

HandGest: Hierarchical Sensing for Robust-in-the-Air Handwriting Recognition With Commodity WiFi Devices

Jie Zhang^{ID}, Member, IEEE, Yang Li^{ID}, Haoyi Xiong, Senior Member, IEEE, Dejing Dou^{ID}, Senior Member, IEEE, Chunyan Miao^{ID}, Senior Member, IEEE, and Daqing Zhang^{ID}, Fellow, IEEE

Abstract—Recent advances in wireless sensing techniques have made it possible to recognize hand gestures using channel state information (CSI) in commodity WiFi devices. Existing WiFi-based gesture recognition systems mainly use learning-based pattern recognition methods to recognize different gestures, however, these methods fail to work well when the locations of transceivers, the relative location and orientation of the hand with respect to transceivers, and/or the hand gesturing size change, leading to inconsistent signal patterns caused by those factors. Although some recent efforts have been made to address the so-called “domain-dependent” gesture recognition problem, they either require prior knowledge on initial locations of the hand and WiFi devices or need to train several classifiers for the specific domains. Different from the state-of-the-art methods, we construct two distinct features from a hand-oriented view (rather than from a transceiver’s view), namely, the dynamic phase vector (DPV) and motion rotation variable (MRV), which are quite consistent in characterizing a big set of handwriting gestures, despite significant change in locations of transceivers, the relative location and orientation of the hand with respect to transceivers, and the drawing sizes. We further incorporate a hierarchical sensing framework and develop HandGest—a real-time handwriting gesture recognition system using commodity WiFi devices, to precisely recognize a great number of “in-the-air” handwritings based on the aforementioned two

Manuscript received 23 November 2021; revised 14 March 2022; accepted 10 April 2022. Date of publication 25 April 2022; date of current version 23 September 2022. This work was supported in part by the National Natural Science Foundation of China A3 Foresight Program under Grant 62061146001, and in part by PKU-Baidu Fund under Grant 2019BD005. (*Corresponding author: Daging Zhang.*)

Jie Zhang is with the Key Laboratory of High Confidence Software Technologies, Ministry of Education, and the School of Computer Science, Peking University, Beijing 100871, China, and also with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: zhangjie.saee@pku.edu.cn).

Yang Li is with the Key Laboratory of High Confidence Software Technologies, Ministry of Education, and the School of Computer Science, Peking University, Beijing 100871, China (e-mail: liyang@stu.pku.edu.cn).

Haoyi Xiong and Dejing Dou are with the Big Data Laboratory, Baidu Inc., Beijing 100193, China, and also with the National Engineering Laboratory of Deep Learning Technology and Applications, Beijing 100193, China (e-mail: haoyi.xiong.fr@ieee.org; doudejing@baidu.com).

Chunyan Miao is with the School of Computer Science and Engineering and the Joint NTU-UBC Research Center of Excellence in Active Living for the Elderly (LILY), Nanyang Technological University, Singapore 639798 (e-mail: ascymiao@ntu.edu.sg).

Daqing Zhang is with the Key Laboratory of High Confidence Software Technologies, Ministry of Education, and the School of Computer Science, Peking University, Beijing 100871, China, and also with the Telecom SudParis, Institute Polytechnique de Paris, 91011 Évry, France (e-mail: dqzhang@sei.pku.edu.cn).

Digital Object Identifier 10.1109/JIOT.2022.3170157

domain-independent features and a pipeline of specific features. Extensive experiments have been done in practical settings with 20 volunteers, evaluation results demonstrate that HandGest outperforms state-of-the-art methods on a large number of handwritings with different transceivers’ location, different initial hand locations and orientations, as well as different drawing sizes. Given its superior performance, we believe that HandGest paves a new way to enhance the real-world practicality of WiFi-based gesture recognition.

Index Terms—Dynamic phase vector (DPV), handwriting recognition, hierarchical sensing, motion rotation variable (MRV).

I. INTRODUCTION

IN THE past decade, the feasibility of gesture-based human–computer interaction (HCI) has been widely explored by researchers, aiming at developing advanced interactive technologies with intelligent devices [1]–[3]. Different from traditional touch-based interfaces, contactless hand-gesture-based HCI supports the natural interaction through hand drawing in the air without dedicated on-body hardware. Existing state-of-the-art methods can be grouped into the following categories: camera-based, acoustic-based, WiFi-based methods, etc. [4]. Blessed by the ubiquitousness of WiFi signals at low costs, WiFi-based solutions, capturing the variations of WiFi signals caused by hand movements, have attracted growing attention from both academia and industries [5]–[7].

To capture gestures in the space covered by WiFi signals, channel state information (CSI) has been used by existing studies [8]–[10]. Specifically, CSI characterizes the frequency response of the wireless channel between the transceivers at orthogonal frequency-division multiplexing (OFDM) subcarriers, and it would vary greatly along with moving objects in the space [11]–[13]. Given a commodity WiFi device pair, one could retrieve CSI from the network interface card (NIC), such as Intel 5300 NIC, with CSITOOL [14]. Based on raw CSI measurements, most state-of-the-art methods adopt learning-based methods [15]–[17] to first extract features from CSI, and then recognize different gestures through classification, where the primitive features, such as CSI amplitude and statistical feature, have been frequently used to classify gesture with respect to signal variations.

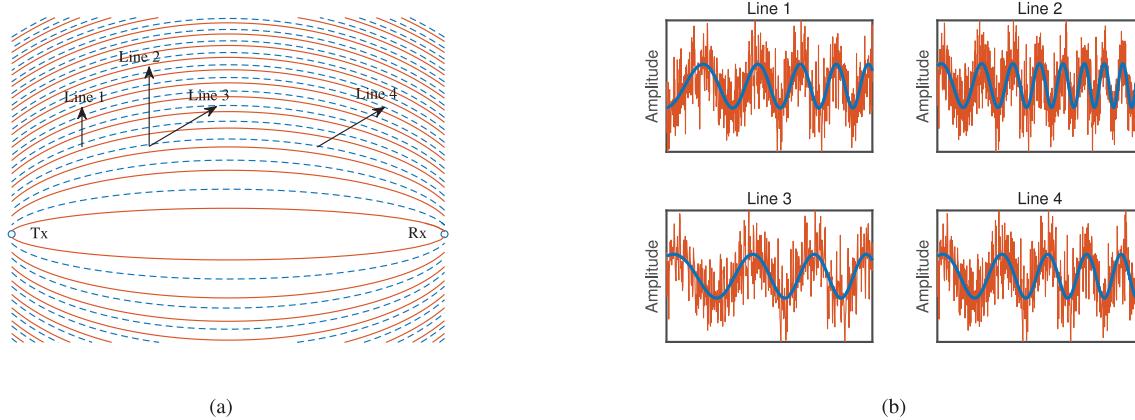


Fig. 1. Effects of initial locations, orientations, and drawing sizes on gesture patterns: (a) drawing lines with different initial locations, directions, and lengths and (b) patterns of different lines.

However, these methods fail to recognize hand gestures robustly in practical settings, as they usually cannot tolerate the change of the relative location and orientation of the hand with respect to transceivers, and the drawing size. Specific constraints or prior knowledge on those factors are required by these methods for hand gesture recognition, as signal patterns of hand gestures are usually inconsistent and they are indistinguishable with primitive features. To show the inconsistency of signal patterns with varying factors, in Fig. 1, we visualize CSI amplitudes of four hand moving straight lines with different initial locations, directions and lengths. As shown in Fig. 1(a), lines 1 and 2 share the common moving direction, but with different initial locations and lengths. As a result, CSI amplitudes of lines 1 and 2 are significantly different. Lines 2 and 3 share the same initial location, but with different moving directions and lengths such that the CSI amplitude of line 2 has more peaks and valleys than line 3 [see Fig. 1(b)]. Moreover, when the hand moves another straight line with different lengths, i.e., line 4, it appears to share similar CSI amplitude with line 1. Apparently, all primitive features extracted from CSI measurements are affected by the above-mentioned three factors. The inconsistency of signal patterns would fail most of the existing WiFi-based gesture/activity recognition methods, which highly rely on the assumption that there are pair-to-pair feature mappings between signal variations and gestures/activities. This is one fundamental problem hindering the existing WiFi-based gesture/activity recognition systems from achieving consistent and high accuracy [18].

To recognize gestures under the change of the aforementioned factors (or, namely, “configurations” in [19] and “domains” in [20]), some “domain adaptation” techniques have been proposed. Specifically, WiAG [19] first leveraged a translation function to generate virtual samples with realistic CSI measurements of hand gestures for other possible configurations. After that, several classifiers were trained using the generated virtual samples for pattern recognition under different configurations. Note that an additional smart phone had been requested to collect data from users, so as to model the translation function. Widar3.0 [20] proposed a domain-independent body-coordinate velocity profile (BVP) to unify

the characterization of gestures in the same type with at least three pairs of WiFi devices, where the locations of the users and WiFi devices are requested for BVP construction. In summary, state-of-the-art methods usually acquire additional information to tackle the domain adaptation issue, which limits their practicability and applicability in real-world application scenarios.

Our Work and Contributions: In this work, we study a specific hand gesture recognition task, namely, “in-the-air” handwriting recognition, aiming at recognizing the digits and letters written by a hand in the air. While the state-of-the-art learning-based solutions either are intolerant with changing factors/domains/configurations, or rely on additional prior knowledge, data annotation, devices and transceiver deployments for hand gesture recognition, we propose to address the domain-dependent issue from a different perspective.

While all existing gesture recognition methods apply the transceiver-oriented view to receive CSI and extract primitive features, it is difficult to recognize the hand gestures robustly as the pattern of the same gesture may vary from the view of WiFi devices due to initial locations and orientations of the hand, and drawing sizes. Differently, we characterize the relative motion of handwriting using a hand-oriented view and construct two distinct features—dynamic phase vector (DPV) and motion rotation variable (MRV)—that provide consistent and distinguishable representations of handwriting gestures in a domain-independent way. Specifically, the use of DPV and MRV accurately categorizes various handwritings into a big number of groups, thus it is quite straightforward to distinguish gestures from different groups accurately. In order to differentiate gestures in the same group, a hierarchical sensing framework is proposed to apply with the fine-grained feature to refine the classification results.

In this way, we construct HandGest that incorporates the two distinct features and a hierarchical sensing framework to recognize the hand gestures robustly. Note that HandGest is just an initial attempt to address the fundamental problem in CSI-based gesture/activity recognition—the location and orientation-dependent problem—in the wireless sensing field, where we use digits and letters recognition as an example.

With the aforementioned problem solved, HandGest provides researchers in the community critical insights in system design while leading a way to move on. The main contributions can be summarized as follows.

- 1) Two hand-centric features, i.e., DPV and MRV, are constructed to convert every handwriting gesture to a unique profile from a hand-oriented view, which is not affected by the initial locations and orientations of the hand, and drawing sizes.
- 2) A hierarchical sensing framework is proposed for robust handwriting recognition in real-world application scenarios, where DPV, MRV, and fine-grained feature are fused in a hierarchical decision-making pathway to ensure the distinguishability of numerous handwriting gestures.
- 3) We design and implement HandGest on commodity WiFi devices and conduct comprehensive experiments in real-world settings. Experimental results of handwriting gestures performed by 20 volunteers with different initial locations, drawing orientations/sizes, and WiFi transceiver settings in three typical indoor environments show that the average recognition accuracy of HandGest outperforms state-of-the-art methods, indicating its robustness and great generalization performance.

The remainder of this article is organized as follows. Section II reviews recent advances of WiFi-based gesture recognition. The preliminaries and backgrounds of gesture sensing with CSI ratio are given in Section III. Section IV presents the methodologies of our work, including the design of DPV and MRV features, and the hierarchical sensing framework. System implementation is shown in Section V. Performance evaluation and further analysis are reported in Section VI, followed by some discussions in Section VII. Finally, conclusions are presented in Section VIII.

II. RELATED WORK

WiFi-based gesture recognition can be grouped into two categories depending on whether the initial locations and orientations of handwritings have been specified or included as the input for recognition.

A. WiFi-Based Gesture Recognition That Depends on the Initial Hand Locations and Orientations

Most of the existing WiFi-based gesture recognition works assume the initial hand locations, orientations and drawing sizes of handwritings are known or specified. They barely perform well with the changes of the initial location and orientation of the hand, and drawing sizes. Specifically, WiFinger [15] adopted multidimensional DTW (MD-DTW) to calculate the similarity between the testing CSI measurements and each of the known gesture profiles, and the one with the highest similarity score was identified as the recognized gesture. QGesture [17] adopted a phase correction algorithm and a novel estimation algorithm to reduce the phase noise of CSI measurements and impacts of environmental dynamics for high accuracy, but the user must perform two

predefined actions to separate gesture from daily activities. WiMU [21] was designed for multiuser gesture recognition, which first generated virtual samples for various plausible combinations of simultaneously performed gestures, and then classified those gestures utilizing the Jaccard similarity-coefficient-based method. Yang *et al.* [22] designed a modified OpenWrt-based platform, in which deep neural networks were adopted to extract the spatiotemporal patterns of gestures. WiDraw [23] extracted Angle-of-Arrival (AoA) values from CSI and average received signal strength (RSS) measurements for hand motion tracking, but it required dense transmitters to guarantee its high accuracy. WiGest [24] defined three unique signal states, including the rising edge, falling edge and pause, and parsed the combinations of those signal states to identify various gestures. While the aforementioned works have made some progresses in recognition, they fail to solve the location/orientation-dependent problem—a fundamental problem hindering the wireless sensing techniques to be applied in real settings.

B. Location/Orientation-Independent WiFi-Based Gesture Recognition

In order to tackle the location/orientation-dependent problem of WiFi-based gesture recognition, some works propose advanced location/orientation-independent feature or/and resort to machine/deep learning methods. WiAG [19] proposed a modified translation function to generate virtual samples of all the gestures in the possible configurations only using the samples of all the gestures in one configuration. Widar3.0 [20] constructed a domain-independent BVP to guarantee each type of gestures has its unique velocity profile in the established body coordinate system, which could achieve satisfactory recognition accuracy on relatively complex gestures. However, it requires to know the locations of user and WiFi devices in advance, and a large number of data for training deep neural network, which limit its practicability and scalability in real-world application scenarios. WiHand [25] extracted location-independent gesture signals from the raw WiFi signals through the low rank and sparse decomposition algorithm, and then utilized support vector machine (SVM) as classifier to identify the gestures. Regani *et al.* [26] proposed a unique hand gesture model and derived the correspondence between the time reversal resonating strength decay and relative distance moved by the hand in the through-the-wall scenario. They achieved 87% recognition accuracy on six uppercase English alphabets, but the hand movement must be straight line during performing gestures in the air. Additionally, Ma *et al.* [27] developed a novel deep neural network for gesture recognition, which could simultaneously learn high-level features and a transferrable similarity evaluation ability, making the learned knowledge transferrable between the training dataset and new testing scenarios. While state-of-the-art methods intend to solve the location/orientation-dependent problem, they actually are with poor robustness and generalization performance, especially when the initial location, orientation, drawing size, etc., are not constrained.

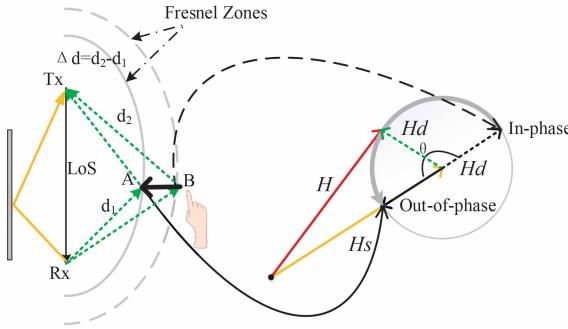


Fig. 2. Moving hand in Fresnel zones.

III. GESTURE SENSING WITH CSI RATIO

Given K subcarriers, denoted as $\mathbf{H}(f_1)$, $\mathbf{H}(f_2), \dots, \mathbf{H}(f_K)$ over the frequencies f_1, f_2, \dots, f_K , CSI aggregates the real-time characterization of the K subcarriers of radio propagation, which can be mathematically represented as

$$\mathbf{H} = [\mathbf{H}(f_1), \mathbf{H}(f_2), \dots, \mathbf{H}(f_K)]. \quad (1)$$

Specifically, a single subcarrier $\mathbf{H}(f_k)$ of CSI can be expressed as

$$\mathbf{H}(f_k) = |\mathbf{H}(f_k)| e^{j\angle \mathbf{H}(f_k)} \quad (2)$$

where $|\mathbf{H}(f_k)|$ denotes the amplitude, and $\angle \mathbf{H}(f_k)$ denotes the corresponding phase, respectively.

In an environment covered by WiFi signals, the radio would propagate over multiple paths, including Line-of-Sight (LoS) path of WiFi transceivers, and reflection paths caused by the surrounding environment, from the transmitter to the receiver. To be specific, when the hand moves in the air, CSI would change dynamically due to the change of the reflection path length, where we can further decompose the CSI into the static component and the dynamic component [28]. The static component remains the same as the subcarriers propagating along with the LoS path and surrounding environment do not change, while the dynamic component varies along with the hand movement. In this way, the decomposition of CSI into static and dynamic components could be modeled as follows:

$$\mathbf{H}(f, t) = \mathbf{H}_s(f, t) + \mathbf{H}_d(f, t) = \mathbf{H}_s(f, t) + A(f, t) e^{-j2\pi \frac{d(t)}{\lambda}} \quad (3)$$

where $\mathbf{H}_s(f, t)$ is the static phasor component, $\mathbf{H}_d(f, t)$ is the dynamic phasor component, $A(f, t)$, $e^{-j2\pi(d(t)/\lambda)}$ and $d(t)$ are the attenuation, phase shift and path length, respectively.

As depicted in Fig. 2, the dynamic phasor component rotates with the change of the reflection path length, caused by the hand movement. In terms of the angle of rotation, when the length of the reflection path changes within a wavelength, the dynamic phasor component rotates on an angle smaller than 2π (i.e., a circle). In terms of the direction of rotation, the rotation direction of the dynamic phasor component also couples with the hand moving direction [29].

As we all know, the signal-to-noise ratio (SNR) of the CSI measurements is usually low due to the hardware and environmental noise, and random offset noise in the CSI phase. It seems difficult to fully capture some small-scale hand movements with the raw CSI signals. Thus, we adopt CSI ratio

to remove the uncertain impulse noise in CSI amplitude and the random offset noise in the CSI phase [30], [31]. The CSI ratio of two antennas can be represented as

$$\begin{aligned} \frac{H_1(f, t)}{H_2(f, t)} &= \frac{e^{-j\theta_{\text{offset}}}(H_{s,1} + A_1 e^{-j2\pi \frac{d_1(t)}{\lambda}})}{e^{-j\theta_{\text{offset}}}(H_{s,2} + A_2 e^{-j2\pi \frac{d_2(t)}{\lambda}})} \\ &= \frac{A_1 e^{-j2\pi \frac{d_1(t)}{\lambda}} + H_{s,1}}{A_2 e^{-j2\pi \frac{d_1(t)+\Delta d}{\lambda}} + H_{s,2}} \\ &= \frac{A_1 e^{-j2\pi \frac{d_1(t)}{\lambda}} + H_{s,1}}{A_2 e^{-j2\pi \frac{\Delta d}{\lambda}} e^{-j2\pi \frac{d_1(t)}{\lambda}} + H_{s,2}} \end{aligned} \quad (4)$$

where $\mathbf{H}_1(f, t)$ and $\mathbf{H}_2(f, t)$ denote the CSI measurements of the first and second antennas, and $\mathbf{H}_{s,1}$ and $\mathbf{H}_{s,2}$ are the corresponding static phasor components, $\Delta d = d_2(t) - d_1(t)$, respectively.

According to (4), we can find that the CSI ratio can cancel most of the random phase offsets, and significantly improve SNR. Thus, the CSI ratio should be very sensitive to hand motions and is capable of fully utilizing the orthogonal information of the CSI amplitude and phase. We conduct an empirical study to verify whether CSI ratio can capture hand movements. We draw a digit “7” in the air [see Fig. 3(a)], CSI ratio illustrated in Fig. 3(b) and (c) keeps rotating in the complex plane with the hand movement, where the length of hand movements corresponds to the overall phase change of CSI ratio (in Fig. 3(b) and (c), CSI ratio starts from cool color and ends at warm color). Moreover, when the hand moves in another direction, the rotation direction of the CSI ratio also changes. For example, in Fig. 3(c), the CSI ratio first rotates counterclockwise when the reflection path becomes shorter, while it rotates clockwise when having the reflection path lengthened. This case study indicates that the CSI ratio can be utilized for sensing hand movements at the scale of signal wavelength.

IV. METHODOLOGY

In this section, we first conduct comprehensive empirical studies to understand the invariance of CSI ratio for handwriting recognition tasks, and then introduce the proposed two hand-centric features, i.e., DPV and MRV, fine-grained feature, and hierarchical sensing framework for robust handwriting recognition, respectively.

A. Understanding CSI Ratio Invariance for Handwriting Recognition

CSI-based handwriting pattern is usually inconsistent, where the existing solutions based on primitive features fail to recognize the same handwriting when the initial hand locations, the orientations, and the sizes of handwriting change. In this section, we aim at uncovering the invariant patterns of CSI ratio for consistent handwriting recognition under the varying conditions.

To understand the inconsistency in handwriting recognition, we draw digit “3” twice [see Fig. 4(a) and (c)]. Apparently, their corresponding CSI amplitudes illustrated in Fig. 4(b)

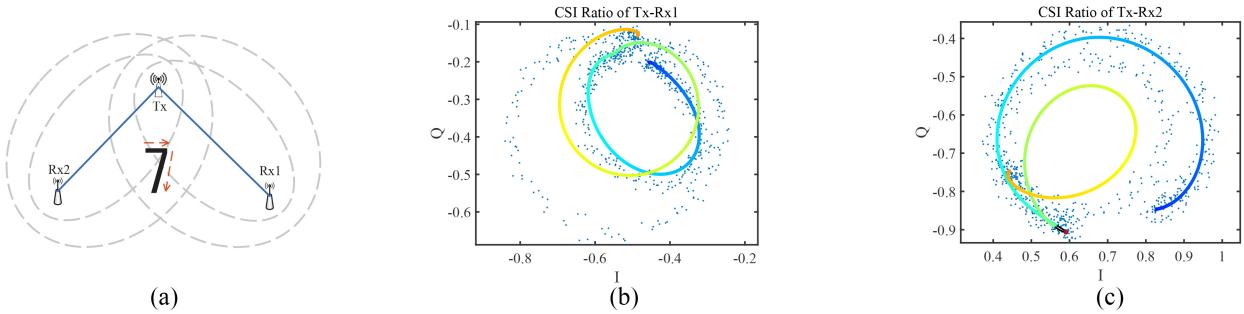


Fig. 3. Empirical study of CSI ratio for sensing hand movement: (a) drawing digit “7” in the air, (b) CSI ratio of Tx–Rx1, and (c) CSI ratio of Tx–Rx2.

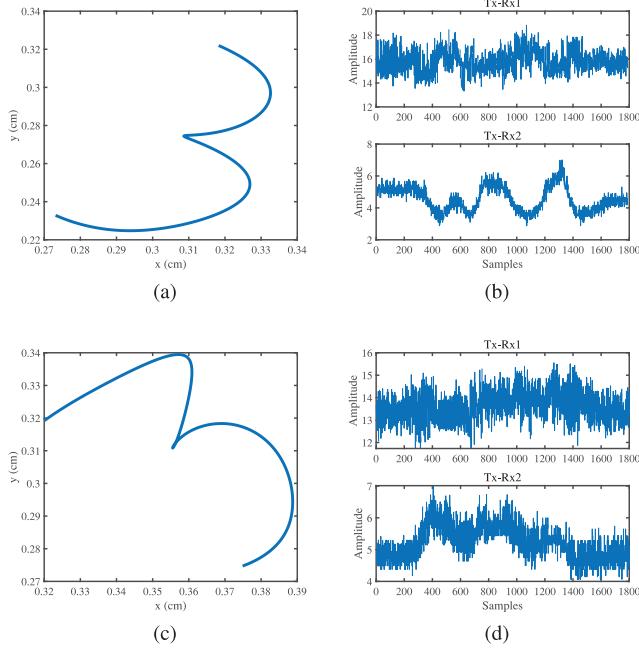


Fig. 4. Empirical study of drawing digit “3”: (a) drawing digit “3” for the first time, (b) CSI amplitude of the first digit “3”, (c) drawing digit “3” for the second time, and (d) CSI amplitude of the second digit “3”.

and (d) are significantly different. Actually, this is a common phenomenon, as the hand movements to draw digit 3 are slightly different every time. CSI amplitude could be considered as a kind of primitive feature, which characterizes hand movements from a transceiver-oriented view, highly depending on the initial locations and orientations of the hand, and drawing sizes, easily leading to inconsistency in recognition. Oppositely, when we characterize hand movements from a hand-oriented view, recording every instantaneous status of hand movement at each time step, we can find that the hand first rotates clockwise, then turns around, and latter rotates clockwise. This pattern exists consistently when the hand writes the digit “3” [see Fig. 5], no matter how initial locations of the hand and the orientations/sizes of handwriting change.

Remark 1: Without loss of generality, we take digits “0–9” as an example to illustrate the proposed hand-centric features, fine-grained feature, and hierarchical sensing framework in the following sections. We use these ten digits as examples for handwriting recognition and uncover every digit with a unique and consistent pattern from a hand-oriented view.

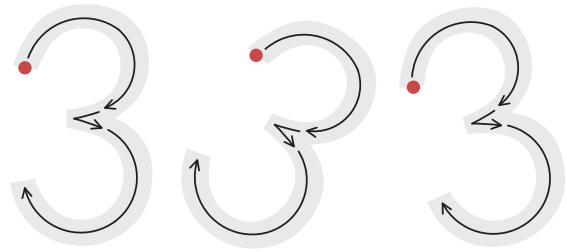


Fig. 5. Hand moving directions of drawing digit “3” in different manners.

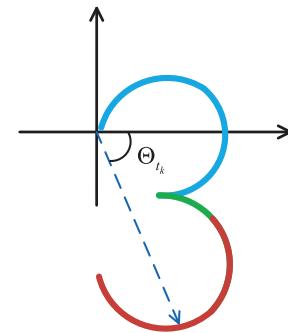


Fig. 6. Illustration of DPV feature.

Remark 2: We consider the moving hand as a mass point in the space. When we handwrite digits in front of WiFi transceivers, CSI signals are reflected by several parts of human body, such as hand, arm and even shoulder. Furthermore, the CSI signals of more than two times reflection are usually weak, it is reasonable to ignore them. In this way, we only consider the CSI signals that directly propagate to the moving hand and then are reflected by the hand to the WiFi receivers.

B. Hand-Centric Feature—Dynamic Phase Vector

Practically, when performing the same handwriting, it is impossible to request users to keep the same initial locations/orientations of the hand, and sizes of handwriting every time. From a transceiver-oriented view, the change of such factors makes the representation of hand movement different. Therefore, it is worth constructing a hand-centric feature that can characterize the hand movement in a hand-oriented view.

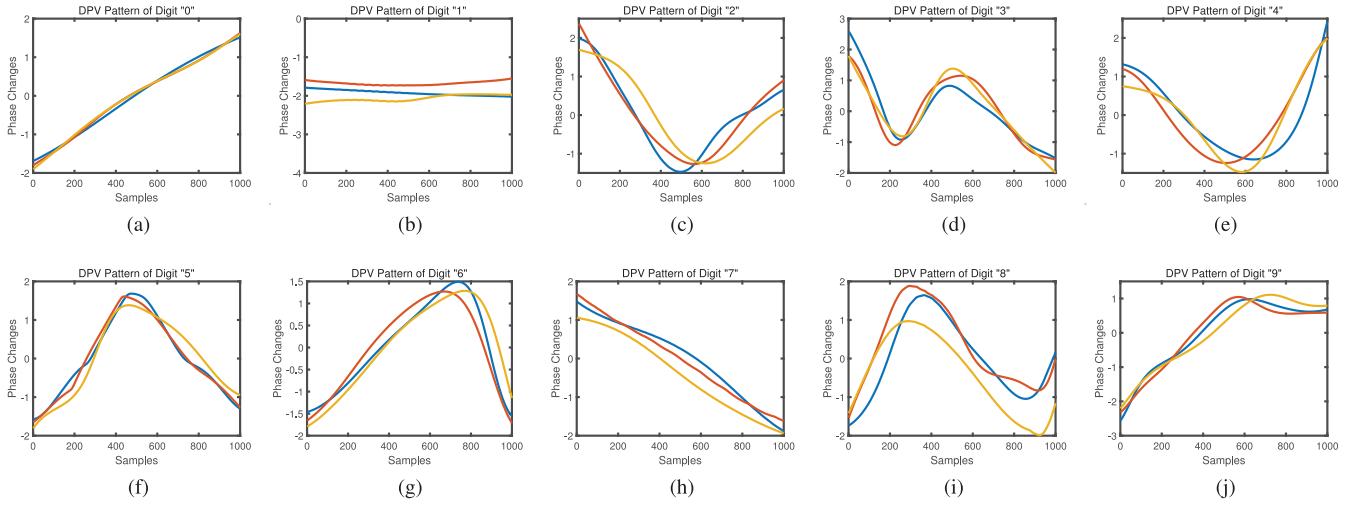


Fig. 7. Pattern templates of DPV of the ten digits. (a) Digit “0.” (b) Digit “1.” (c) Digit “2.” (d) Digit “3.” (e) Digit “4.” (f) Digit “5.” (g) Digit “6.” (h) Digit “7.” (i) Digit “8.” (j) Digit “9.”

As shown in Fig. 6, we establish a coordinate system with initial location of the hand as the origin. We define a modified hand-centric feature, named DPV, as the phase variation between the initial location of the hand and the horizontal axis of each point on the trajectory. Mathematically, dynamic phase sequence can be represented as

$$\Theta = \{\Theta_{t_1}, \Theta_{t_2}, \dots, \Theta_{t_n}\} \quad (5)$$

where Θ_{t_k} stands for the dynamic phase at time step t_k , $k \in [1, n]$.

Note that with the hand movement in ideal settings, CSI of every subcarrier should vary consistently with the same pace while the reflection path caused by the moving hand could be measured by any of them. However, due to multipath propagation, random noise, etc., such consistency may not always exist during the whole hand moving process in practical scenarios. Based on the observation, we find that the subcarrier, whose phase change difference has the lowest variance, can characterize the hand movement better than others in a sliding window. Therefore, to refine DPV, we select the subcarrier with the lowest variance in each sliding window

$$\Upsilon_{t_k, Q} = \operatorname{argmin}_{q \in [1, K]} \operatorname{var}(\Upsilon_{t_k, q}) \quad (6)$$

where $\Upsilon_{t_k, q}$ denotes the dynamic phase of the specific q th subcarrier at time step t_k , $q \in [1, K]$, and $\Upsilon_{t_k, i}$ is the dynamic phase of the i th subcarrier, respectively.

Fig. 7 illustrates the pattern templates of DPV of the ten digits in three independent experiments of different settings, where every pattern template refers to the phase of DPV changing over time when handwriting the digit in a common way. Apparently, DPV is capable of preserving characteristics of all digits through capturing the hand movement in the hand-oriented view. Though these digits are not fully distinguishable using the DPV features, DPV is still able to categorize them into several subgroups with consistent patterns. For example, digits “0,” “1,” “3,” “7,” and “8” have their own unique patterns, but digits “2” and “4” have similar DPV patterns, digits “5,” “6,” and “9” have similar DPV patterns, thus digits “2”

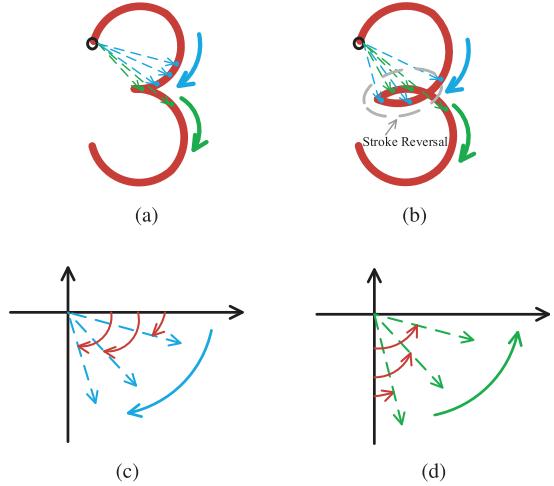


Fig. 8. Theoretical analysis of representation capability of DPV for the hand movement. (a) Drawing digit “3” without stroke reversal. (b) Drawing digit “3” with stroke reversal. (c) DPV rotates clockwise. (d) DPV rotates counterclockwise.

and “4,” digits “5,” “6,” and “9” should be two independent subgroups.

We still take digit 3 as an example to demonstrate the representation capability of DPV for the hand movement. As illustrated in Fig. 8(a), when we draw digit 3 without stroke reversal, its DPV first rotates clockwise and then counterclockwise [see Fig. 8(c) and (d)]. When we draw digit 3 with stroke reversal, the corresponding DPV also first rotates clockwise and then counterclockwise [see Fig. 8(b)–(d)]. According to Fig. 6, digit 3 consists three components in DPV feature space, sequentially denoted by blue, green and red, corresponding to the three-segment curves in Fig. 7(d). DPV can be regarded as a kind of coarse-grained hand-centric feature, which has excellent representation capability for the hand movement and is not sensitive to the local characteristics. Thus, it can efficiently reduce the effects of different handwriting styles on the same digit. Fig. 9 illustrates a practical case of drawing digit 3 without and with stroke reversal, we can find that the two corresponding DPV patterns are consistent, verifying the

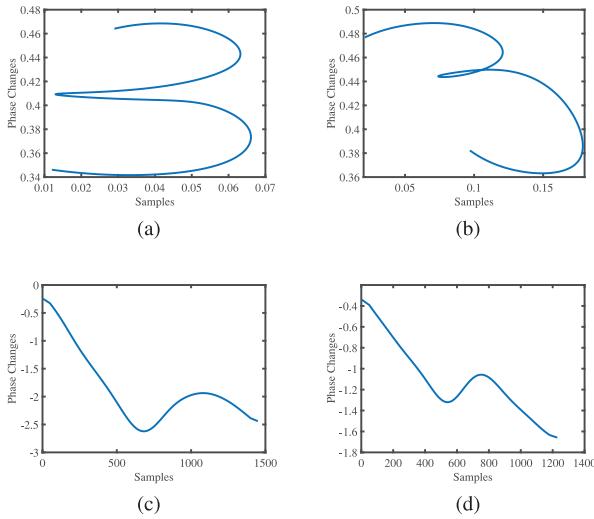


Fig. 9. Empirical study of drawing digit “3” without and with stroke reversal: (a) digit “3” without stroke reversal, (b) digit “3” with stroke reversal, (c) DPV pattern of digit “3” without stroke reversal, and (d) DPV pattern of digit “3” with stroke reversal.

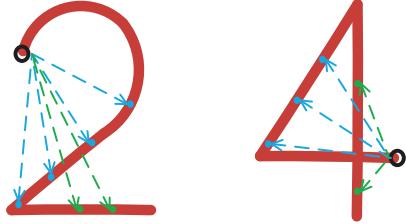


Fig. 10. Similarity between digits “2” and “4” in DPV feature space.

above analysis about the representation capability of DPV for the hand movement.

C. Hand-Centric Feature—Motion Rotation Variable

In this section, we introduce another hand-centric feature, named MRV. As was discussed, DPV represents digits with consistent patterns, however, it also fails to provide distinguishable characterizations to the digits “2” and “4,” and digits “5,” “6” and “9.” For example, in Fig. 10, the DPVs of both digits 2 and 4 first rotate clockwise and then counterclockwise. It means that the essential difference between digits 2 and 4 is the different segment sequence, but DPV cannot characterize these detailed segments (including straight line, arc, etc.) well.

To tackle the problem, we propose MRV, as the angle variation of adjacent sampling points in a certain period of time [see Fig. 11]

$$\Delta\theta = \{\Delta\theta_{t_1}, \Delta\theta_{t_2}, \dots, \Delta\theta_{t_n}\} \quad (7)$$

where $\Delta\theta_{t_k}$ is the angle variation at time step t_k (θ_{t_k} and $\theta_{t_{k-1}}$ represent the instantaneous angles at time step t_k and t_{k-1})

$$\Delta\theta_{t_k} = \theta_{t_k} - \theta_{t_{k-1}}. \quad (8)$$

According to (8), $\Delta\theta_{t_k}$ can be either positive or negative depending on the hand moving direction. Specifically, if hand rotates clockwise, $\Delta\theta_{t_k} > 0$, and if hand rotates counterclockwise, $\Delta\theta_{t_k} < 0$. Thus, MRV also can be utilized to show the hand moving direction.

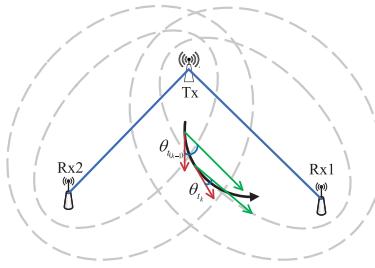


Fig. 11. Illustration of MRV feature.

Fig. 12 illustrates the pattern templates of MRV of the ten digits in three independent experiments of different settings, where every pattern template refers to the angle of MRV changing over time when handwriting the digit. In Section IV-B, we have demonstrated that DPV cannot separate digits 2 and 4, and they can be identified using MRV with distinguished patterns. Similarly, digit 6 has different MRV pattern compared with digits 5 and 9. Accordingly, the ten digits can be furthered classified with the help of MRV. Unfortunately, patterns of some digits are still similar, motivating us to construct targeted fine-grained feature to try to separate them.

D. Fine-Grained Feature

In this section, we introduce fine-grained feature designed for the indistinguishable digits by DPV and MRV. As mentioned above, two digits are still indistinguishable, but they could be divided into one subgroup. Thus, we can construct case-by-case fine-grained feature depending on the characteristics of the digits in the subgroup to finally separate all of them. The constructed fine-grained feature is actually the qualitative representation of primitive features, which is enough for the rest of unrecognized digits in the subgroup.

According to Figs. 7 and 12, digits 5 and 9 have similar DPV and MPV patterns. Experimentally, the velocity directions of the last segment of digits 5 and 9 are significantly different relative to link Tx-Rx1. The velocity direction of the last segment of digit 5 is always positive, and the corresponding one of digit 9 is negative. Thus, the velocity direction of the last segment can be a potential fine-grained feature for separating digits 5 and 9. Practically, we randomly select some digits 5 and 9 to verify this observation, which are demonstrated in Fig. 13. According to Fig. 13, we can find that the last segment of velocity directions of the selected digit 5 are always positive, and the corresponding ones of the selected digit 9 are always negative, indicating that velocity direction is enough for only recognizing digits 5 and 9. The velocity direction of the last segment of handwriting is only a kind of case-by-case fine-grained feature proposed in this article. If more gestures are supposed to be recognized using the proposed method in a location/orientation-independent manner, more features need to be constructed following the idea.

E. Hierarchical Sensing Framework

With the above three types of features, in this section, we present a hierarchical sensing framework for the handwriting

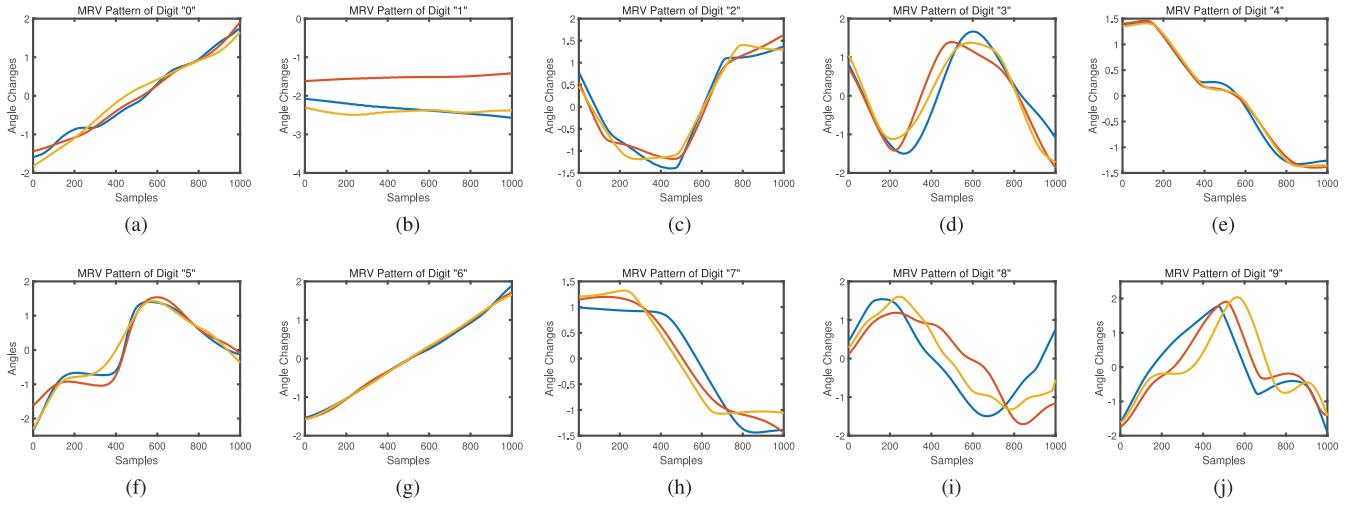


Fig. 12. Pattern templates of MRV of the ten digits. (a) Digit "0." (b) Digit "1." (c) Digit "2." (d) Digit "3." (e) Digit "4." (f) Digit "5." (g) Digit "6." (h) Digit "7." (i) Digit "8." (j) Digit "9."

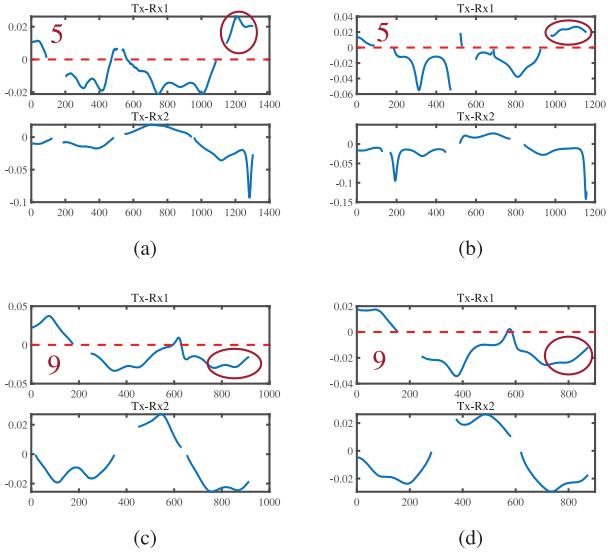


Fig. 13. Empirical study of identifying digits "5" and "9" using targeted fine-grained feature. Fine-grained feature of the (a) first digit "5", (b) second digit "5", (c) first digit "9", and (d) second digit "9".

recognition. The objective of the hierarchical sensing framework is to enhance the discriminant capability of the features for pattern recognition, while the design of the framework is motivated by the following observations.

- 1) Compared to traditional primitive features, the proposed hand-centric features, i.e., DPV and MRV, play important roles in maintaining pattern consistency of digits, which make the digit patterns not change significantly due to the difference of initial location and orientation of the hand, drawing size, etc.
- 2) Each or several digits have their own common characteristics. A certain feature, including both traditional primitive feature and hand-centric feature, may have good representation capability only for specific digits, but relatively poor representation capability for other digits. Thus, it is impossible to completely separate all the digits only utilizing any of the feature. For example,

DPV cannot identify digits 2 and 4, and MRV does not function well for digits 0 and 6.

- 3) Every feature has different representation capability for different digits, if all the features are utilized equally, the features with poor representation capability for specific digits may provide redundant information, leading to performance loss.

In this way, we design the hierarchical sensing framework as illustrated in Fig. 14. Accordingly, all the digits are divided into several subgroups by DPV in the first layer, named *first coarse-grained classification*. In the next layer, MRV-based *second coarse-grained classification* is implemented for the grouped digits. After that, unrecognized digits with similar MRV patterns have been categorized into smaller subgroup. In the final layer, targeted finer-grained feature is designed for the subgroup to separate the digits completely, i.e., *fine-grained classification*. Fig. 15 illustrates the hierarchical sensing framework for English letters, which also employs a three-layer structure with DPV-based first coarse-grained classification, MRV-based second coarse-grained classification, and fine-grained classification.

V. SYSTEM IMPLEMENTATION

In this section, we detail the design and implementation of HandGest, which consists of data collection, data preprocessing, feature construction, and hierarchical sensing components. Fig. 16 depicts its overview architecture, in which data collection component collects raw CSI measurements from the WiFi receivers and divides them to produce CSI ratio streams, data preprocessing component is mainly for denoising CSI ratio data through Savitzky–Golay (SG) filter, feature construction component constructs DPV, MRV, and fine-grained feature based on the denoised CSI ratio, and hierarchical sensing component identifies handwriting gestures.

A. Data Collection

In this component, we collect CSI measurements from the two WiFi receivers equipped with Intel 5300 NICs with three

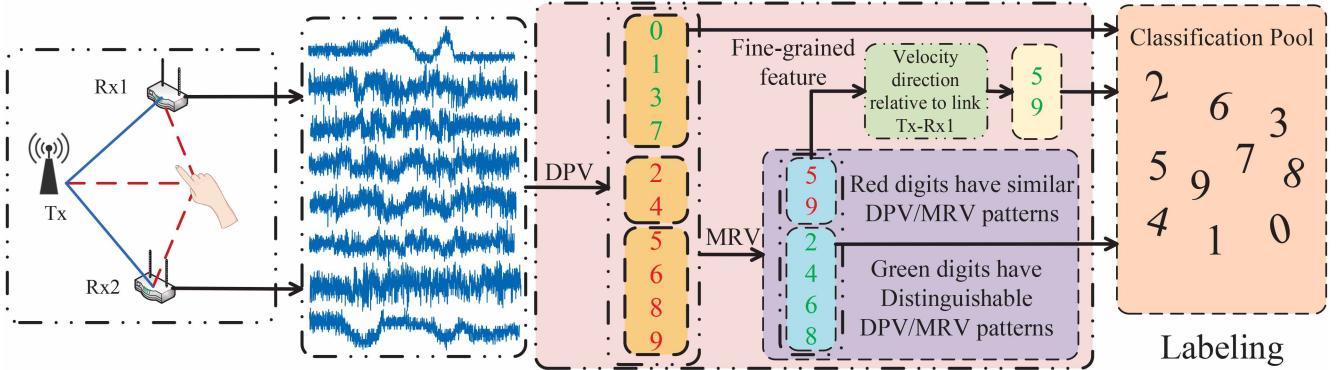


Fig. 14. Hierarchical sensing framework for digits.

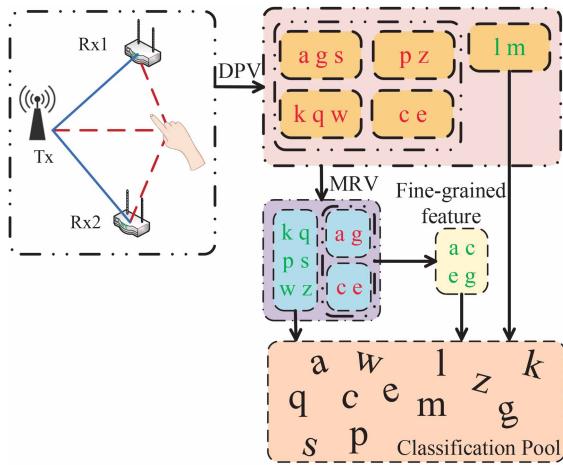


Fig. 15. Hierarchical sensing framework for letters.

antennas for packet receiving. CSI measurements contain 30 subcarriers, each of them is a 3×30 complex matrix. We send all the CSI data to a Dell XPS 7590 laptop via TCP/IP socket. Finally, we divide the two complex CSI measurements from the two antennas at the same WiFi receiver to obtain CSI. The sampling rate is set as 500 Hz for a trade-off between recognition resolution and time consumption of HandGest.

B. Data Preprocessing

Compared with raw CSI, the SNR of CSI ratio is higher, but it still needs to be smoothed for revealing the circular shape of CSI ratio clearly for phase change extraction in the complex plane. Considering that SG filter can fit successive subset of data points with a low-degree polynomial and little distortion through linear least-square manner, we adopt it to denoise and smooth the CSI ratio data [32], [33].

C. Feature Construction

In this component, we first construct DPV by calculating the phase variation between the starting point of the hand movement and the horizontal axis of each point on the trajectory at each time step, establishing a dynamic phase sequence.

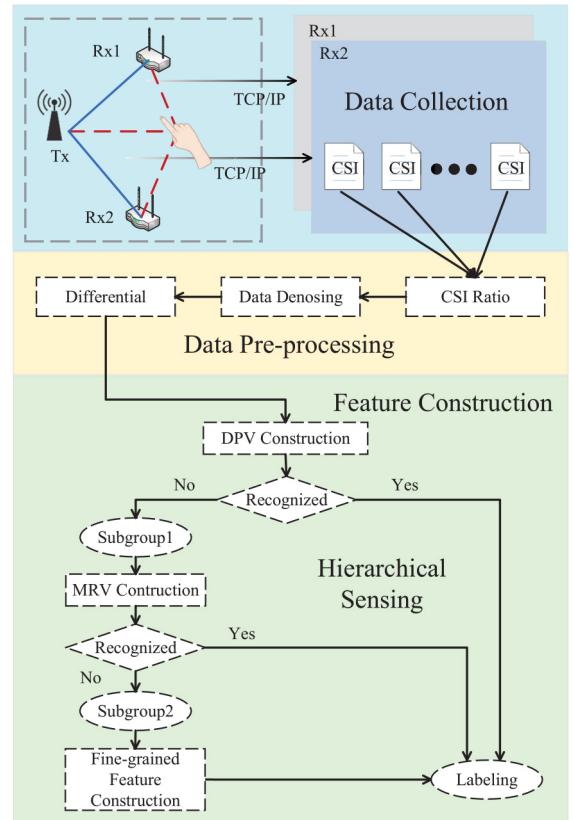


Fig. 16. Overview architecture of HandGest.

As mentioned above, HandGest does not have specific constraints on initial locations of the hand, and it cannot directly calculate the trajectory of hand movement, thus we construct DPV by

$$\Theta_{t_k} = \arctan\left(\frac{\Phi_{t_k}^1}{\Phi_{t_k}^2}\right) \quad (9)$$

where $\Phi_{t_k}^1$ and $\Phi_{t_k}^2$ denote the phase of CSI ratio of links Tx-Rx1 and Tx-Rx2.

DPV has good representation capability for characterizing the hand movement, which can reduce the effects of drawing inconsistency of the same handwriting gestures, but it may not identify all the digits only using DPV. Thus, we then calculate

the angle variation of adjacent sampling points in a certain period of time to construct MRV

$$\Delta\Phi_{t_k}^1 = \Phi_{t_k}^1 - \Phi_{t_{k-1}}^1 \quad (10)$$

$$\Delta\Phi_{t_k}^2 = \Phi_{t_k}^2 - \Phi_{t_{k-1}}^2 \quad (11)$$

$$\Delta\theta_{t_k} = \arctan\left(\frac{\Delta\Phi_{t_k}^1}{\Delta\Phi_{t_k}^2}\right). \quad (12)$$

Differently, MRV is sensitive to the local characteristics of hand movement, which is discriminative to the gestures with similar DPV patterns. Finally, the targeted fine-grained feature is proposed to identify the unrecognized gestures by DPV and MRV, such as velocity relative to link Tx–Rx1, direction of displacement relative to link Tx–Rx1, etc.

D. Hierarchical Sensing

In this component, all the constructed features are embedded into the modified hierarchical sensing framework to identify handwriting gestures, enhancing discriminant capability of those features in different layers. Specifically, first coarse-grained classification separates some of the gestures using DPV, in which all the handwriting gestures have their own consistent patterns. In addition, second coarse-grained classification is performed for the handwriting gestures with similar DPV patterns using MRV. Finally, fine-grained classification is implemented for the unrecognized handwriting gestures by DPV and MRV. In the offline preparation phase, to minimize the efforts required for training and calibration, our method only needs to collect the WiFi signals for each handwriting once as the training sample. In the online recognition phase, when an unknown handwriting is drawn and processed, HandGest extracts the two features from the WiFi signals and searches the nearest example as the recognition label using DTW by calculating the Euclidean distance of DPV and MRV features between the unknown handwriting and the template handwritings. Note that no matter the handwritings are performed by whom and in which location/orientation settings, the proposed method is capable of producing a robust model that can accurately recognize the handwritings performed by different persons in varying location/orientation settings.

VI. PERFORMANCE EVALUATION

In this section, we conduct comprehensive experiments to evaluate the performance of HandGest in various scenarios. Specifically, we first introduce experimental settings, including experimental environments, deployment of WiFi devices, participants, etc. Furthermore, we compare HandGest with state-of-the-art methods. Finally, we show experimental results of HandGest with different initial locations of the hand, drawing sizes, drawing speeds, drawing manners, LoS distances of WiFi transceivers, angles between WiFi transceivers, sampling rates, static environments, dynamic environments, etc.

A. Experimental Settings

The experimental environments include an office, a meeting room, and an empty hall of Peking University, layouts of

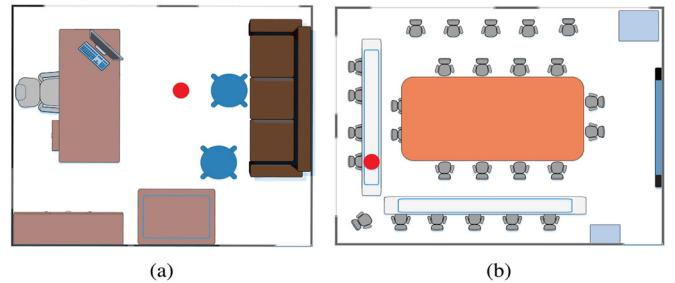


Fig. 17. Layouts of (a) office and (b) meeting room.

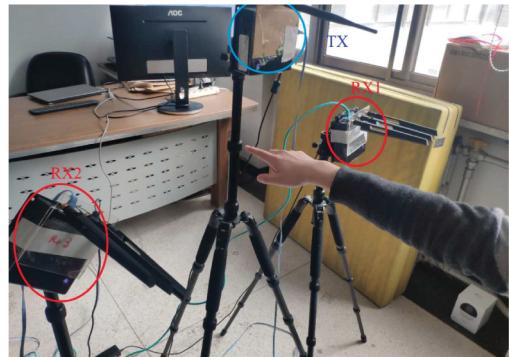


Fig. 18. Experimental setting in the office.

the office and meeting room are illustrated in Fig. 17. The size of the office is 5 m×5 m, the size of the meeting room is 6 m×8 m, and the size of the empty hall is 8m×12m, respectively. Accordingly, both of the office and meeting room are multipath-rich due to their complex layouts. The tested handwriting gestures include selected digits and letters.

In each of the three indoor environments, we build a 2-D Fresnel zones with two pairs of WiFi transceivers, utilizing GigaByte mini PCs with an Intel 5300 NIC and external omnidirectional antennas for data transmitting and receiving, as shown in Fig. 18 (experimental locations of HandGest are denoted by the red points in Fig. 17). HandGest runs in 5.24-GHz frequency with 40 MHz, and the packet transmission rate is set as 500 packets per second, installing CSITOOL on the mini PCs for data collection.

We recruit 20 volunteers to participate our experiments, including ten males and ten females. There are three experienced participants (i.e., #1–#3), 12 of them have no prior knowledge about the proposed methods for CSI-based gesture recognition (i.e., #4–#15), and the rest of them know nothing about WiFi sensing (i.e., #16–#20).

B. Overall Comparisons

In this section, we compare HandGest with selected state-of-the-art CSI-based gesture recognition methods, including WiFinger [15], Widar3.0 [20], FingerDraw [29], and AirDraw [34], and machine/deep learning methods commonly utilized in WiFi sensing, including SVM, convolutional neural network (CNN), long short-term memory (LSTM), and CNN+LSTM. We train SVM, CNN, LSTM, and CNN + LSTM using the data collected from the randomly

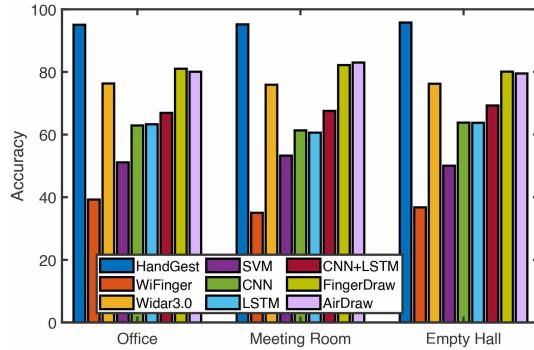


Fig. 19. Comparison results of HandGest and state-of-the-art methods.

selective 15 volunteers, and test them using the data collected from the rest 5 volunteers.

According to Fig. 19, HandGest outperforms all the comparison methods. In addition, the recognition accuracy of Widar3.0 is about 73%, which is higher than WiFinger, SVM, CNN, LSTM, and CNN + LSTM. Because Widar3.0 utilizes a modified BVP to describe the power distribution over different velocities of body parts involved in gesture performing, and train the deep learning model. However, Widar3.0 may not generate enough resolution ratio to well identify centimeter-level gestures. Furthermore, it needs at least three wireless links, and the locations of user and WiFi devices are required, which may be practically inconvenient. HandGest achieves better recognition accuracy than FingerDraw and AirDraw. The main reasons include: 1) both FingerDraw and AirDraw calculate the displacement information of the hand/finger to reconstruct the drawing trajectory, but the accumulated error may lead to poor drawing trajectory and 2) when we draw the digits in 180°, Microsoft Azure OCR used in FingerDraw and AirDraw for classification cannot recognize such kind of digits accurately. Differently, HandGest characterizes the relative motions of hand gestures, thus the accumulated error has no effect on recognition accuracy, and it can achieve better recognition accuracy under different orientations and initial locations of the hand, and drawing sizes. Machine/deep-learning-based methods do not achieve good performance, because they assume that all the gestures have their unique patterns. As mentioned above, pattern consistency highly depends on the initial location of the hand in Fresnel zones and the number of Fresnel zones it moves across. In order to present the recognition accuracy of each individual digit, we select samples collected from different environments and volunteers to construct a confusion matrix of digits (see Fig. 20). Accordingly, the overall average recognition accuracy is around 95%.

C. Evaluation of Robustness and Generalization Performance

In this section, we evaluate the robustness and generalization performance of HandGest under various factors.

1) *Impact of Different Initial Locations of the Hand:* We print templates for all the selected digits and letters on a cardboard, and hang them in the sensing area between the

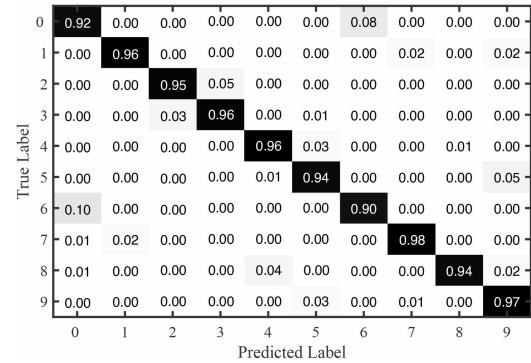


Fig. 20. Confusion matrix of digits.

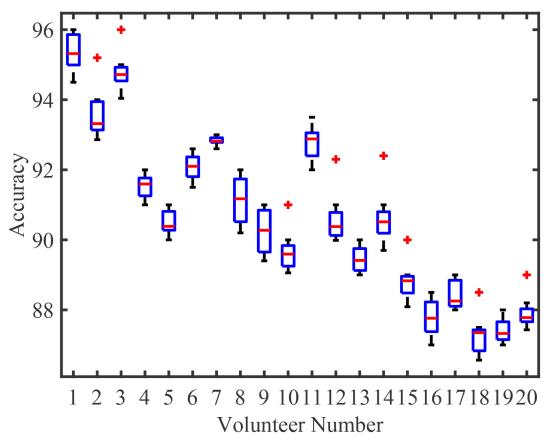


Fig. 21. Comparison results of different participants without restrictions on initial location of the hand.

two pairs of WiFi transceivers, making participants draw at different initial locations by adjusting the locations of those templates. In this manner, there are no special constraints on the initial locations of the hand, all the participants can perform handwriting gesture at any locations naturally between the WiFi transceivers. Fig. 21 depicts the corresponding overall accuracies, which are not affected by the different initial locations of the hand. We can find that #1–#3 achieve the highest accuracies with more than 95%, and the accuracies of others concentrate at 90%, indicating the robustness and great generalization performance of HandGest with both experienced and unexperienced participants.

2) *Impact of Different Drawing Sizes:* We print three templates on a cardboard in three different sizes (5 cm×7 cm for small size handwriting gestures, 7 cm×10 cm for middle size handwriting gestures, and 10 cm×12 cm for large size handwriting gestures) and hang them in front of the volunteers as the ground truth of drawing sizes. We select one volunteer to draw digits and letters in the meeting room, and the corresponding results are listed in Table I. According to Table I, we can find that the overall accuracies of different drawing sizes are comparable only with little fluctuations.

3) *Impact of Different Drawing Speeds:* We ask a volunteer to perform handwriting gestures in three different speeds, including low speed about 3 cm/s, normal speed about 5 cm/s, and high speed about 10 cm/s, the detailed results are listed

TABLE I
COMPARISON RESULTS OF DIFFERENT DRAWING SIZES

	Small size	Middle size	Large size
Accuracy	95.35%	95.79%	92.97%

TABLE II
COMPARISON RESULTS OF DIFFERENT DRAWING SPEEDS

	Low speed	Normal speed	High speed
Accuracy	94.17%	95.10%	90.79%

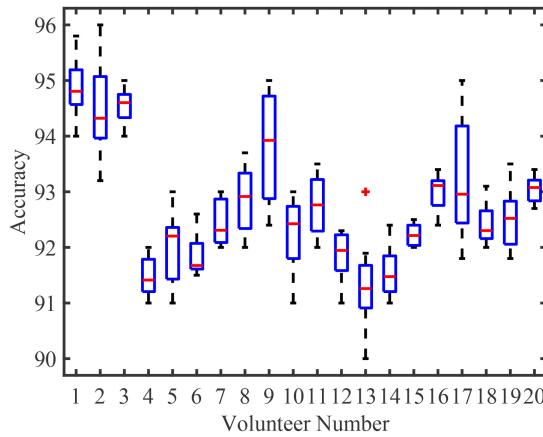


Fig. 22. Comparison results of different volunteers without restrictions on drawing manners.

in Table II. Accordingly, both of the low speed and normal speed achieve better accuracies than high speed. We conjecture that the hand cannot perform the handwriting gestures in a standard manner with high speed, leading to performance degradation.

4) *Impact of Different Drawing Manners:* Similar to the evaluation of different initial locations of the hand, there are still no specific constraints on drawing manners, but with fixed initial locations. As we all know, it is difficult to perform one specific handwriting gesture in the same manner every time. The overall accuracies are depicted in Fig. 22, in which all the volunteers can achieve satisfactory results as long as they follow the standard styles of digits and letters, indicating that the proposed hand-centric features and hierarchical sensing mechanism can reduce or even eliminate the effects of drawing inconsistency on recognition accuracy.

5) *Impact of Different LoS Distances:* We set five different LoS distances of WiFi transceivers, including 60, 75, 100, 125, and 150 cm, and ask a volunteer to perform handwriting gestures in each setting. The comparison results are listed in Table III, we can find that the accuracies under short LoS distances are little better than the one under longer LoS distance. The main reason is that the reflected WiFi signals of the hand under short LoS distance scenarios are stronger than the ones under longer setting, which can reduce the effects of other moving objects in the surroundings.

TABLE III
COMPARISON RESULTS OF DIFFERENT LOS DISTANCES

	50 cm	75 cm	100 cm	125 cm	150 cm
Accuracy	95.41%	95.62%	94.63%	93.50%	91.30%

TABLE IV
COMPARISON RESULTS OF DIFFERENT ANGLES BETWEEN WiFi TRANSCEIVERS

	60 degree	90 degree	120 degree
Accuracy	94.71%	96.03%	95.09%

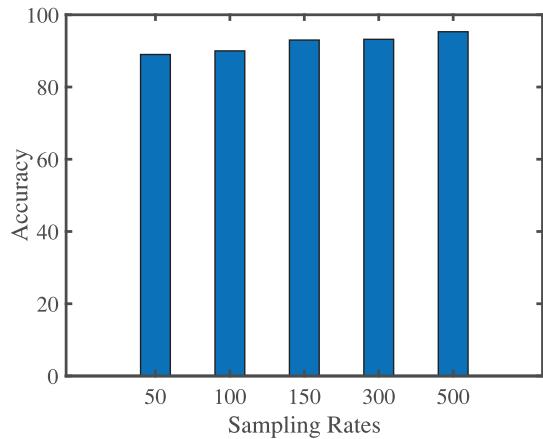


Fig. 23. Comparison results of different sampling rates.

6) *Impact of Different Angles Between WiFi Transceivers:* We set three different angles between WiFi transceivers, including 60°, 90° and 120°. According to the results listed in Table IV, HandGest can achieve similar accuracies with different angles between WiFi transceivers, indicating that HandGest does not need additional efforts on deployment of angles between devices.

7) *Impact of Different Sampling Rates:* Sampling rate is actually important to guarantee the satisfactory accuracy, thus we set five different sampling rates to evaluate its effects, including 50, 100, 150, 300, and 500 Hz. Fig. 23 depicts the corresponding comparison results, when the sampling rate is less than 100 Hz, the accuracies are worse than the ones with higher sampling rates, the corresponding accuracies become stable with small fluctuations, indicating that the lower sampling rate is not enough for capturing hand movements. In practical scenarios, we should do a tradeoff between recognition accuracy and system running time, because high sampling rate may lead to HandGest relatively time-consuming.

8) *Impact of Different Static Environments:* As mentioned above, we implement HandGest in three typical indoor environments, including an office, a meeting room, and an empty hall. The experimental results are detailed in Table V, we can find that there is no obvious difference among them. Because HandGest recognizes handwriting gestures depending on the reflected WiFi signals in the limited sensing area, the changes

TABLE V
COMPARISON RESULTS OF STATIC ENVIRONMENTS

	Office	Meeting room	Empty hall
Accuracy	95.73%	95.21%	96.02%

TABLE VI
COMPARISON RESULTS OF DYNAMIC ENVIRONMENTS

	Office	Meeting room
Door open	93.42%	95.30%
Door closed	95.99%	95.77%
Window open	94.93%	94.06%
Window closed	95.11%	94.63%

TABLE VII
COMPARISON RESULTS OF PEOPLE MOVING AT DIFFERENT DISTANCES

	Office	Meeting room	Empty hall
2m	94.31%	94.90%	95.55%
1.5m	95.20%	95.36%	95.02%
1m	93.32%	92.01%	94.60%
0.5m	90.20%	90.62%	91.93%

of multipaths superimposed in different indoor environments have negligible effects on the recognition accuracy.

9) *Impact of Dynamic Environments*: We conduct experiments in dynamic environments to evaluate the performance of HandGest in dynamic scenarios. Specifically, we open and close door and window while the user performs handwriting in front of WiFi transceivers. The experimental results are listed in Table VI, which indicate the performance of HandGest has not been affected seriously. In addition, we ask people to move around the WiFi transceivers. According to Table VII, HandGest still could achieve good performance when people move far from 0.5 m away. When people are too close to the WiFi devices, we cannot achieve the satisfactory DPV and MRV patterns, because the movement of the surrounding people significantly affects the received signals.

VII. DISCUSSION

In this section, we discuss several open issues of this work, including the flexibility of the proposed hierarchical sensing framework, interpretability of feature extraction for WiFi sensing, additional benefits of the proposed hierarchical sensing mechanism, and limitations of HandGest.

A. Flexibility of the Hierarchical Sensing Framework

We demonstrate the proposed hierarchical sensing frameworks for recognizing digits and letters in Figs. 14 and 15, respectively. In the three-layer framework, DPV-based first

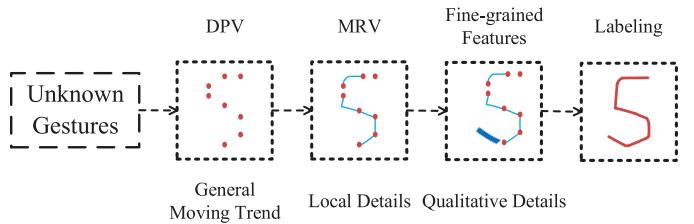


Fig. 24. Interpretable feature extraction of handwriting gesture recognition.

coarse-grained classification and MRV-based second coarse-grained classification are performed to ensure the digits and letters have the unique profiles, and targeted fine-grained features are constructed in the third layer for specific digits and letters, making the distance among handwriting gestures with similar DPV or MRV patterns smaller and the distance among handwriting gestures with different DPV or MRV patterns larger. The above experiments show that such a kind of hierarchical sensing framework can recognize digits and letters with robustness and great generalization performance. It should be noted that all the modules in this framework can be withdrawn and adjusted, and new modules and more modified features can be embedded depending on recognition granularity and type of handwriting, indicating its good flexibility.

B. Interpretability of Feature Extraction for WiFi Sensing

Data-based methods attract much more attention in WiFi sensing field due to the easier implementation, which aim at finding fixed feature mapping between signal variation and activity using machine/deep learning techniques [35]–[39]. In simple scenarios, data-based methods may achieve satisfactory performance when the relationship between signal variation and activity is consistent. It may be difficult to find pair-to-pair feature mappings only using machine learning due to the rich multipath and/or other random interference. In this situation, deep learning is usually adopted to extract more discriminative features through the complicated multilayer structure. However, when two different digits have similar patterns, deep learning may fail to extract discriminative features, such as digits 5, 6, and 9 have similar DPV patterns [40]. In addition, deep learning suffers from time-consuming training process due to fine-tuning of a large number of parameters, and this complication makes it difficult to theoretically understand and interpret the feature extraction process [41], [42]. Thus, the “black-box” deep learning makes people lose understanding the essence of WiFi sensing. Differently, the proposed hierarchical sensing mechanism enables feature extraction of WiFi sensing interpretable and visualized, which is very useful for WiFi sensing, as shown in Fig. 24. In the first layer, DPV represents the general movement of the hand, it may be some instantaneous points at each time step, describing the outline of specific gesture. In the second layer, MRV has good representation capability for the local characteristics of specific handwriting gesture, making its texture clearer. Finally, fine-grained feature can describe qualitative details well, we can recognize the profile of this handwriting gesture.

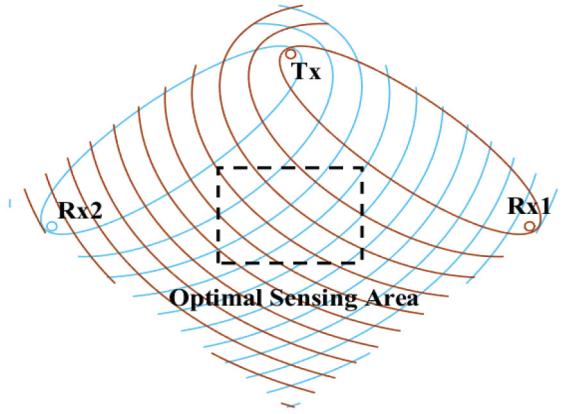


Fig. 25. Intersection of Fresnel zones of the WiFi transceivers.

C. Additional Benefits of the Hierarchical Sensing Mechanism

In this article, we propose the hierarchical sensing mechanism for handwriting recognition, but we envision that it should be a general methodology to guarantee the robustness and generalization performance for many WiFi sensing applications. In addition, we also can identify handwriting gestures with similar trajectories using the hand-centric features embedded in the hierarchical structure. For example, we can draw digit “0” tall and slender, and letter “o” short and fat. Some existing methods directly utilize quantitative features, such as velocity and accumulated displacement, making them achieve satisfactory accuracy only when the hand moves in specific areas of the two pairs of WiFi transceivers. Because the intersected Fresnel zones of the two pairs of WiFi transceivers are orthogonal at specific areas between them (illustrated as optimal sensing area in Fig. 25), and it is sparse and dense alternately in other areas. In the optimal sensing area, quantitative relationship of the features can be held. However, the performance of the proposed hierarchical sensing mechanism is not limited by the so-called “optimal sensing area” due to the embedded hand-centric features recording instantaneous status of the moving hand at each time step from a hand-oriented view.

D. Limitations

HandGest focuses on handwriting recognition of one hand, it is difficult to separate the WiFi signals reflected by multiple moving hands simultaneously using two pairs of commodity WiFi devices. It is possible to capture multiple moving hands by increasing additional WiFi devices to construct 3-D Fresnel zones in a multiview manner. Moreover, we treat the moving hand as a mass point due to the weak reflection of other parts of human body. However, it should be multiple points reflection when our body shakes vigorously during performing handwritings, which is a challenging issue. While the proposed hierarchical sensing framework is extensible, but more fine-grained features are needed to be identified if additional types of handwritings are required for recognition.

VIII. CONCLUSION

In this article, we designed and implemented HandGest, the real-time handwriting recognition system using commodity WiFi devices. Specifically, we first constructed two hand-centric features, i.e., DPV and MRV, to ensure the consistency and discriminability of handwriting, in the feature spaces, without strict constraints of initial locations and orientations of the hand, drawing sizes, etc. Furthermore, we proposed a hierarchical sensing framework that only needs two pairs of commodity WiFi devices and achieves good robustness and great generalization performance in practical scenarios through a layer-by-layer recognition procedure. Comprehensive experiments on a large number of handwritings performed by 20 volunteers in different real-world settings demonstrated that HandGest outperforms state-of-the-art solutions without knowing or constraining the initial locations and orientations of the hand, drawing sizes, etc. The main objective of this work was to address the fundamental problem in CSI-based gesture/activity recognition—the location and orientation-dependent problem. The proposed HandGest is just an initial attempt to solve this problem, where we used the ten-digit recognition as an example. We envision that the proposed hierarchical sensing mechanism is a general methodology for both gesture recognition and other CSI-based sensing applications, which may fill in the gap between laboratory prototype systems and real-world deployments.

REFERENCES

- [1] D. Zhang, H. Wang, and D. Wu, “Toward centimeter-scale human activity sensing with Wi-Fi signals,” *Computer*, vol. 50, vol. 1, pp. 48–57, Jan. 2017.
- [2] Y. Ma, G. Zhou, and S. Wang, “WiFi sensing with channel state information: A survey,” *ACM Comput. Surveys*, vol. 52, no. 3, pp. 1–36, 2019.
- [3] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, “Wireless sensing for human activity: A survey,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, 3rd Quart., 2020.
- [4] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, “SignFi: Sign language recognition using WiFi,” *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–21, 2018.
- [5] X. Wang, X. Wang, and S. Mao, “RF sensing with the Internet of Things: A general deep learning framework,” *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 62–67, Sep. 2018.
- [6] J. Zhang, Y. Li, and W. Xiao, “Integrated multiple kernel learning for device-free localization in cluttered environments using spatiotemporal information,” *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4749–4761, Mar. 2021.
- [7] 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 gestures,” in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2014, pp. 617–628.
- [8] W. Wang, A. X. Liu, M. Shahzad, K. Li, and S. Lu, “Understanding and modeling of WiFi signal based human activity recognition,” in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2015, pp. 65–76.
- [9] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, “Keystroke recognition using WiFi signals,” in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, London, U.K., 2015, pp. 1–13.
- [10] Y. Zou, W. Liu, K. Wu, and L. M. Ni, “Wi-Fi radar: Recognizing human behavior with commodity Wi-Fi,” *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 105–111, Oct. 2017.
- [11] J. Zhang, W. Xiao, and Y. Li, “Data and knowledge twin driven integration for large-scale device-free localization,” *IEEE Internet Things J.*, vol. 8, no. 1, pp. 320–331, Jan. 2021.
- [12] Z. Wang, B. Guo, Z. Yu, and X. Zhou, “Wi-Fi CSI-based behavior recognition: From signals and actions to activities,” *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 109–115, May 2018.

- [13] F. Wang, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "Respiration tracking for people counting and recognition," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5233–5245, Jun. 2020.
- [14] 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, p. 53, 2011.
- [15] S. Tang and J. Yang, "WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, Paderborn, Germany, 2016, pp. 201–210.
- [16] O. Zhang and K. Srinivasan, "Mudra: User-friendly fine-grained gesture recognition using WiFi signals," in *Proc. 12th Int. Conf. Emerg. Netw. Exp. Technol. (CoNEXT)*, Irvine, CA, USA, 2016, pp. 83–96.
- [17] N. Yu, W. Wang, A. X. Liu, and L. Kong, "QGesture: Quantifying gesture distance and direction with WiFi signals," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–22, 2018.
- [18] Y. Zeng, D. Wu, J. Xiong, and D. Zhang, "Boosting WiFi sensing performance via CSI ratio," *IEEE Pervasive Comput.*, vol. 20, no. 1, pp. 62–70, Jan.–Mar. 2021.
- [19] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th ACM Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, Niagara, NY, USA, 2017, pp. 252–264.
- [20] Y. Zheng *et al.*, "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th ACM Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, Seoul, South Korea, 2019, pp. 313–325.
- [21] R. H. Venkatnarayanan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proc. 16th ACM Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, Munich, Germany, 2018, pp. 401–413.
- [22] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from WiFi: A siamese recurrent convolutional architecture," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10763–10772, Dec. 2019.
- [23] 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. (MobiCom)*, Paris, France, 2015, pp. 77–89.
- [24] H. Abdelnasser, M. Youssef, and K.A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Hong Kong, 2015, pp. 1472–1480.
- [25] Y. Lu, S. Lv, and X. Wang, "Towards location independent gesture recognition with commodity WiFi devices," *Electronics*, vol. 8, no. 9, pp. 1069–1088, 2019.
- [26] S. D. Regani, B. Wang, and K. J. R. Liu, "WiFi-based device-free gesture recognition through-the-wall," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Toronto, ON, Canada, 2021, pp. 8017–8021.
- [27] X. Ma, Z. Zhao, L. Zhang, Q. Gao, M. Pan, and J. Wang, "Practical device-free gesture recognition using WiFi signals based on metalearning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 228–237, Jan. 2020.
- [28] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2329–2342, Nov. 2015.
- [29] D. Wu *et al.*, "FingerDraw: Sub-wavelength level finger motion tracking with WiFi signals," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–27, 2020.
- [30] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, 2019.
- [31] T. Needham, *Visual Complex Analysis*. Oxford, U.K.: Oxford Univ. Press, 1998.
- [32] R. W. Schafer, "What is a Savitzky–Golay filter?" *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.
- [33] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "WiDir: Walking direction estimation using wireless signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, Heidelberg, Germany, 2016, pp. 351–362.
- [34] Z. Han, Z. Lu, X. Wen, J. Zhao, L. Guo, and Y. Liu, "In-air handwriting by passive gesture tracking using commodity WiFi," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2652–2656, Nov. 2020.
- [35] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.
- [36] Y. Yang, J. Cao, X. Liu, and X. Liu, "Door-monitor: Counting in-and-out visitors with COTS WiFi devices," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1704–1717, Mar. 2020.
- [37] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New Delhi, India, 2018, pp. 1–16.
- [38] F. Wang, W. Gong, and J. Liu, "On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2035–2047, Apr. 2019.
- [39] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, "Robust extreme learning machine for modeling with unknown noise," *J. Franklin Inst.*, vol. 357, no. 14, pp. 9885–9908, 2020.
- [40] D. Wu, D. Zhang, C. Xu, H. Wang, and X. Li, "Device-free WiFi human sensing: From pattern-based to model-based approaches," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 91–97, Oct. 2017.
- [41] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, "Non-iterative and fast deep learning: Multilayer extreme learning machines," *J. Franklin Inst.*, vol. 357, no. 13, pp. 8925–8955, 2020.
- [42] T. Zheng, Z. Chen, S. Ding, and J. Luo, "Enhancing RF sensing with deep learning: A layered approach," *IEEE Commun. Mag.*, vol. 59, no. 2, pp. 70–76, Feb. 2021.



Jie Zhang (Member, IEEE) received the Ph.D. degree from the University of Science and Technology Beijing, Beijing, China, in 2019.

He was a Visiting Research Fellow with the University of Leeds, Leeds, U.K., in 2018. From 2019 to 2021, he was a Postdoctoral Research Fellow with Peking University, Beijing. He is currently an Associate Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing and a Marie Skłodowska Curie Research Fellow with the School of Electronic and Electrical Engineering, University of Leeds. His current research interests include pervasive computing, body sensor network, wireless sensor networks, and machine learning.

Dr. Zhang was a recipient of the Marie Skłodowska-Curie Actions Individual Fellowships, European Commission.



Yang Li received the B.E. degree from Northwestern Polytechnical University, Xi'an, China, in 2020. He is currently pursuing the Ph.D. degree with the School of Computer Science, Peking University, Beijing, China.

His research interest includes wireless sensing.



Haoyi Xiong (Senior Member, IEEE) received the Ph.D. degree in computer science from Télécom SudParis, Évry, France, and the Université Pierre et Marie Curie, Paris, France, in 2015.

From 2016 to 2018, he was a Tenure-Track Assistant Professor with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO, USA, and a Postdoctoral Fellow with the University of Virginia, Charlottesville, VA, USA, from 2015 to 2016. He is currently a Principal Architect with the Baidu Inc., Beijing, China, and also a Graduate Faculty Scholar affiliated with the University of Central Florida, Orlando, FL, USA. He has published more than 70 papers in top computer science conferences and journals, including UbiComp, ICML, ICLR, RTSS, KDD, AAAI/IJCAI, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON COMPUTERS, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE INTERNET OF THINGS JOURNAL, and ACM Transactions on Knowledge Discovery From Data. His current research interests include AutoDL and ubiquitous computing.

Dr. Xiong was a co-recipient of the 2020 IEEE TCSC Award for Excellence in Scalable Computing (Early Career Researcher) and the prestigious Science and Technology Advancement Award (First Prize) from the Chinese Institute of Electronics in 2019.



Dejing Dou (Senior Member, IEEE) received the B.E. degree from Tsinghua University, Beijing, China, in 1996, and the Ph.D. degree from Yale University, New Haven, CT, USA, in 2004.

He is the Head of Big Data Laboratory and Business Intelligence Lab, Baidu Research, Beijing. He is also a Full Professor (on leave) with the Computer and Information Science Department, University of Oregon, Eugene, OR, USA, and has led the Advanced Integration and Mining Lab since 2005. He has been the Director of the NSF IUCRC

Center for Big Learning, Gainesville, FL, USA, since 2018. He was a Visiting Associate Professor with Stanford Center for Biomedical Informatics Research, Stanford, CA, USA, from 2012 to 2013. He has published more than 160 research papers, some of which appear in prestigious conferences and journals, such as AAAI, IJCAI, ICML, NeurIPS, ICLR, KDD, ICDM, ACL, EMNLP, CIKM, ISWC, JIIS, and JoDS, with more than 5000 Google Scholar citations. His research areas include artificial intelligence, data mining, data integration, NLP, and health informatics.

Prof. Dou has received over five million PI research grants from NSF and NIH. His DEXA'15 paper received the Best Paper Award. His KDD'07 paper was nominated for the best research paper award. He is on the editorial boards of *Journal on Data Semantics*, *Journal of Intelligent Information Systems*, and *PLoS One*. He has been serving as a program committee member for major international conferences and as a program co-chair for five of them. He is a Senior Member of AAAI and ACM.



Daqing Zhang (Fellow, IEEE) received the Ph.D. degree from the University of Rome "La Sapienza," Rome, Italy, in 1996.

He is currently a Chair Professor with the School of Computer Science, Peking University, Beijing, China, and Telecom SudParis, Évry, France. He has published over 300 technical papers in leading conferences and journals. His current research interests include context-aware computing, urban computing, mobile computing, big data analytics, and pervasive elderly care.

Prof. Zhang was a recipient of the Ten-Years CoMoRea Impact Paper Award at IEEE PerCom 2013, the Honorable Mention Award at ACM UbiComp 2015 and 2016, the Ten-Years Most Influential Paper Award at IEEE UIC 2019, and the Distinguished Paper Award at ACM UbiComp 2021. He served as the general or program chair for over 17 international conferences, giving keynote talks at more than 20 international conferences. He is an Associate Editor of *IEEE Pervasive Computing*, *ACM Transactions on Intelligent Systems and Technology*, and *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*.



Chunyan Miao (Senior Member, IEEE) received the B.S. degree from Shandong University, Jinan, China, in 1988, and the M.S. and Ph.D. degrees from Nanyang Technological University (NTU), Singapore, in 1998 and 2003, respectively.

She is currently a Professor with the School of Computer Science and Engineering, NTU, and the Director of the Joint NTU-UBC Research Center of Excellence in Active Living for the Elderly (LILY). Her research focuses on infusing intelligent agents into interactive new media (virtual, mixed, mobile, and pervasive media) to create novel experiences and dimensions in game design, interactive narrative, and other real-world agent systems.