Writing in the Air with WiFi Signals for Virtual Reality Devices

Zhangjie Fu, Jiashuang Xu, Zhuangdi Zhu, Alex X. Liu, and Xingming Sun

Abstract—Recently, handwriting recognition approaches has been widely applied to Human-Computer Interface (HCI) applications. The emergence of the novel mobile terminals urges a more man-machine friendly interface mode. The previous air-writing recognition approaches have been accomplished by virtue of cameras and sensors. However, the vision based approaches are susceptible to the light condition and sensor based methods have disadvantages in deployment and highcost. The latest researches have demonstrated that the pervasive wireless signals can be used to identify different gestures. In this paper, we attempt to utilize channel state information (CSI) derived from wireless signals to realize the device-free air-write recognition called *Wri-Fi*. Compared to the gesture recognition, the increased diversity and complexity of characters of the alphabet make it challenging. The Principle Component Analysis (PCA) is used for denoising effectively and the energy indicator derived from the Fast Fourier Transform (FFT) is to detect action continuously. The unique CSI waveform caused by unique writing patterns of 26 letters serve as feature space. Finally, the Hidden Markov model (HMM) is used for character modeling and classification. We conduct experiments in our laboratory and get the average accuracy of the *Wri-Fi* are 86.75 and 88.74 percent in two writing areas, respectively.

Index Terms—Air-write recognition, channel state information

1 Introduction

In this paper, we aim to design a device-free air-writing recognition system for Virtual Reality (VR) devices to take text input from users. An *air-writing recognition* system allows human users to write characters/symbols in the air as text input to VR devices. Inputing text has been a key challenge for VR systems because traditional text input mechanisms, such as keyboards, touch-screens, mouses, are particularly inconvenient for VR users. A system is *device-free* if it does not require users to wear any devices. With such a device-free air-writing recognition system, a user can easily input text for VR devices. The device-free air-write recognition pattern ensures the privacy information preserved [1], [2] because it has no access to collect individual sensitive information such as fingerprint, face.

Prior work on air-writing recognition falls into three categories from the perspective of data acquisition: vision-based [3], sensor-based [4], [5], and WiFi signals based [6], [7]. Vision based schemes use cameras to capture texts written by human users. The key limitation of such schemes is that they are susceptible to light conditions. Sensor based schemes use hand-held sensors, such as Wiimote and Leap Motion [8], [9], to capture user hand movements. The key limitation of such schemes is that they are not device-free and thus inconvenient to use. WiFi signal based schemes use WiFi signals to recognize user handwritings. Sun et al. use the Arrived of Angle (AoA) information derived from WiFi signals to track user's hand movement [7]. The key limitation of this scheme is that the air-writing recognition is realized by a ready-made handwrite recognition application called MyScript so it lacks of the research on the features of letters and classification.

The main contributions can be summarized as follows

- We propose *WriFi*, a WiFi based air-writing recognition system that uses the commercial off-the-shelf (COTS) WiFi devices to recognize human handwritings.
- Compared to the gesture recognition, the increased diversity and complexity of characters of the alphabet make it challenging.
- The Principle Component Analysis (PCA) is used for denoising effectively and the energy indicator derived from the Fast Fourier Transform (FFT) is to detect action continuously.
- The unique CSI waveform caused by unique writing patterns of 26 letters serve as feature space.
- We conduct experiments in our laboratory and get the average accuracy of the *Wri-Fi* are 86.75 and 88.74 percent in two writing areas, respectively.

2 Related Work

In this section, we discuss related work in three fields: wireless sensing, handwriting recognition, and hand-gesture recognition.

- *Z. Fu, J. Xu, and X. Sun are with the School of Computer Science and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China. E-mail: fzj@nuist.edu.cn, {xjsww1214, sunnudt}@163.com.*
- *Z. Zhu and A.X. Liu are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824. E-mail: {zhuzhuan, alexliu}@cse.msu.edu.*

Manuscript received 2 Sept. 2017; revised 6 Feb. 2018; accepted 12 Apr. 2018. Date of publication 4 May 2018; date of current version 7 Jan. 2019. (Corresponding author: Zhangjie Fu.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2018.2831709

2.1 Wireless Sensing

In recent years, researchers have proposed to use Wi-Fi signals for sensing purposes. Prior work has used Wi-Fi signals to detect the presence and location of an indoor human [10], [11], [12], to recognize gestures [13], [14], human activities and speech [15], [16], [17]; to count the number of people [18], [19]; and to identify human users [20], [21]. In 2013, Zhou et al. utilized the RSSI values to detect the presence of indoor humans [10]. In 2015, Abdelnasser et al. presented a system called *WiGest*, which uses RSSI fluctuations to recognize seven gesture families to control a media player application [14]. Due to the advent of CSI Tool in [22], we have access to the measurement of fine-grained CSI values, which are distinguished from the coarse valued RSSI. In 2017, Wang et al. proposed a fall detection system called *WiFall*, which utilizes finer-grained CSI to assist in elderly health supervision [23]. After that, Wang et al. presented a system called *WiHear* based on the claim that Wi-Fi signals can be used to determine what a person is saying by building mouth models for each syllable [17]. Wang et al. used CSI values to identify daily human activities by constructing a speed model that quantifies the inherent dynamic correlations between CSI variations and different motion speeds [15]. Ali et al. proposed a gesture-recognition system called *Wikey* to recognize the 37 keys on computer keyboards [24] . In 2016, Wang et al. proposed a novel authentication method by analyzing the CSI caused by people's unique walking gait [21]. Zeng analyzed the people's unique walking gaits to achieve identity authentication and proposed a system named *WiWho* [20].

2.2 Handwriting Recognition

Handwriting is a fundamental input mechanism for human-computer interaction. There are two main research goals in handwriting recognition. One is to identify a character that is written or printed on a surface or paper, and the other is to recognize characters/symbols that are written in an imaginary air box. Prior work on air-writing recognition falls into three categories from the perspective of data acquisition: vision-based [25], sensor-based [8], [9], and WiFi signals based [6], [7].

2.2.1 Vision-Based Handwriting Recognition

In conventional approaches, the user writes on the given paper or the touch screen. The schemes capture the manuscript and convert it to picture form [25], [26]. Then, the system extracts features of the character/symbol and perform the recognition based on the extracted features. Zhang et al. utilized the Kinect, a commercial feeling peripheral developed by Microsoft, to propose a new writing experience by continuously capturing the depth pictures of the writing procedure and analyzing the pixel values, shapes, etc [9]. In another common solution, the user writes on the touch screen.

2.2.2 Sensor-Based Air-Handwriting Recognition

In the last decade, the advancement of various types of sensors promoted the development of air-writing recognition approaches. The main challenge in air-handwriting recognition is the absence of the visual and haptic feedback.

Fig. 1. The illustration of the *WriFi* system.

In 1993, researchers integrated and attached various kinds of sensors to a glove called Data Glove, aiming to recognize 26 English letters [27]. The system used 26 different static gestures to represent the 26 letters in the alphabet. The user wears the Cyber Glove when performing static gestures. In practice, it is difficult for the user to remember the dictionary of 26 static gestures. In recent years, with the launch of handheld sensors and game machines, sensor-based handwriting recognition methods have been realized using somatosensory devices. In [3], researchers leveraged the Wiimote and Worldviz PPT-X4 to acquire 6-DoF hand motion data to recognize isolated characters, words and streams of words written in the air. In [4], [5], [28], air-writing recognition was realized with a glove attached with inertial sensors.

2.3 Hand Gesture Recognition

The hand gesture recognition schemes aim to recognize gestures, contributing to the smart-home system, sign language interpretation and motion sensing games. Hand gestures recognition solutions can also be categorized into vision-based [25], sensor-based [8], Wi-Fi signal-based [29], and ultrasound based [30]. Wang et al. used the depth camera from Kinect to address hand gesture recognition [31]. The scheme in [8] combined the Leap Motion controller and a depth camera. The gesture recognition solutions proposed by [30], [32] used ultrasound to remotely control a mobile phone. The research in [29] decoded the CSI of wireless signals to recognize slight hand gestures. The *WiDraw* system utilized wireless signals to depict the trajectory of hand movement. The *WiSee* in [33] used the Doppler Shift effect of wireless signals to recognize nine whole-home gestures.

3 *WriFi* System Overview

In this section, we illustrate how the *WriFi* system works. The *Wri-Fi* consists of a router as the sender to emit WiFi signals, and a laptop with dedicate network card as the receiver to receive WiFi signals, as shown in Fig. 1 and a user writes beside the Line of Sight (LoS) path. When a user writes characters/symbols in the air, the Chanel State Information (CSI) values that the receiver receives are impacted due to the multi-path effect of wireless signals. Because different character/symbols written by the user hand impact the CSI values differently, based on how the CSI values change, we recognize the character that the user writes.

In the data acquisition step, users perform air-writing in a unistroke and overlapped style, and the system collects CSI measurement at the mean time. In the data preprocessing step, *WriFi* adopts four steps to de-noise the collected CSI singals: normalization, covariance matrix calculation, eigen decomposition, and principal component calculation. In the action detection stage, *WriFi* applies a sliding window to the CSI stream to calculate the FFT-based motion energy

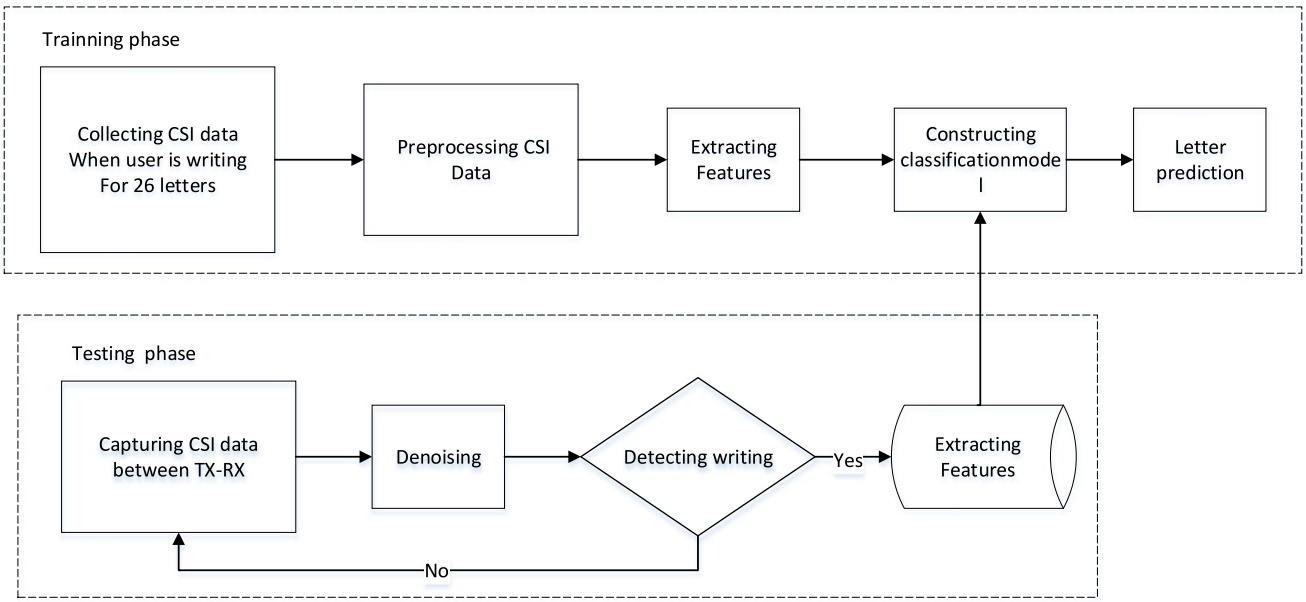


Fig. 2. *Wrif-Fi* workflow.

as metric to discriminate four common indoor behaviours. In the feature extraction step, *WriFi* extracts the four kinds of waveform characteristics of 26 letters and resolves a letter into stroke-level to analyze the speed variation when writing ligatures. In the training step, *WriFi* builds 26 Hidden Markov Model (HMM)s, one for each individual English letter using features extracted from the denosied CSI values. In the testing step, *WriFi* first uses an energy indicator to detect the presence of a writing action, and then recognize this letter by fitting the letter to an HMM model that has the maximum likelihood to generate it. Fig. 2 displays the workflow of the WriFi system.

3.1 Data Acquisition

Unistroke and Overlapped Writing Style. We make two important assumptions in this paper. The first assumption is that *WriFi* users adopt the same writing style called unistroke. Unistroke is a writing style that requires the user to complete a letter in one continuous stroke. Fig. 3 illustrates eight different letters written in the unistroke writing style.

As writing habits vary from people to people, if users writes in different manners, such as stroke-by-stroke manner and reserved-stroke manner, we have to train separate models for the same letter. Therefore, We unify the writing style as unistroke style to improve the effectiveness of the model-training procedure so that we only need to train one model for each letter. The second assumption is that all the individual letters are written in the overlapped manner. Otherwise, if the user writes from the left to right in the air, the other parts of the body, along with finger and arms, will move as well. As a result, the data we obtain is not pure enough which contains the writing component but also the component of other parts movement. Fig. 4 shows the overlapped manner for the handwritten letters 'A' and 'B'. These letters are written on the touch screen of an iPad as illustrations.

In conclusion, we assume our writing styles as unistroke and overlapped.

CSI Measurement. WriFi collects Channel state information measurements from the targeted device. CSI is a physical-layer information that are available in many commodity network interface cards, such as Intel 5300 and Atheros 9390. It describes the multi-path fading, Doppler shift and scattering effect in the specific scenario and characterizes the quality of the communication link. The MIMO-OFDM (Multiple In Multiple Out/ Orthogonal Frequency Division Multiplexing) system in the standard of IEEE 802.11/n consists of multiple transmit-receive TX-RX antenna pairs to increase the capacity of the communication link for subcarrier-scale information. The MIMO-OFDM modulation scheme is introduced to ensure finer-grained estimated physical-layer information for the CSI. The OFDM technique divides the channels into a specific number of orthogonal subcarriers, which reduces the

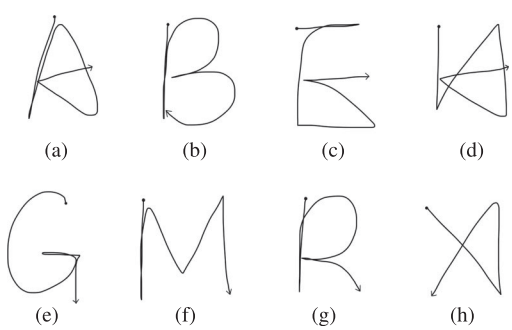


Fig. 3. Unistroke writing style for eight letters.

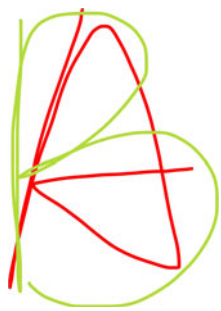


Fig. 4. A illustration for two overlapped letters written on iPad.

Inter- Symbol Interference (ISI), and the transmitted signal is modulated at different frequencies and carried in the OFDM subcarrier groups. The relation between the transmitter and receiver ends is shown below

$$Y_i = H_i \times X_i + N_i. \tag{1}$$

In which X_i and Y_i stand for transmit and receive vector, respectively. N_i is additive white Gaussian noise vector at ith subcarrier. H_i is a $N_{Rx} \times N_{Tx}$ channel gain matrix, quantifying the multi-subcarrier the representation for, N_{Tx} and N_{Rx} are the amount of TX and RX antennas, respectively. The gain matrix can be estimated by the following equation:

$$H_i = \frac{Y_i}{X_i}. \tag{2}$$

The element H_i in estimated complex valued CSI matrix for each subcarrier is define as equation

$$H_i = |h_i| e^{j\angle H}. \tag{3}$$

In which, $|h_i|$ is the amplitude and phase at i th OFMD subcarrier. In our case, we select 2 transmit antennas and 3 receive antennas, respectively. Thus, the MIMO system contains 2×3 communication links. Each CSI stream is divided into 30 orthogonal subcarriers. Thus, there are 30 matrices with dimension 2×3 in one packet. It also can be represented by the following format:

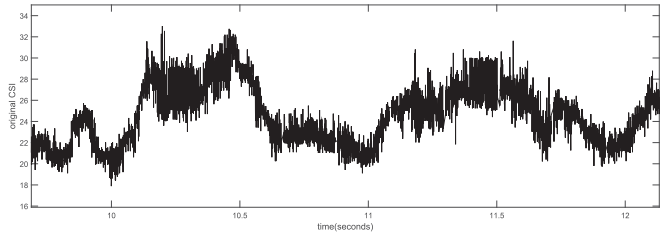
$$\begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,30} \\ H_{2,1} & H_{2,1} & \cdots & H_{2,30} \\ \vdots & \vdots & \ddots & \vdots \\ H_{6,1} & H_{6,2} & \cdots & H_{6,30} \end{bmatrix}. \tag{4}$$

In which $H_{i,j}$ is the element in the information gain matrix H, representing the CSI value of the jth subcarrier in the ith antenna pair in one CSI stream.

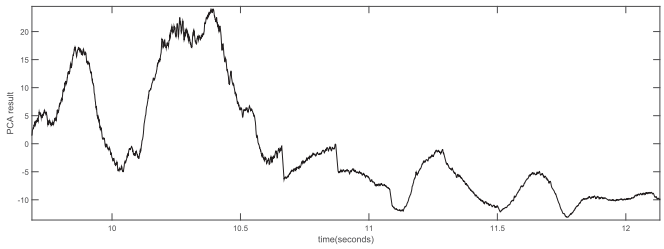
3.2 Data Preprocessing

We cannot use the original CSI values to recognize motion characters accurately. As we can infer from Fig. 5, the captured CSI time-series streams are noisy and contain many outliers. In theory, a low pass filter such as the Butterworth filter, should perform well in eliminating high frequency noise. However, these kind of traditional low pass filters cannot remove burst and impulse noise well in CSI time-series. The bursty and impulse noise, which derive from the transmission power and rate, changes unpredictably between the transmitter and receiver ends. Moreover, the use of a rigorous low-pass filter may result in the loss of useful features. To address this problem, we use the Principle Component Analysis to track the entire CSI time-series. PCA is an effective dimension reduction algorithm. It calculates correlations among dimensions to find the correlated components and eliminate redundant noises. Previous work proved that the motion is correlated in CSI streams, while the burst noises are uncorrelated with motion data [15]. Our de-nosing approach consist of four procedures:

(1) Normalization: for the consecutive CSI stream, we get $N \times S_n$ dimensional CSI matrix $H_{t,r}$, which can be formed by following representation



(a) The Captured Original CSI Waveform.



(b) The Captured Denoised CSI Waveform.

Fig. 5. The original CSI waveform and the denoised CSI waveform for writing the letter 'A'.

$$H_{t,r} = \left[H_{t,r}(1)|H_{t,r}(2)|\cdots|H_{t,r}(N)\right]^T. \tag{5}$$

In which N is the number of CSI packets and S_n is the total number of subcarriers. $H_{t,r}(m)$ is the $S_n \times 1$ dimensional vector indicating the m th CSI packet at the (t, r) antenna pair. We normalize the $H_{t,r}(m)$ to obtain the $Z_{t,r}(m)$ with zero mean and unit variance

$$H_{m,k} = \frac{H_{m,k} - \bar{H}_k}{S_k}. \tag{6}$$

In which $H_{m,k}$ is the array element in $H_{t,r}$, $\bar{H}_k$ is the mean of the kth dimension, and S_k is the standard deviation of jth dimension.

(2) Calculating Covariance Matrix: we calculate the covariance for the normalized $Z_{t,r}(i)$ among dimensions and construct the covariance matrix with dimension of $n \times n$, in which n is the number of subcarriers. In our case, we choose n to be 180.

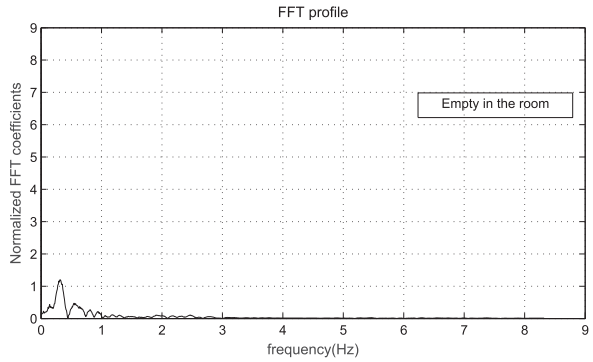
(3) Eigen decomposition: we calculate the eigenvalues e and the corresponding eigenvectors q for the covariance matrix. After that, we calculate the contribution scores, rank the eigenvalues and choose top 10 eigenvalues and the corresponding eigenvectors.

(4) Calculating principle components: the principle components can be decomposed by $H \times q$.

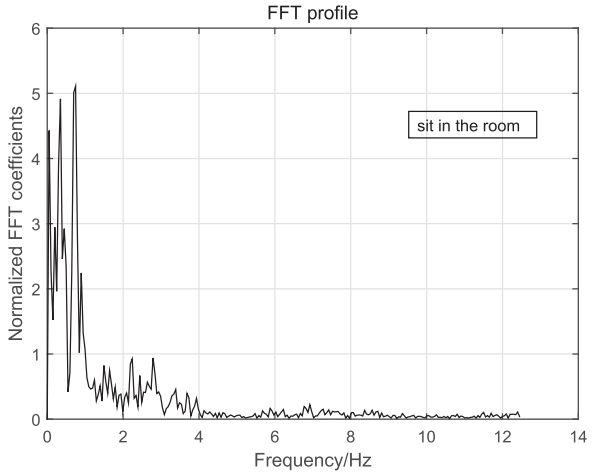
Using PCA, we reduce the dimensionality of the sample data by means of mapping m dimensional to n dimensional principle components (where $n < m$). The features are mutually linearly independent, and n is a linear representation of m. As a result, we obtain the denoised data with dimension $N \times 10$. Therefore, the computational cost is greatly decreased. Figs. 5a and 5b display the original CSI time-series waveform and the PCA-based denoised waveform for writing the character 'A', respectively.

3.3 Writing Detection

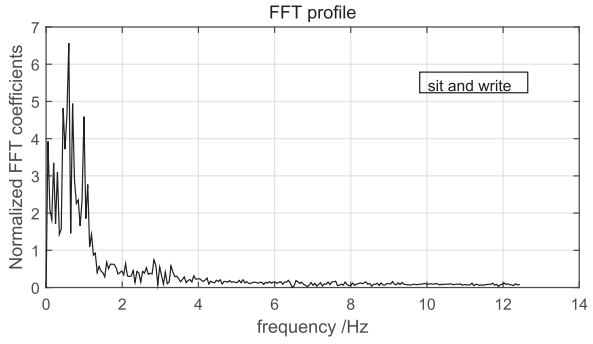
Since our work does not depend on any wearable sensor or motion tracker, it is difficult for us to check whether the



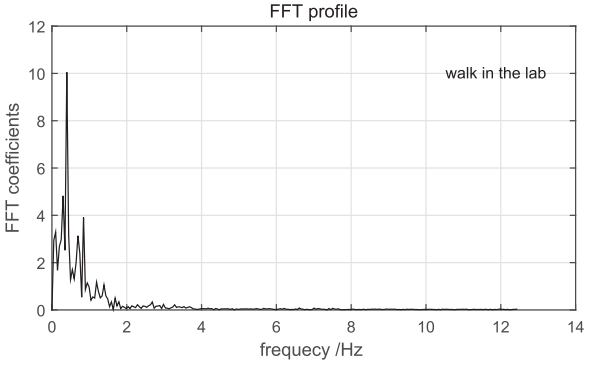
(a) The FFT profile of the empty room



(b) The FFT profile of the user writting in the air



(c) The FFT profile of the user with no writing



(d) The FFT profile of the user walking into the room

Fig. 6. The energy for four states from two users.

user is writing or not. Wang et al. proved that the energy intensity and frequency differ greatly for different motions, such as walking, running, sitting and empty [15]. Therefore,

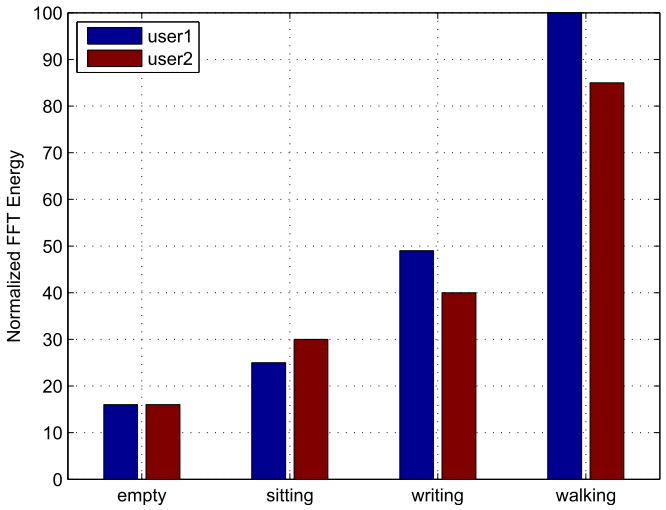


Fig. 7. The energy for four states from two users in a 2s window.

we obtain the Fast Fourier Transform profiles of different states to verify the feasibility of using frequency domain characteristics in motion detection. We take four scenarios into consideration

1) No one is in the room.
2) A user is in the room with no air-writing action.
3) A user is in the room and is writing in the air.
4) A user is walking into the room.

Fig. 6 depicts the FFT frequency profiles for the above scenarios, respectively. We can learn from the figure that in the empty room scenario, the energy is quite low compared with the scenarios of sitting, writing or walking.

We introduce an energy indicator to detect the presence of writing motion. It uses a time window to track the FFT of the denoised CSI values. The indicator calculates energy as follows:

$$E = \sum_{i=1}^{n/2} mag^2. \tag{7}$$

In which E is the calculated energy, n is the length of the time window, and mag is the normalized FFT coefficients calculated over the time window per 200 ms. Fig. 7 shows the instant FFT profiles of four indoor scenarios. We can see that different scenarios generate different energy levels.

To approximately find the start and end point of actions, *WriFi* persistently monitors the difference of short-term motion energy in two consecutive windows. Once the difference is greater than positive threshold we predefined, *WriFi* considers there exists action in current scenarios. Once an action is ended, the motion energy in current window decreases sharply. Thus, if the difference is smaller than the negative threshold, *WriFi* consider the action has ended. In practice, we set two seconds for people to perform 4 activities in order to range the motion energy level.

3.4 Feature Extraction

After detecting the starting and ending time of writing, we extract features from the denoised CSI time series containing the writing action. Our feature extraction approaches are based on two perspectives. First, we analyze a written letter as a whole unit, and generate a unique profile for each

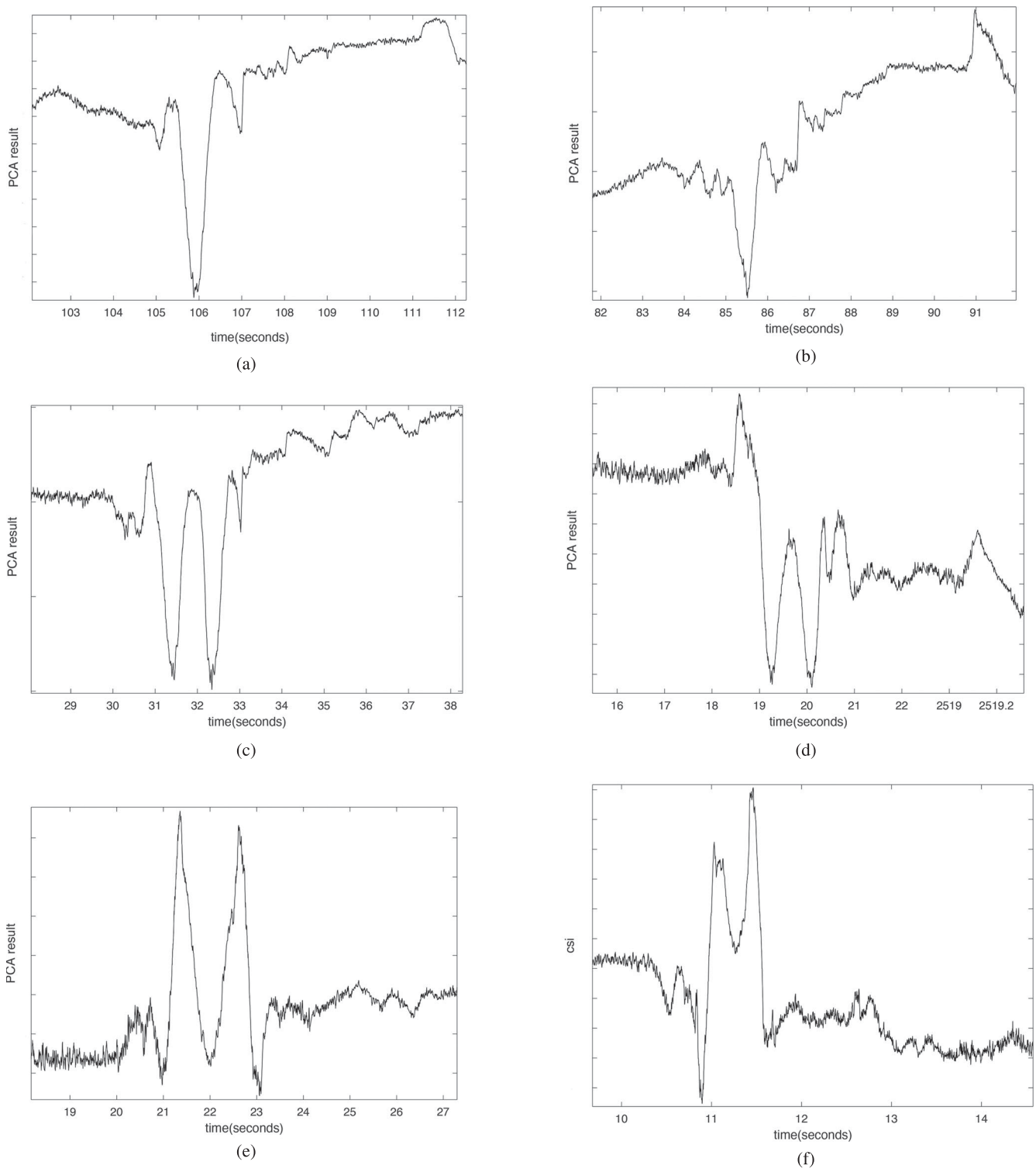


Fig. 8. The CSI waveforms for three letters from different subjects.

letter extracted from the CSI waveform. Second, we take a finer granularity and analyze the strokes that compose each letter. Especially, we use time-frequency analysis to extract features that represent changes in the motion speed caused by these strokes.

Letter Analysis. For letter analysis, we concentrate on the profile of each letter as a whole. We build profiles for each letter based on the fact that the structure of each letter varies comparatively. For instance, the structure of letter 'I' is much simpler than that of letter 'E'. The difference in the structure of 26 letters leads to the difference in the writing duration and writing trajectory, which therefore causes the difference in the CSI waveforms when writing each letter. The Fig. 8 shows the CSI waveforms of writing three letters by two different users. Figs. 8a and 8b show the CSI waveforms for writing the letter 'L', Figs. 8c and 8d show the CSI waveforms for writing the letter 'P', and Figs. 8e and 8f show the CSI waveforms for writing the letter 'E'. We can

observe from the figure that the waveforms are similar when writing the the same letter by different users, but are highly different when writing different letters even by the same user. Therefore, we track the CSI time-series to extract the profile of each letter. We choose four characteristics as the profile, which represent the shape and trend of the CSI waveform:

- peak factor: the ratio of the peak and root mean square value.
- wave factor: the ratio of the root mean square and mean value.
- coefficient dispersion: the ratio of the standard deviation and mean value.
- autocorrelation coefficient.

Stroke Analysis. For stroke analysis, we focus on features that represent strokes contained in each letter. In the unistroke writing style, all strokes of a letter are jointed by ligature, and the writing directions and speed change instantly when turning to another stroke. We use a time-frequency analysis tool to analyze the changes in writing speed caused by different strokes. Especially, we use Discrete Wavelet Transform (DWT) to decompose the CSI waveform into 10 levels. We calculate the total energy from each level for the entire writing action, and the energy distributed in each level represents the current writing speed.

In conclusion, we choose above features from two perspectives at two different granularities for further letter recognition.

(The system in [6] calculated Dynamic Time Warping (DTW) distances to match the waveform template for classification. However, the limitations lie in three aspects: heavy computation; it ignores the temporal dynamic characteristics of writing signal; it is too dependent on the initial writing waveform.)

3.5 Modeling and Training

we propose a Hidden Markov Model based writing recognition approach. Especially, we employ 26 Gaussian Mixture Model-Hidden Markov Models (GMM-HMMs) to recognize 26 air-writing English letters. HMM is one of the commonly used methods of dynamic gesture recognition, this is because the characteristics of dynamic hand gestures can extract contain time information, the HMM can provide time scale deformation which owns flexible and efficient training and recognition algorithms. For training, we use the denoised CSI time-series as training samples to build 26 HMMs in total. Each model corresponds to an individual English letter, and is trained using samples of writing that letter. For testing, we evaluate an unknown sample against HMMs of all letters, and chose the model that has the highest likelihood to generate the sample.

3.5.1 Hidden Markov Model

Hidden Markov Model is a suitable tool to recognizing activities with time-evolving feature vectors. It is widely used in applications like speech recognition and gesture recognition [34], [35]. HMM is used based on the assumption that the activity being modeled is a Markov process. This process consists of visible observations, and each observation corresponds to a hidden state that is invisible to the observer [36]. The probabilities of transitions from one state to another are called transition probabilities. In our paper, we observe that an English letter is composed of strokes, and when writing an letter, the transitions among strokes can cause subtle changes in movement speed, which correspondingly cause changes in CSI waveforms. Therefore, we consider writing a letter as a Markov process, in which the visible CSI waveforms are observations, and the transition among strokes are hidden states. The probabilities of transitions among different strokes are determined by the unique structure of the letter itself.

TABLE 1
Writing Duration of Motion Characters by Five Subjects Written in the $25 * 25cm$ Imagery Box

	avg		avg		avg
A	2.58	J	1.71	S	1.85
B	2.52	K	2.43	T	1.76
C	1.65	L	1.59	U	1.74
D	2.14	M	2.53	V	1.42
E	2.89	N	2.16	W	2.20
F	2.63	O	1.72	X	1.88
G	2.42	P	1.97	Y	1.92
H	2.38	Q	2.15	Z	1.89
I	1.34	R	2.38		

3.5.2 Model Parameterization

Different activities contain different number of hidden states. To determine the number of states for HMMs of each letter, we set the number of states in proportion to the writing duration for each letter. We record the durations for writing each of the 26 letters by different users. Table 1 displays the average writing durations of five users in the 25×25 cm writing region. We can see that letters with more complex structures have more writing durations.

For each HMM, we train its transition probability using Expectation Maximization (EM) algorithm. First, we initialize the parameter set λ by estimation. λ contains a state transmission matrix $a_{i,j}$, a initialization probability π, and a confusion matrix $b_{i,k}$. In the expectation stage, the $P(O|\lambda)$ is obtained by the initialized parameters. In the maximization stage, the parameter set λ is re-estimated and updated to $\overline{\lambda}$ to achieve $P(O|\lambda) > P(O|\overline{\lambda})$ with backward algorithm. Ultimately, the maximum likelihood and optimized parameters for modelling can be obtained after iterations.

4 Implementation and Evaluation

4.1 Implementations

In our experiment, the whole system is composed of a TP-Link router which serves as the Access Point (AP) and a laptop which serves as the Detect Point (DP). Both of them are off-the-shelf commercial devices. The router is TP-Link WRD6300. The laptop is Dell M411R with 4 GB RAM, equipped with Intel 5300 802.11n NIC and installed with OS Ubuntu 12.04 LTS. The AP and DP have two and three external omnidirectional antennas, respectively. The CSI values is provided by the modified driver of the NIC, which is developed by [22]. The dimension of the obtained CSI values are determined by the number of transmit-receive antenna pairs. In our case, the CSI is In addition, we use an

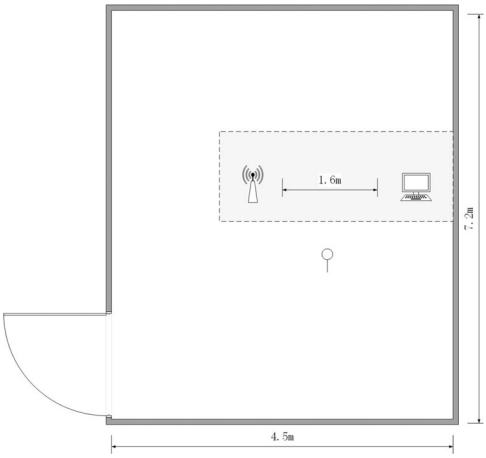


Fig. 9. The picture of the experimental scenario.

extra laptop to send packets to the receiver end via the router using *iperf* tool [37]. The sample rate is 2,500 samples/s. The high sample rate ensures that abundant detailed writing data are obtained for analysis. The system runs on Matlab2014b. We select the 5 GHz-frequency band with shorter wavelength, which has a higher frequency resolution.

4.2 Evaluation Setup

Our experiment is conducted in a university laboratory, which is 7.5 m in length and 4.5 m in width We choose five right-handed students (two females and three males) aged 20-23 as volunteers to collect the CSI data. All letter are written with the index finger on or above the appointed area of the white board, which is located beside the TX-RX pair. The letters are written with hand and arm movements other than wrist movements. In addition, volunteers are asked to keep other parts of the body steady. Fig. 9 displays the layout of the experimental scenario. The distance between the AP and the DP is 1 m. The distance is short because we want to obtain a high Signal-to-Noise Ratio (SNR). It takes approximately 15 minutes or each volunteer to write a single letter 100 times. Every single letter in the alphabet is collected 200 times from each volunteer as training samples. Half of the samples of each letter are written in the size of 25×25 cm, and the remaining 100 samples of the letter are written in the size of 35×35 cm. We note that this is a reasonable size for volunteers to write in the air. We construct two models for each letter to observe the impact of writing size on letter- recognition accuracy, which will be elaborated in this section.

4.3 Writing Detection Analysis

To evaluate the detection accuracy, we introduce two metrics: the true positive rate (TPR) and false alarm rate (FAR). The TPR is the ratio of the number of correctly detected writing actions to the total number of detections. The FAR is the ratio of the number of incorrect writing-action detections of other states to the total number of detections.

We invite five volunteers to carry out the detection test. We conduct experiments on four indoor scenarios (empty, sitting, writing, walking), each repeated 100 times. For the writing scenario, we randomly select 10 letters of the alphabet randomly to analyze writing detection accuracy. All the activities are behaved next to the LoS path for 30 centimeters.

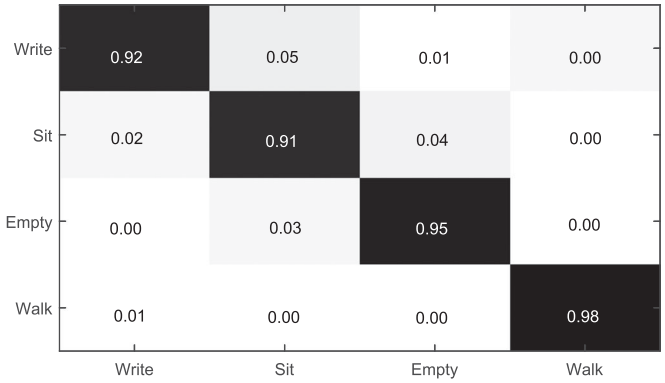


Fig. 10. The confusion map for four scenarios.

The grey-scale colormap from the confusion matrix for the four indoor situations is shown in Fig. 10. We can learn from the colormap that *WriFi* can detect the writing action with a TP rate of 92 percent and can distinguish empty and writing scenarios without any mistakes. We can also infer from the colormap that the FAR of *WriFi* is 6 percent. The detection accuracy of distinguishing the sitting and writing scenarios is 91 percent. In addition, the *WriFi* system can effectively distinguish the walking scenario from the others, which have lower energy. In conclusion, the average detection TPR of the four indoor scenarios is 94.0 percent.

To evaluate the robustness of action detection stage of *WriFi*, we invite other 5 untrained volunteers for detecting test. They are required to perform sit, write (same characters as the trained people wrote) and walk for 100 time, respectively, which is the same as the five trained people. Fig. 11 displays the detection result of untrained people in the distance of 30 cm away from LoS path. The TPR of detecting walking reaches the highest average accuracy, 93.80 percent and the TPR of detecting writing and sit are 88.80, 88.6 percent among five untrained users. In conclusion, the average detecting accuracy reached 90.40 percent We can conclude that the action detection method is reliable to detect and discriminate common indoor activities.

4.4 Writing Recognition Analysis

We take two factors that affect the accuracy of letter recognition into consideration: the training-sample size and the writing- region size.

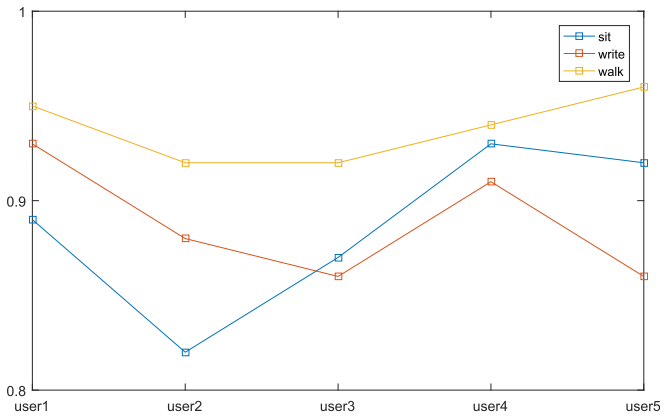


Fig. 11. The detection result for untrained five users among three indoor activities.

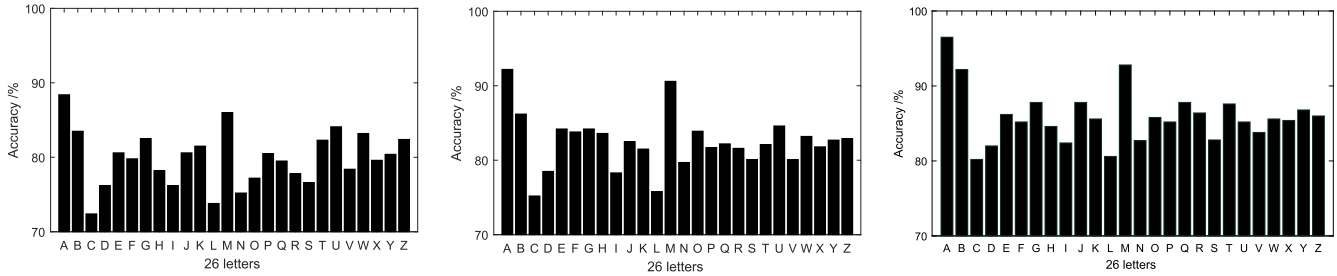


Fig. 12. The cross validation results for 26 letters in three training samples. (a) 30 samples, (b) 70 samples, and (c) 100 samples.

4.4.1 Impact of the Training Sample Size on Recognition Accuracy

In this section, we evaluate the accuracy of the HMM model with different sample sizes. It is worth mentioning that the writing region size used in this section is 25×25 cm. We conduct three groups of experiments to determine the relationship between the accuracy and the number of training samples. We build HMMs with 30, 70, and 100 samples per letter from all voluneteers, respectively.

First, we collect 30 samples for each letter from five voluneteers, and build HMM models. Then, we test the performances of the constructed models with leave-one-out cross validation. In the next experiments, we increase the number of samples from 30 to 70 and 100, respectively. Similarly, we build HMMs for each letter.

The overall recognition accuracy for all letters in the 30 samples reaches 80.21 percent. The average accuracy of cross validation increases to 83.69 percent when the number of samples for each letter rises is increased from 30 to 70. Figs. 12a and 12b show the average results of leave-one-out cross-validation for the 26 letters of the alphabet. The average accuracy of the cross validation results for 100 samples reaches 86.75 percent over all letters from all volunteers. Fig. 12c shows the average results of 100 samples for each letter. We present the grey-scale colormap of the confusion matrix obtained after leave-one-out cross validation on 100 training samples for subject 5 in Fig. 13. The colormap records the number of times a certain letter is recognized correctly. The misclassification of characters results from the ambiguity of the letters. The writing patterns are similar between the pair of C and O; D and P; T and J, etc. After examining the average recognition results with different sample sizes, we can conclude from Fig. 12 that the accuracy raises with the number of samples of motion letters.

4.4.2 Impact of the Writing Region Size on Recognition Accuracy

In this section, we will discuss the changes in the accuracy of the system when the writing region is expanded. The expansion of the writing region results in a larger motion amplitude when writing a letter. We consider two sizes of the writing area: 25×25 cm and 35×35 cm. As mentioned above, we ask volunteers to write in two different areas, so we train two separate models for two regions. We still use the data of 100 samples from five subjects for training and leave-one-out cross validation.

The average accuracy across increases from 86.75 to 88.74 percent when the writing area is expanded. Fig. 14 shows the comparison of the average cross validation results for 26 letters in the two writing regions. The rise in accuracy indicates that the movement range has an impact on the recognition accuracy: the CSI values are so susceptible that the accuracy fluctuates more observably when the range of the user action becomes larger. Moreover, the longer writing duration and larger movement range allow the system to capture the writing action more effectively. In practice, however, people are not willing to write in over large space. Thus, to obtain high accuracy, we need to recruit more people to produce training samples.

Compared to the *Wi-Wri* system in [6], the recognition accuracy of the proposed approach is higher. First, we collected more samples for each letter, and different writing

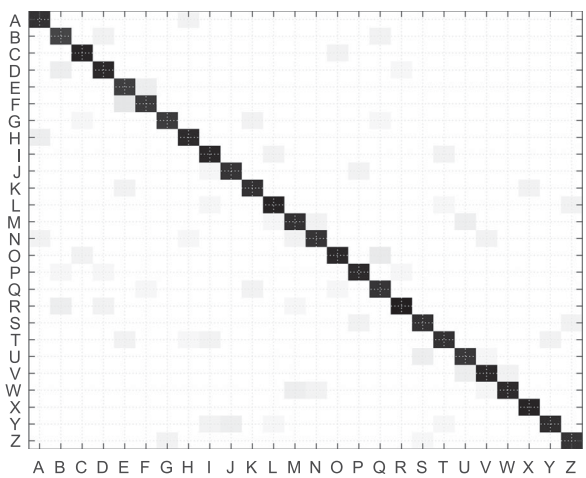


Fig. 13. The confusion matrix for the cross validation result from user5.

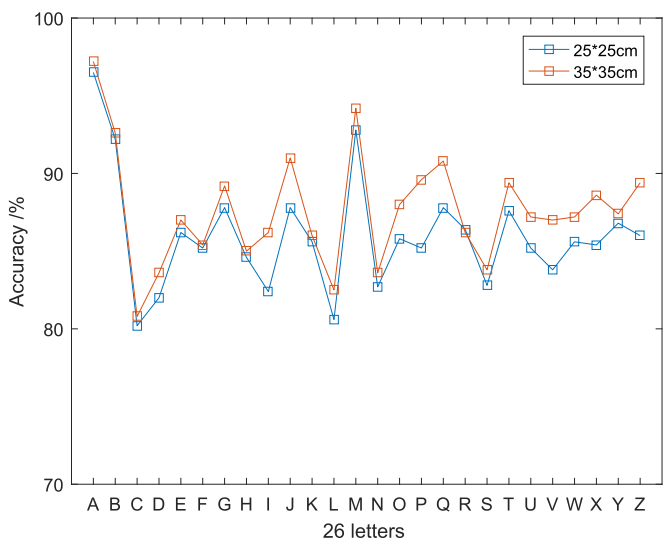


Fig. 14. The cross validation results for 26 letters in two writing areas.

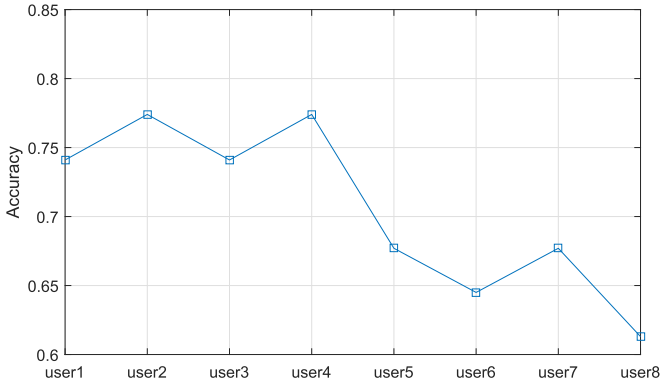


Fig. 15. The real-time accuracy for the sentence covering 26 letters.

region also impacts the recognition accuracy. In addition, *Wi-Wri* uses the warped waveforms as feature for classification, which reserves the most information while breaks the ligature which joints connective strokes and ignores the detailed information.

In contrast to WiDraw system, the accuracy of WiDraw reaches 96 percent. By contrast, our system accuracy is lower than WiDraw. This is because WiDraw utilized a commercial application to recognize the air-written letters which owns robust algorithm and affluent data as support. In this paper, we proposed a recognition system which do not rely on the off-the-shelf applications.

5 Usability Study

The air-writing recognition system aims to realize the motion-based text input scheme on smart devices. To evaluate the performance of the system, we take two factors into consideration: the recognition accuracy and efficiency in real time. In addition, the detection range should also be considered. According to the formula $c = \lambda f$, c is the velocity of the electromagnetic waveform and the f is 5 GHz in our case. the minimum detectable range is 6 centimeters. We select a sentence covering the 26 letters for real-world accuracy testing, in which the letters of the alphabet are distributed evenly.

5.1 Real-World Recognition Accuracy of WriFi

In this section, we will discuss the recognition accuracy of the system in practice. We select a sentence covering all 26 letters for evaluation: *The five boxing wizards jump quickly*. The sentence contains 31 letters. All five volunteer are required to write the sentence letter by letter. The volunteer is required to suspend writing among letters for 2 seconds as a cushioning time. The average accuracy of real-time recognition reaches 74.84 percent. We introduce a common metric, the Character Error Rate (CER), to evaluate the performance on text input accuracy. Fig. 15 displays the accuracies for the five users on the above sentence. The accuracy includes cases in which the system mistakenly detects the writing action as sitting.

In addition, we also invited three untrained students to participate into the real-time accuracy tests to evaluate the robustness of *wri-Fi*. They were required to obey the overlapped unistroke writing rule and wrote the sentence above letter by letter. As a result, the accuracy of three untrained students reaches 64.5 percent.

5.2 Real-World Recognition Efficiency of WriFi

We introduce the letters per minute to calculate the number of letters the system recognizes in one minute. It is a common metric for examining the efficiency of a text input system. We ask the volunteer stop for a 2-second interval before writing the next letter. They are required to write all the letters in alphabetic orderly and record the entire time spent. The average time the system takes to recognize all 26 letters is 124.82s across all five volunteer. Thus, the number of letters recognized per minute is 12.50 (=26/124.82 s). For the virtual keyboard, which requires the user to point and click, the Words Per Minute (WPM) of text input is 8.42, which is higher than the WPM of the air-writing approach. However, air-writing is a more intuitive and primitive way to input text.

6 Discussion

There are two limitations in our air-write system.

First, since the CSI derived from Wi-Fi signals is susceptible to the surrounding environment, our air-write system can perform well in the little interference place. In the collecting data phase, there are two people in the lab, one for collecting data and another one is to guide the subject to write. The guidance people either cannot act too intensely to reduce the interference to Wi-Fi signals.

Second, the writing rule is formulated by ourselves and is suitable for majority of the subjects. However, The untrained people may be unaccustomed to the writing rule. At present, the system can only work well in the stable environment. The accuracy is affected by the writing habitat, surrounding human motion.

7 Conclusion

In this paper, we make the following contributions. First, we propose a novel air handwriting recognition framework without any other specialized sensors which meet the device-free, low-cost and pervasive concept. We demonstrate that Wi-Fi signals can also be used in distinguishing 26 individual capitalized English letters. We denoise the collected CSI data using PCA based denoising algorithm. The system detects the writing action persistently by FFT based energy indicator. After obtain the motion character data, the unique writing patterns of 26 letters result in the unique CSI waveforms, we extract the waveform characteristics to depict the waveform shape and DWT coefficients to represent current writing speed. The features provide the observation for building HMMs. As a result, the average accuracy of detection method reaches to 94.0 percent. In the case of 100 training samples written in larger space, the cross validation result reaches the highest, 88.74 percent. In the future, we will continue to add the 10 digits 0-9 into the air-write recognition research. We will train separate models for the same letter to accommodate the diversity of the writing habits. Furthermore, the word-based air-write recognition will be explored deeply as well. Due to the Wi-Fi connection is not stable sometimes, so we consider to use the directional antenna for more stable connection and higher signal intensity.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grant 61772283, U1536206,

U1405254, 61602253, 61672294, 61502242; by the National Key R&D Program of China under grant 2018YFB1003205; by the Jiangsu Basic Research Programs-Natural Science Foundation under grant numbers BK20150925 and BK20151530; by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund; by the Major Program of the National Social Science Fund of China (17ZDA092), Qing Lan Project, Meteorology Soft Sciences Project; by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China.

REFERENCES

[1] C.-Z. Gao, Q. Cheng, P. He, W. Susilo, and J. Li, "Privacy-preserving naive bayes classifiers secure against the substitution-then-comparison attack," *Inf. Sci.*, vol. 444, pp. 72–88, 2018.

[2] J. Li, X. Chen, S. S. Chow, Q. Huang, D. S. Wong, and Z. Liu, "Multi-authority fine-grained access control with accountability and its application in cloud," *J. Netw. Comput. Appl.*, vol. 112, pp. 89–96, 2018.

[3] M. Chen, G. AlRegib, and B.-H. Juang, "Air-writing recognition—Part I: Modeling and recognition of characters, words, and connecting motions," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 403–413, Jun. 2016.

[4] C. Amma, D. Gehrig, and T. Schultz, "Airwriting recognition using wearable motion sensors," in *Proc. 1st Augmented Human International Conf.*, 2010, Art. no. 10.

[5] T. Költringer, P. Isokoski, and T. Grechenig, "TwoStick: Writing with a game controller," in *Proc. Graph. Interface*, 2007, pp. 103–110.

[6] X. Cao, B. Chen, and Y. Zhao, "Wi-Wri: Fine-grained writing recognition using Wi-Fi signals," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, 2016, pp. 1366–1373.

[7] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 77–89.

[8] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and Kinect devices," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 1565–1569.

[9] X. Zhang, Z. Ye, L. Jin, Z. Feng, and S. Xu, "A new writing experience: Finger writing in the air using a Kinect sensor," *IEEE MultiMedia*, vol. 20, no. 4, pp. 85–93, Oct.–Dec. 2013.

[10] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Towards omnidirectional passive human detection," in *Proc. IEEE INFOCOM*, 2013, pp. 3057–3065.

[11] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 269–282, 2015.

[12] 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 Int. Conf. Mobile Syst. Appl. Services*, 2012, pp. 183–196.

[13] W. He, K. Wu, Y. Zou, and Z. Ming, "WiG: WiFi-based gesture recognition system," in *Proc. 24th Int. Conf. Comput. Commun. Netw.*, 2015, pp. 1–7.

[14] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 1472–1480.

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

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

[17] 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.

[18] Y. Wang, J. Yang, Y. Chen, H. Liu, M. Gruteser, and R. P. Martin, "Tracking human queues using single-point signal monitoring," in *Proc. 12th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2014, pp. 42–54.

[19] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only WiFi power measurements," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1381–1393, Jul. 2015.

[20] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in *Proc. 15th Int. Conf. Inf. Process. Sensor Netw.*, 2016, Art. no. 4.

[21] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 363–373.

[22] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, pp. 53–53, 2011.

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

[24] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 90–102.

[25] A. Schick, D. Morlock, C. Amma, T. Schultz, and R. Stiefelhagen, "Vision-based handwriting recognition for unrestricted text input in mid-air," in *Proc. 14th ACM Int. Conf. Multimodal Interaction*, 2012, pp. 217–220.

[26] L. Jin, D. Yang, L.-X. Zhen, and J.-C. Huang, "A novel vision-based finger-writing character recognition system," *J. Circuits Syst. Comput.*, vol. 16, no. 03, pp. 421–436, 2007.

[27] S. S. Fels and G. E. Hinton, "Glove-talk: A neural network interface between a data-glove and a speech synthesizer," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 2–8, Jan. 1993.

[28] C. Amma, M. Georgi, and T. Schultz, "Airwriting: Hands-free mobile text input by spotting and continuous recognition of 3D-space handwriting with inertial sensors," in *Proc. 16th Int. Symp. Wearable Comput.*, 2012, pp. 52–59.

[29] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, "AudioGest: Enabling fine-grained hand gesture detection by decoding echo signal," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 474–485.

[30] D. M. van Willigen, E. Mostert, and M. A. Pertijs, "In-air ultrasonic gesture sensing with MEMS microphones," in *Proc. IEEE SENSORS*, 2014, pp. 90–93.

[31] C. Wang, Z. Liu, and S.-C. Chan, "Superpixel-based hand gesture recognition with Kinect depth camera," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 29–39, Jan. 2015.

[32] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 82–94.

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

[34] M. J. Gales, "Maximum likelihood linear transformations for HMM-based speech recognition," *Comput. Speech Language*, vol. 12, no. 2, pp. 75–98, 1998.

[35] H.-K. Lee and J.-H. Kim, "An HMM-based threshold model approach for gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 961–973, Oct. 1999.

[36] Z. Liu, Z. Wu, T. Li, J. Li, and C. Shen, "GMM and CNN hybrid method for short utterance speaker recognition," *IEEE Trans. Ind. Informat.*, vol. PP, no. 99, p. 1, 2018.

[37] A. Tirumala, F. Qin, J. Dugan, J. Ferguson, and K. Gibbs, "iPerf: The TCP/UDP bandwidth measurement tool," 2005. [Online]. Available: http://dast.nlanr.net/Projects

Zhangjie Fu received the PhD degree in computer science from the College of Computer, Hunan University, China, in 2012. He is currently an associate professor with the School of Computer Science and Software, Nanjing University of Information Science and Technology, China. He was a visiting scholar of computer science and engineering with the State University of New York at Buffalo from 2015 to 2016. His research interests include cloud security, outsourcing security, digital forensics, network, and information security. His research has been supported by NSFC, PAPD, and GYHY. He is a member of the ACM.

Jiashuang Xu received the BE degree in network engineering from the Nanjing University of Information Science and Technology, China, in 2015. She is currently working toward the MS degree in computer science and technology in the School of Computer Science and Software, Nanjing University of Information, Science and Technology, China. Her research interests include networking, information security, and dependable systems.

Zhuangdi Zhu received the bachelor's degree from the College of Elite Education, Nanjing University of Science and Technology, in 2015. She is working toward the PhD degree in the Computer Science Department, Michigan State University. She joined IBM Research China, as a research assistant in 2014. Her research interests include systems, networking, and Internet of Things

Alex X. Liu received the PhD degree in computer science from the University of Texas at Austin, in 2006. He is currently an associate professor with the Department of Computer Science and Engineering, Michigan State University, East Lansing. His research interests focus on networking, security, and dependable systems. He received the IEEE and IFIP William C. Carter Award in 2004 and an NSF CAREER Award in 2009. He received the MSU College of Engineering Withrow Distinguished Scholar Award in 2011.

Xingming Sun received the BS degree in mathematics from Hunan Normal University, China, in 1983, the MS degree in computing science from the Dalian University of Science and Technology, China, in 1988, and the PhD degree in computing science from Fudan University, China, in 2001. He is currently a professor with the Department of Computer and Software, Nanjing University of Information Science and Technology, China. In 2006, he visited the University College London, United Kingdom. He was a visiting professor with the University of Warwick, United Kingdom, between 2008 and 2010. His research interests include network and information security, database security, and natural language processing.