GMM-HMM

HMM is widely used in speech recognition, gesture recognition, and other fields [16]-[18]. Its application in activity recognition is based on one basic assumption: the activity to be modeled as a Markov process. This process comprises visible observations, with each observation corresponding to a hidden state which the observer cannot see [19]. In this paper, it is observed that human activity is composed of body movements which will cause the CSI information change. Therefore, the human motion is considered to be modeled as HMM, where the visible CSI phase value is the observation value and the limb transition is a hidden state. The possibility of transition between different limbs depends on the particular structure of the action itself.

Human activity is a continuous motion in time and space, and the complex motion patterns can be fitted by multiple Gaussian density functions. HMM can be divided into discrete HMM and continuous HMM based on the characteristics of different observed variables. The difference between the two lies in the model parameter B. The observation of the former is a discrete random variable, and the corresponding model parameter B is a probability distribution matrix. In continuous HMM, the observation is a continuous random variable, and the corresponding model parameter B is composed of the observation probability density function of the state. Under normal circumstances, the distribution of each state can be fitted with a mixed Gaussian distribution. The probabilistic model can deal with data with strong noise, at the same time has good robustness, and performs well on high-dimensional data. In particular, the GMM has strong coding capabilities for continuous and complex motion trajectory data such as human motions. Therefore, the GMM-HMM can be used to imitating and learn complex human activities. The model frame is shown in Fig. 4.

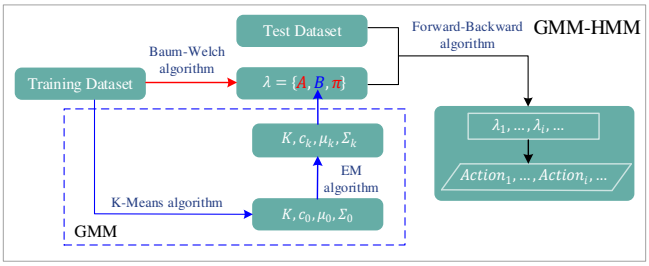


FIGURE 4. GMM-HMM Model frame.

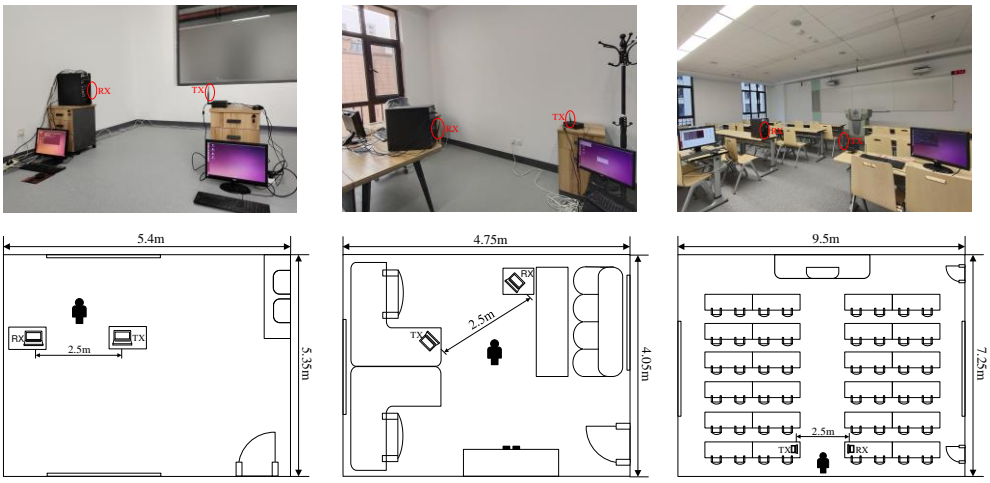
It is generally believed that the initial values of the parameters π and A will not have much influence on the model results. When the basic constraints are met, random values or uniform values can be used. However, the different initialization methods of parameter B will greatly affect the model results and usually choose a more complicated initialization method according to different applications. This article uses the GMM method in part A of Section III to initialize the model parameters. In particular, Σ is set to the diagonal covariance matrix ('diag'), the Gauss number K is set as 3, and the maximum number of iterations is set as 20. Except that the calculation of probability distribution matrix B is slightly complex, the three basic algorithms of GMM-HMM are the same as those of HMM.

After obtaining the initial model parameters, the Baum-Welch algorithm described in part B of Section III is used to train the model parameters with collected training datasets and optimize the model parameters iteratively. For each activity sequence, a corresponding model will be trained. Then, the test datasets are input to the trained model in the testing phase. The Forward-Backward algorithm is utilized to obtain the best model. After the observation sequence and model parameters are obtained, the model that maximizes the probability of the observation value can be solved with each model corresponding to a specific activity. Subsequently, the recognition of the action is achieved.

IV. EXPERIMENTS AND EVALUATION

The proposed system is assessed on the self-collected datasets, the factors affecting the recognition accuracy are discussed, and the system performance is compared with previous work.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access



(a) Washroom (b) Office (c) Classroom

FIGURE 5. Experimental scenarios. The height of the WiFi device is 1.2m.

A. EXPERIMENTAL SETUP

Our system is implemented on COTS hardware. A LENOVO desktop with three external antennas is used as a receiver pinging packets from a Mini desktop with one antenna. Both of them are equipped with Intel 5300 NIC and an open-source driver modified by Halperin et al. [20] to promote the correct reception of the network card. Then, the Linux 802.11n CSI Tool software, working under the Ubuntu 14.04 LTS system, is used to obtain the CSI data from the driver. The constructed 1×3 MIMO system operates in a 5.320 GHz environment and collects samples at 100 Hz. After completing the extraction of CSI information, the CSI data is preprocessed by Matlab2017a and input into HMM to complete training.

A total of 600 (5actions×30samples× 4volunteers) samples are collected by recruiting four volunteers with different heights and shapes to imitate five different activities (*calling, squatting, walking, stand-fall, walk-fall*) in a pre-configured environment (a washroom as illustrated in Fig. 5(a)). The volunteers are two females and two males ranging in age from 22 to 25. The datasets are extracted in the washroom of size 5.4m×5.35m and each sample consumes 10 seconds. When each sample is collected, the specified action is only performed from the 4th to the 7th second and the volunteer remains still the rest of the time to prevent overlapping of the action data during data collection. The source data is collected and named as $subjecti\_aj\_sk$, which represents the k^{th} record of the j^{th} action performed by experimenter i. The performance of the system is then evaluated in diverse environments, such as the washroom, an office of size 4.75m×4.05m (as shown in Fig. 5(b)), and a classroom of size 9.5m×7.25m (as shown in Fig. 5(c)).

Note that unless otherwise specified, 1) the rate of transmission is 100Hz. 2) for all samples of each action, 2/3 samples are randomly selected as training data, and the remaining 1/3 samples are applied as test data. 3) A total of 20 experiments of each action are performed and keep a record of the accuracy of every experiment. Then the average values of the recorded accuracies of each action are calculated as the final results to pursue a fair degree of precision.

Truth \ Prediction	Calling	Squatting	Walking	Stand-fall	Walk-fall
Calling	96.8	0.0	0.0	3.2	0.0
Squatting	2.9	94.1	0.0	3.0	0.0
Walking	0.0	0.0	98.9	0.0	1.1
Stand-fall	1.2	0.0	0.0	98.8	0.0
Walk-fall	0.0	0.0	0.4	0.0	99.6

FIGURE6. Confusion matrix of 5 activities performed by 4 volunteers.

B. EVALUATION

1) ACCURACY OF ACTIVITY CLASSIFICATION

The confusion matrix shown in Fig. 6 is established to assess the performance of the proposed system, in which each element denotes the ratio used to classify actual activity and predicted activity. The results show that the recognition accuracy of each activity is larger than 97%, indicating that the system achieves a high classification accuracy overall activities. Four activities achieve over 96% classification accuracy except for *squatting* activity since *squatting* is very similar to *falling* and easy to be misjudged.

2) COMPARED WITH PREVIOUS WORK

We compare our recognition system with previous work, Random forest, HMM, LSTM in terms of *lie down, fall, walk, run, sit down, stand up*[21]. To make sure it is a fair and convincing comparison, the datasets from reference [21] are used in the proposed system. In other words, all the methods participating in the comparison use the same datasets. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access

TABLE 1. Identification precision of different methods

Method	Average Accuracy
Random forest	64.7%
HMM	73.3%
LSTM	90.5%
Proposed method	97.8%

TABLE 2. Mean accuracy of different methods

Method	Average Accuracy
CNN+MFCC	78.0%
LSTM+MFCC	78.9%
HMM+MFCC	80.0%
Proposed method	99.0%

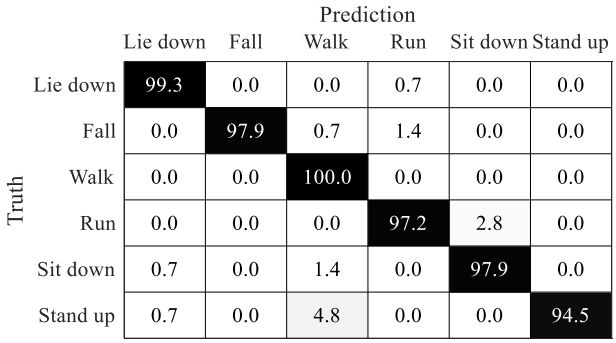


FIGURE 7. Confusion matrix of 6 activities in literature [21] using the proposed system.

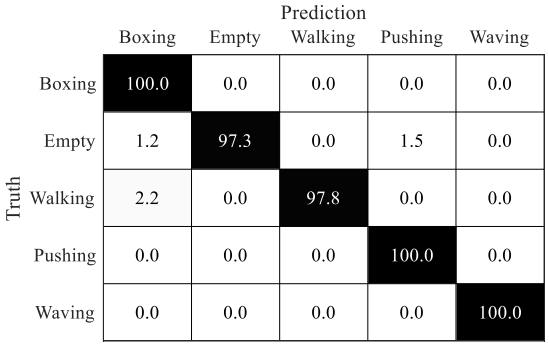


FIGURE 8. Confusion matrix of 5 activities in literature [22] using the proposed system.

participating in the comparison use the same datasets. The details of the parameters and datasets are shown below.

For the Random forest method, the PCA is applied on the CSI amplitude and STFT is used to extract features (the first 25 frequency components are used as the feature vector). Then, a Random forest with 100 trees is used for activities classification. The extracted features using STFT are also applied in HMM and the MATLAB toolbox is used for HMM training. The raw CSI amplitude with 90-dimension (3 antennas × 30 sub-carriers) is used as the feature vector for evaluating the performance of LSTM and the number of hidden units is set to be 200 where only one hidden layer is considered. Furthermore, samples of datasets[1] are collected at 1KHz sampling rate in an indoor office area where the Tx and Rx are located 3 m apart in LOS condition and a total of 720 samples (6actions×20samples× 6volunteers) are collected.

Fig. 7 and Table 1 give the confusion matrix and comparing results, respectively. It can be observed that the proposed system has the highest average accuracy of 97.8%, for the selected feature is phase difference and the GMM algorithm is utilized to initialize HMM parameters, which can identify activities more accurately.

In addition, we also compared the system with references [22] in terms of *boxing, empty, walking, pushing, waving*. Reference [22] proposes Mel frequency cepstral coefficient (MFCC) feature extraction for audio signals for CSI time series classification and MFCC features are used in CNN, LSTM, and HMM classification methods. ITI datasets consist of five activities with 50 training samples for each one. The data were collected from a person moving in a 3.1 m by 7.0 m

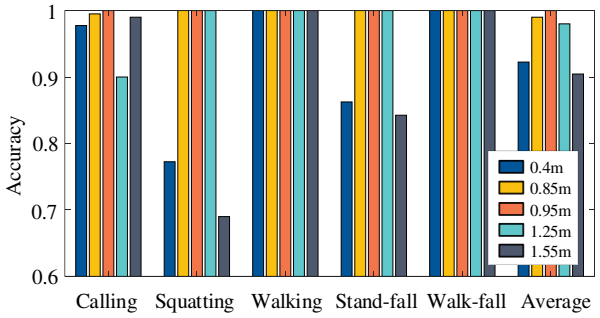


FIGURE 9. Detection accuracies at different heights.

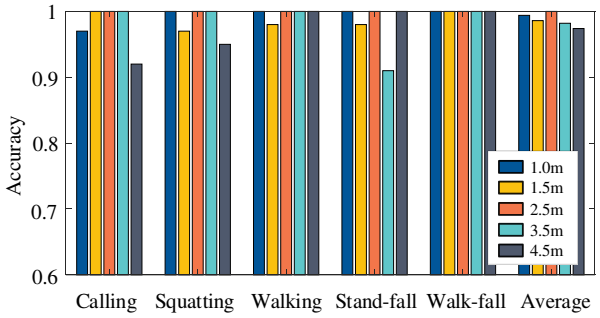


FIGURE 10. Detection accuracies at different distances.

office room, in multiple positions and directions. It can be observed from Figure 8 and Table 2, which represent confusion matrices and comparison results respectively, the identification accuracy of our system is 99.0%, which is superior to other methods when using ITI datasets.

3) OPTIMAL LOCATION PARAMETERS DETERMINATION FOR WIFI DEVICES

[1] https://drive.google.com/file/d/19uH0_z1MBLtmMLh8L4BlNA0w-XAFKipM/view?usp=sharing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access

TABLE 3. The system robustness in different environments

Train	Test	Accuracy					
		Calling	Squatting	Walking	Stand-fall	Walk-fall	Average
Model trained in classroom	Test in office	99.8%	100%	98.1%	99.6%	99.6%	99.4%
	Test in washroom	99.0%	100%	98.1%	96.9%	97.7%	98.3%
Model trained in office	Test in classroom	99.2%	100%	98.0%	98.2%	97.7%	98.6%
	Test in washroom	97.6%	99.4%	97.4%	96.7%	92.3%	96.7%
Model trained in washroom	Test in classroom	97.2%	99.5%	97.2%	85.5%	100%	95.9%
	Test in office	99.5%	100%	96.0%	96.3%	91.0%	96.6%

When an action occurs, the position of the person changes on the vertical plane and the horizontal plane. Therefore, the height of the device and the TX-RX distance will affect the recognition accuracy. To find the height and distance for WiFi devices that can provide the most accurate recognition accuracy, two experiments are conducted using the activity data of a volunteer (30 samples) collected in the washroom: diverse device heights with 0.4m, 0.85m, 0.95m, 1.25m, and 1.55m, and different TX-RX distances including 1.0m, 1.5m, 2.5m, 3.5m, and 4.5m. Fig. 9 and Fig. 10 show the evaluation results at different heights and distances for WiFi devices. It can be observed that when the two elements are set at 0.95m and 2.5m, the system can achieve the best identification accuracy for all the activities. Therefore, 0.95m and 2.5m are selected as the height of the device and the TX-RX distance in the following experiments, respectively.

4) IMPACT OF ENVIRONMENTAL INTERFERENCE

The CSI characteristics are different for diverse environments. The identification precision of our system in the classroom, office, and washroom, as well as washroom with interference, is evaluated, respectively. Thirty samples of each activity for one volunteer are collected in every environment so that the total number of all samples in the experiment is 600. As shown in Fig. 11, the average recognition accuracies in the above environments are 100%, 100%, and 100%, 98.2%. The accuracy for the washroom in the second case is slightly lower due to there is an additional subject making pretty small movements at the edge of the experimental field in the process of data collection, which causes some interference. The results show that the proposed method can achieve high accuracy in different environments while it is difficult to correctly collect data in an environment with interference.

5) THE MODEL ENVIRONMENT ROBUSTNESS

To assess the robustness of the proposed model in different environments, we tested models trained in each environment using data collected from other environments. For example, models trained in the classroom will be used to test data collected from bathrooms and offices to assess model robustness. Thirty samples of each activity are collected from a volunteer in each environment. As shown in TABLE 3, the accuracy of the trained model in other different environments is above 95.9%. which indicates that the proposed method is robust to different environments.

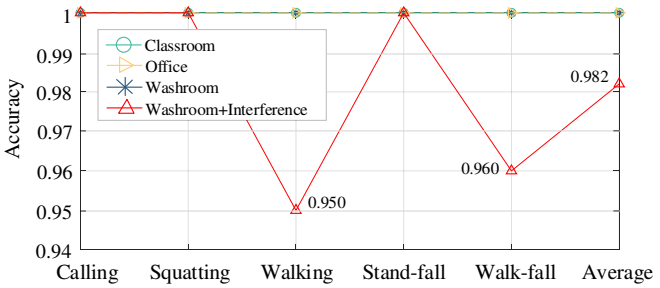


FIGURE 11. Recognition accuracy in different environments

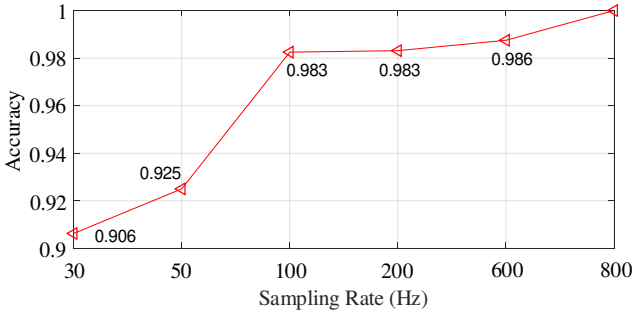


FIGURE 12. Impact of sampling rates on accuracy

6) IMPACT OF DIFFERENT TRANSMISSION RATES

The impact of transmission rates on human recognition accuracy is further investigated. Fig. 12 shows the average identification precision of five activities collected from a volunteer in the washroom at six diverse sampling rates. It can be observed that severely degraded performance happens when the sampling rate is around 50 Hz and accuracy improves with a higher sampling rate for noticeable changes of CSI can be captured during movement (maximum accuracy is reached at 800Hz), but the increase in accuracy is not obvious beyond 100Hz. Therefore, 100Hz is selected in the following experiments as our sampling rate to obtain a good compromise between computational cost and precision.

7) IMPACT OF TRAINING-SAMPLE SIZE AND HUMAN DIVERSITY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3082627, IEEE Access

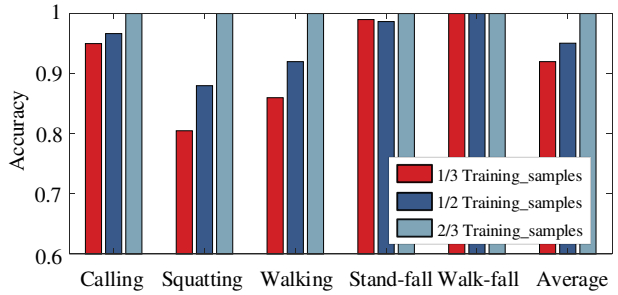


FIGURE 13. Impact of training datasets dimensionalities on accuracy.

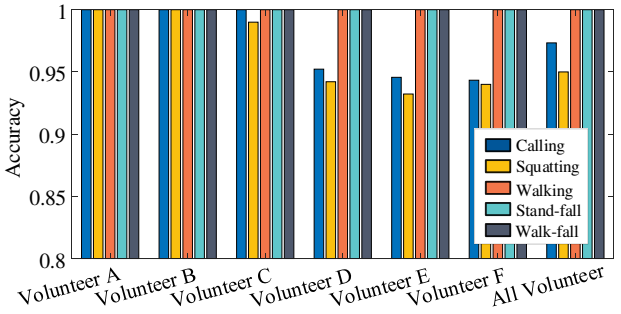


FIGURE 14. Impact of human diversity on accuracy.

Two proof schemes are designed to evaluate and analyze the system performance on our own collected datasets. The different number of training samples has an important influence on the accuracy of activity recognition. The results of three different proportions of the training datasets of five activities collected from a volunteer in the washroom are shown in Fig. 13. The results suggest that increase the size of the training datasets appropriately can get a better recognition accuracy.

The diversity of people not only increases the diversity of CSI but also increases the difficulty of identifying activities for people who have different movement patterns, such as speed, ranges, and styles. There are six volunteers in the experiment, three of whom do targeted training before the experiment and three of whom do not. Fig. 14 shows that volunteers A, B, and C who exercise regularly achieve accuracy of 100%, 100%, and 99.8%, respectively. The remaining volunteers D, E, and F who are not trained reach 97.8%, 97.5%, and 97.6%, which are slightly lower than the former. The accuracy of the fusion data of the six volunteers is 98.5% which is better than the three untrained volunteers. So that, targeted practice of simulated activities before performing experiments could perhaps improve accuracy standard.

V. CONCLUSION

In this paper, an activity recognition system using commodity WiFi devices is proposed, the factors affecting accuracy are explored, and the ability and robustness to recognize human activities are demonstrated. The results show that the proposed system achieves an average accuracy of greater than 97% on self-collected datasets. It is worth noting that the recognized activities include two types of falling actions (*stand-fall* and *walk-fall*), which reminds us that the system has great potential to be a practical, non-intrusive solution for activity recognition and fall detection. The solutions of how to identify more fine-grained human activities, simultaneously identify the activities of multiple people, and improve the system robustness in complex environments are urgent problems to be solved in engineering applications. Those above challenges will be considered in our future work.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. Ilias Kalamaras and Mr. Konstantinos Votis [22] for useful discussions and for providing us the ITI datasets. We would also like to thank the authors in Reference [21] for sharing their datasets as open source. This work is supported in part by the National Natural Science Foundation of China (Grant No. 61471293).

REFERENCES

[1] W. Liu, X. Gao, L. Wang, and D. J. W. P. C. Wang, "BFP: Behavior-free passive motion detection using PHY information," vol. 83, no. 2, pp. 1035-1055, 2015.

[2] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2016, pp. 363-373.

[3] 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," presented at the Proceedings of the 20th annual international conference on Mobile computing and networking, Maui, Hawaii, USA, 2014. [Online]. Available: https://doi.org/10.1145/2639108.2639143.

[4] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-Free Fall Detection by Wireless Networks," IEEE Transactions on Mobile Computing, vol. 16, no. 2, pp. 581-594, 2017, doi: 10.1109/TMC.2016.2557792.

[5] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192-1209, 2013, doi: 10.1109/SURV.2012.110112.00192.

[6] C. S. Evangeline and A. J. S. R. Lenin, "Human health monitoring using wearable sensor," 2019.

[7] X. Yang and Y. Tian, "Super Normal Vector for Human Activity Recognition with Depth Cameras," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 5, pp. 1028-1039, 2017, doi: 10.1109/TPAMI.2016.2565479.

[8] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," vol. 46, no. 2 %J ACM Comput. Surv., p. Article 25, 2013, doi: 10.1145/2543581.2543592.

[9] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-Free Human Activity Recognition Using Commercial WiFi Devices," IEEE Journal on Selected Areas in Communications, vol. 35, no. 5, pp. 1118-1131, 2017, doi: 10.1109/JSAC.2017.2679658.

[10] T. Z. Chowdhury, C. Leung, and C. Y. Miao, "WiHACS: Leveraging WiFi for human activity classification using OFDM subcarriers' correlation," in 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 14-16 Nov. 2017 2017, pp. 338-342, doi: 10.1109/GlobalSIP.2017.8308660.

[11] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the Mona Lisa: spot localization using PHY layer information," presented at the Proceedings of the 10th international conference on Mobile systems, applications, and services, Low Wood Bay, Lake District, UK, 2012. [Online]. Available: https://doi.org/10.1145/2307636.2307654.

[12] J. Zong, B. Huang, L. He, B. Yang, and X. Cheng, "Device-Free Crowd Counting Based on the Phase Difference of Channel State Information," in 2020 IEEE International Conference on Information Technology,Big Data and Artificial Intelligence (ICIBA), 6-8 Nov.

2020 2020, vol. 1, pp. 1343-1347, doi: 10.1109/ICIBA50161.2020.9276804.
[13] R. W. Schafer, "What Is a Savitzky-Golay Filter? [Lecture Notes]," IEEE Signal Processing Magazine, vol. 28, no. 4, pp. 111-117, 2011, doi: 10.1109/MSP.2011.941097.
[14] H.-S. Chen, H.-T. Chen, Y.-W. Chen, and S.-Y. Lee, "Human action recognition using star skeleton," presented at the Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, Santa Barbara, California, USA, 2006. [Online]. Available: https://doi.org/10.1145/1178782.1178808.
[15] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989, doi: 10.1109/5.18626.
[16] L. Hyeon-Kyu and J. H. Kim, "An HMM-based threshold model approach for gesture recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 961-973, 1999, doi: 10.1109/34.799904.
[17] M. J. F. Gales, "Maximum likelihood linear transformations for HMM-based speech recognition," Computer Speech & Language, vol. 12, no. 2, pp. 75-98, 1998/04/01/ 1998, doi: https://doi.org/10.1006/csla.1998.0043.
[18] S. Karpagavalli and E. Chandra, "Phoneme and word based model for tamil speech recognition using GMM-HMM," in 2015 International Conference on Advanced Computing and Communication Systems, 5-7 Jan. 2015 2015, pp. 1-5, doi: 10.1109/ICACCS.2015.7324119.
[19] Z. Liu, Z. Wu, T. Li, J. Li, and C. Shen, "GMM and CNN Hybrid Method for Short Utterance Speaker Recognition," IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3244-3252, 2018, doi: 10.1109/TII.2018.2799928.
[20] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11n traces with channel state information," vol. 41, no. 1 %J SIGCOMM Comput. Commun. Rev., p. 53, 2011, doi: 10.1145/1925861.1925870.
[21] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A Survey on Behavior Recognition Using WiFi Channel State Information," IEEE Communications Magazine, vol. 55, no. 10, pp. 98-104, 2017, doi: 10.1109/MCOM.2017.1700082.
[22] Tegou, T., Papadopoulos, A., Kalamaras, I. et al. Using Auditory Features for WiFi Channel State Information Activity Recognition. SN COMPUT. SCI. 1, 3 (2020). https://doi.org/10.1007/s42979-019-0003-2.

XIAOYAN CHENG received the B.S. degree from the College of Information and Electrical Engineering, China Agricultural University, in 2019. She is currently pursuing the M.S. degree from the School of Information and Communication Engineering, Xi'an Jiaotong University. Her research interests include device-free wireless human activity recognition and fall detection.

BINKE HUANG received the B.Sc. degree in Information and Communication Engineering and the Ph.D. degree in the Electromagnetic and Microwave Engineering from Xi'an Jiaotong University, Xi'an, China in 1998 and 2004, respectively. He is currently an Associate Professor with the School of Information and Communication Engineering, Xi'an Jiaotong University. From 2017 to 2018, he was a Visiting Scholar with the Department of Electrical Engineering (ESAT), KU LEUVEN, Belgium. From 2000 to 2001, he was a Research Assistant with the Department of Electrical engineering, City University of Hong Kong, Kowloon, Hong Kong SAR, P. R. China. His research interests include radar signal processing, modeling of microwave and millimeter wave components, wireless communication antennas, and computational electromagnetics.

JING ZONG received the B.S. degrees from the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, China, in 2018, where she is currently pursuing the M.S. degree. Her research interests include device-free wireless crowd counting and respiration detection.