

WiFinger: Leveraging Commodity WiFi for Fine-grained Finger Gesture Recognition

Sheng Tan and Jie Yang
Florida State University, Tallahassee, Florida, USA
{tan, jie.yang}@cs.fsu.edu

ABSTRACT

Gesture recognition has become increasingly important in human-computer interaction (HCI) and can support a broad array of emerging applications, such as smart home, virtual reality, and mobile gaming. Traditional approaches usually rely on dedicated sensors that are worn by the user or cameras that require line of sight. In this paper, we present fine-grained finger gesture recognition by using a single commodity WiFi device without requiring user to wear any sensors. Our low-cost system, WiFinger, takes advantages of the fine-grained Channel State Information (CSI) available from commodity WiFi devices and the prevalence of WiFi network infrastructures. It senses and identifies subtle movements of finger gestures by examining the unique patterns exhibited in the detailed CSI. In WiFinger, we devise environmental noise removal mechanism to mitigate the effect of signal dynamic due to the environment changes. Moreover, we propose to capture the intrinsic gesture behavior to deal with individual diversity and gesture inconsistency. Our experimental evaluation in both home and office environments demonstrates that our system can achieve over 93% recognition accuracy and is robust to both environment changes and individual diversity. Results also show that our system can work with WiFi beacon signals and provides accurate gesture recognition under NLOS scenarios.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

Keywords

WiFi; Channel State Information (CSI); Gesture Recognition; Finger Gesture

1. INTRODUCTION

In recent years, gesture recognition is gaining increasing importance in human-computer interaction (HCI). Compar-

ing to traditional techniques using peripheral devices such as mouse or keyboard, gesture-based interaction serves as a more convenient and natural means for users to interact with computers. Gesture made with fingers is particularly crucial as our HCI bandwidth is the highest there due to finger dexterity [14]. Recognizing finger gesture is also extremely compelling for interacting with mobile and wearable devices and performing finger control in emerging applications, such as smart home, virtual reality, and mobile gaming. Google's Soli radar chip [3], for example, is recently developed for the wearables to recognize finger gestures.

Prior work in gesture recognition mainly relies on pre-installed depth and infrared cameras (e.g., kinect, leap motion) [1, 4, 5, 24, 27] or dedicated sensors (e.g., RFID, gloves, motion sensors) that are worn by the user [9, 20, 29, 12, 22, 33, 2, 21]. These approaches however require significant deployment overhead and incur non-negligible cost. In addition, the camera-based solution cannot work in non-line-of-sight (NLOS) scenarios. Recently, Radio Frequency (RF) based gesture recognition using either specialized [23, 8, 16, 7, 11] or commodity RF devices [30, 31, 19, 18, 28, 10] have drawn considerable attention as they don't require users to wear any physical sensors and can work under NLOS scenarios. These systems however only provide coarse-grained gesture recognition such as body activities [23, 8, 7, 30, 31, 34] or hand movements [16, 19, 18, 28]. For example, WiDraw [28] utilize a dense deployment of 25 RF devices to recognize large-scale hand movements. While WiKey [10] and the system proposed by Chen *et al.* [11] can recognize specific finger movements of typing, WiKey [10] requires the WiFi packets to be transmitted at outrageously high rate of 2500 packets/second (i.e., highest possible rate) and is very sensitive to environmental changes, and the system [11] relies on specialized software-defined radio to extract radio wave features that are not reported in commodity RF device. Such limitations and the high infrastructure cost make these methods hard to deploy for gesture recognition in a practical and user friendly system.

In this paper, we demonstrate that the commodity WiFi can be exploited for fine-grained finger gesture recognition which is both easily deployable and low-cost. Our proposed system takes advantages of the fine-grained wireless channel measurement of Channel State Information (CSI) and the prevalence of WiFi network infrastructure. First, the detailed physical layer measurement of CSI is internally tracked by IEEE 802.11 MIMO and is readily available in commodity WiFi devices. Such fine-grained CSI is able to detect the minute environment changes that alter signal

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MobiHoc'16, July 04-08, 2016, Paderborn, Germany

© 2016 ACM. ISBN 978-1-4503-4184-4/16/07...\$15.00

DOI: <http://dx.doi.org/10.1145/2942358.2942393>

propagation and multipath. It is thus capable of capturing the subtle movements of fingers to provide fine-grained gesture recognition. Leveraging detailed CSI to recognize gestures doesn't require users to wear any sensors and can work under both LOS and NLOS scenarios. Second, the prevalence of WiFi network infrastructures enables the proposed system to reuse existing WiFi devices and networks without requiring dedicated or specialized hardware. The system could reuse existing WiFi signals, for example, the beacon signals of WiFi networks, to perform finger gesture recognition. Reusing existing WiFi infrastructures not only advances and extends the applications that could be supported by WiFi networks but also enables easy and large-scale deployment of the proposed system due to the proliferation of WiFi devices and networks [6].

In particular, our system, WiFinger, utilizes only a single WiFi device with its connected AP to recognize finger gestures by examining the unique patterns exhibited in the detailed CSI. Accurately discerning the finger gestures is challenging, however, in this single WiFi device recognition system, because the multipath, shadowing, and fading components of signal could be dynamic due to the environment changes. For example, people walking around or moved furniture could change the multipath environment and affect the signal propagation. Such changes could also be sensed by the detailed CSI and may distort the CSI pattern of the finger gesture. Moreover, there exists individual diversity of each user such as the finger length, movement speed, and gesture consistency. Even for the same user, the same finger gesture could be slightly different from time to time due to the lack of consistency.

To handle environmental changes, we propose an environmental noise removal mechanism which employs multipath mitigation and wavelet based denoising to filter out the environmental noises while trying to keep the CSI patterns only resulted from the finger gestures. In particular, the multipath mitigation removes the signal components that arrive at the receiver through longer multipath propagations which are more likely affected by the changed environments, while the wavelet based denoising is used to further remove the high frequency noises while trying to keep sufficient details of CSI pattern for differentiating similar gestures. To deal with the individual diversity and gesture inconsistency, we propose to identify the principal components of the CSI pattern and to choose critical subcarriers that are sensitive to finger gesture for accurate gesture recognition. Specifically, the principal component identification exploits the idea of the intrinsic gesture behavior of the user [25] and extracts the gesture components which are invariant across the same set of finger gestures that one user performed for accurate gesture recognition.

We experimentally evaluate WiFinger in both office and home environments with typical finger gestures including zoom in/out, circle left/right, swipe left/right, and flip up/down. Result shows that our system achieves overall accuracy over 93% and is robust to both environment changes and individual diversity. It also shows that our system can work with WiFi beacon signals and provides accurate gesture recognition under NLOS scenario. The contributions of our work are summarized as follows:

- We show that the commodity WiFi can be reused to capture subtle changes of finger movements for fine-grained gesture recognition. Such approach doesn't

require any dedicated or specialized devices and can work under NLOS scenarios.

- We develop WiFinger which runs on a single WiFi device with its connected AP and extracts detailed CSI to profile the patterns of finger gestures. WiFinger, as a software-only solution that works with commodity WiFi devices, is both easily deployable and low-cost.
- We devise environmental noise removal mechanism to mitigate the effect of the environment changes. Such a method enables the WiFinger's robustness to various environmental interference such as people walking around and furniture changes.
- We exploit the principal component of the CSI pattern and select critical subcarriers for accurate gesture recognition. The principal component extraction makes our system resilient to individual diversity and gesture inconsistency.
- We conduct extensive experiments in both office and home environments with multiple participants under various conditions. The results show that WiFinger achieves over 93% recognition accuracy and can work with existing WiFi beacon traffic.

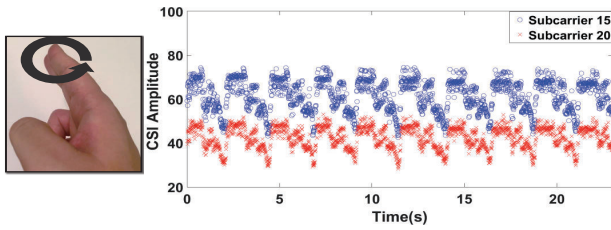
2. RELATED WORK

In general, the approaches for gesture recognition can be divided into three categories: wearable sensor based, camera based, and RF signal based.

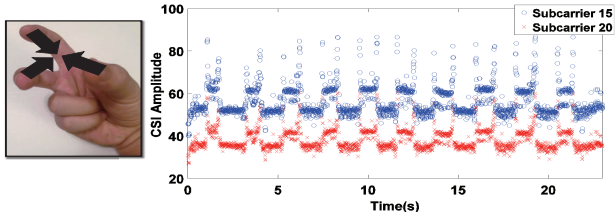
Wearable sensor based. Many research efforts have been done by using dedicated sensors worn by users' hand for gesture recognition. For example, Risq [22] utilizes inertial sensors on a wristband to recognize smoking gestures. Nelson *et al.* [20] developed a system using multi-sensor glove to recognize paralysis patients' gestures. Applications like text input using hand gestures also attract many attentions. PhonePoint Pen [9] for example recognizes human hand writing by holding mobile phone in hands. RFIDraw [29] tracks hand or finger movements by attaching RFID to user's fingers. Other wearable devices such as smartwatch [33] or wearable ring [12, 21] can also be used to enable text input recognition by hand movements. These methods however all require user to wear physical sensors.

Camera based. Early works [24, 27] has laid solid foundation for gesture recognition using dedicated cameras. Recent advancement in imaging technology enables depth or infrared cameras for gesture recognition, such as the ones used in Microsoft Kinect [1], Leap Motion [4] and WiiU [5]. Although these approaches do not require user to wear any sensors, they rely on dedicated hardware which incurs non-negligible cost and installation overhead, and only work under LOS scenario.

RF signal based. The RF signal based methods are most related to our work. Without requiring user to wear any physical sensors, RF based approaches can sense user motion under both LOS and NLOS scenarios. By using specialized RF devices, systems like WiSee [23] WiTrack [7] and Wi-Vi [8] are able to track large scale human movements. AllSee [16] and the system proposed by Chen *et al.* [11] are capable of tracking hand movements and even the specific finger movements of typing. Those systems however all rely on specialized hardware. Although the systems (e.g., E-eyes and CARM) [34, 23, 8, 7, 30, 31, 16, 19, 18, 28] use commodity WiFi devices, they can only identify large scale



(a) CSI measurements of Circle Left



(b) CSI measurements of Zoom In

Figure 1: Illustration of CSI measurements for two different finger gestures.

human activities or hand movements. While WiKey [10] can recognize some specific finger motions, it requires the WiFi packets to be transmitted at outrageously high rate of 2500 packets/second and is very sensitive to environment changes, which largely limits its practical applications.

Comparing to existing approaches, WiFinger can provide fine-grained finger gesture recognition using a single commodity WiFi device with its connected AP. It is a software-only solution that works with WiFi beacon signals, which is both easily deployable and low-cost.

3. SYSTEM DESIGN

To build an easily deployable and low-cost solution for fine-grained gesture recognition, we devise an approach that senses and identifies subtle movements of finger gesture leveraging commodity WiFi. In this section, we discuss the preliminaries, design challenges and goals, system overview, and the core components of the system.

3.1 Preliminaries and Challenges

WiFi has been evolving from providing laptop connectivity to connecting all kinds of mobile and smart devices with higher speed and more advanced technologies. It has resulted in the prevalence of WiFi devices and ubiquitous coverage of WiFi network, which provides the opportunity to extend WiFi’s capabilities beyond communication, particularly in sensing the physical environment. When the wireless signal propagates through space, any environment changes, either small scale or large scale, affect the received wireless signal, which is commonly known as shadowing and small-scale fading. With measurable changes in the received signals, activities in the physical environment could be potentially inferred. In particular, the 802.11a/g/n/ac employs OFDM technology which partitions the relatively wideband 20MHz channel into 52 subcarriers and provides detailed channel state information (CSI) of each subcarrier. The relative “narrow-band” subcarriers are very sensitive to the small movements in physical environment which results in the changes of CSI. On the contrary, the traditionally used received signal strength (RSS) is a coarse-grained informa-

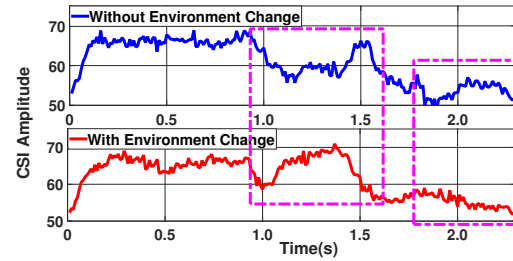


Figure 2: CSI patterns of Circle Left under different environments.

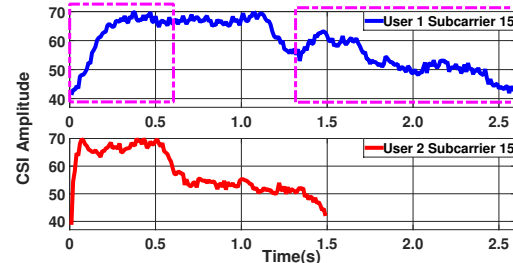


Figure 3: CSI pattern of Circle Left performed by different users.

tion which provides averaged power in the received signal over the whole channel bandwidth and may not capture such changes. We are thus motivated to reuse existing WiFi infrastructure to sense and identify subtle movements of finger gestures by leveraging the detailed CSI provided by the commodity WiFi device.

Figure 1 shows the CSI amplitude of two subcarriers (i.e., subcarrier 15 and 20) when a participant is repeatedly doing two finger gestures (circle left and zoom in) in front of the laptop. The CSI is extracted from the laptop that connected to a commercial AP in a 802.11n network. We observe that the CSI amplitude of those two subcarriers exhibits obvious periodic patterns and each of the finger gestures can be distinguished by its unique CSI pattern. This observation strongly indicates that the detailed CSI extracted from commodity WiFi could be analyzed for fine-grained figure gesture recognition.

Accurately discerning the finger gestures is however challenging because of the interferences from the surrounding environment. The interferences could come from the environment changes such as furniture change and people moving around. Such changes, for example a table/chair is moved to a different place or a person is walking around in the environment, alter the multipath environment which leads to construction or destruction (based on individual subcarrier phase shifts) effect in the combined signals at the receiver. Such effect could also be captured by the detailed CSI (due to subcarrier’s relative “narrow-band” nature) and creates distortion of the CSI pattern. Figure 2 illustrates such CSI pattern distortion at one subcarrier when a participant is doing circle left gesture with and without environment changes. We observe that the CSI patterns in the dash windows are heavily distorted due to one person is walking around in the environment. Such distortion could significantly degrade the accuracy of gesture recognition.

Moreover, the finger gesture is subjected to individual diversity and gesture inconsistency. Different people may have different finger and hand size, movement pace, and habit to perform finger gestures. Even for the same person, she/he

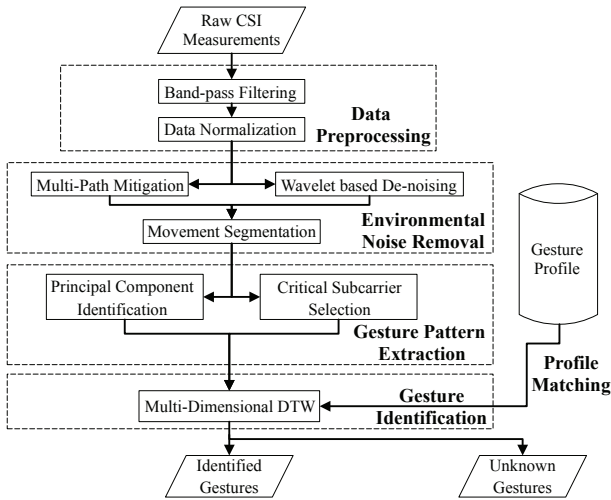


Figure 4: Overview of system flow.

could perform the same gesture slightly different from time to time due to lack of gesture consistency. Figure 3 shows the captured CSI amplitude of one subcarrier for the same gesture performed by two different users. Although the shape of these two CSI traces exhibits certain similarity, the length and some details of the CSI pattern are very different due to different finger movement speed and gesture inconsistency. In particular, the second user perform finger gesture much faster than the first one, and the patterns at the head and tail of the CSI traces have clear difference. The individual diversity and gesture inconsistency thus could seriously affect the robustness of the recognition system.

3.2 Design Goals

To accurately recognize the fine-grained finger gestures by using the detailed CSI from a single commodity WiFi device, the design and implementation of our system involve a number of challenges:

Easily Deployable. The system should be easily deployable on existing commodity WiFi without using any dedicated or specialized hardware or requiring users to wear any physical sensors. It should work with both LOS and NLOS scenarios and only utilize existing WiFi traffic or beacons at the deployed AP without dedicated user generated traffic.

Robust to Environmental Change. The interferences from the surrounding environment could dynamically change the detailed CSI. Our system should be able to provide accurate finger gesture recognition by mitigating interferences such as furniture change, people moving around, and body movements of the user.

Resilient to Individual Diversity and Gesture Inconsistency. Once the system is setup, it should be able to be used by multiple users without user-specific calibration. It thus should be resilient to both individual diversity and gesture variation due to lack of consistency.

3.3 System Overview

The basic idea of our system is to examine the unique pattern exhibited in the CSI measurements that extracted from a single commodity WiFi device. As each finger gesture has its unique CSI pattern (as shown in Fig. 1), the gesture recognition could be done by matching the CSI pat-

tern against the gesture profiles. Figure 4 shows the flow of our system. The system takes the time-series CSI measurements extracted from one single commodity WiFi device with its connected AP as input. It can reuse existing network traffic, such as WiFi beaconing signals, or system-generated periodic traffic (if network traffic is insufficient) to extract the detailed CSI. The measured CSI is then preprocessed to remove out-of-band noise via a butterworth filter and to normalize the CSI traces to the same scale.

The core of our system, WiFinger, consists of the *Environmental Noise Removal* and the *Gesture Pattern Extraction*. Environmental noise removal encompasses two different techniques to address the challenge of environmental interferences. It first employs *Multipath Mitigation* to mitigate the interference stemmed from the environment changes such as moved furniture and/or people moving around. It then utilizes *Wavelet Based Denoising* to further remove the noise by decomposing signals into approximation coefficients and detail coefficients. A dynamic thresholding method is applied to the detail coefficients to remove the noisy components while keeping sufficient details of the CSI pattern. After that, the system performs movement segmentation to separate the CSI measurements to each finger gesture.

Next, our system performs gesture pattern extraction by utilizing *Principal Component Identification* and *Critical Subcarrier Selection*. Principal component identification is used to capture the intrinsic user gesture behaviors by identifying the CSI components which remain stable within the user’s finger gestures. The identified principal components are usually invariant in the presence extensive variations in the user’s gesture, and hence resilient to individual diversity and gesture inconsistency. Our system then uses critical subcarrier selection to choose the subcarriers that have high sensitivity to the subtle movements of fingers gestures for gesture recognition.

At last, our system recognizes gestures by calculating the similarity between the extracted CSI pattern and the pre-constructed gesture profiles. As the CSI of multiple subcarriers can be used for extracting CSI pattern, *Multi-Dimensional Dynamic Time Warping (MD-DTW)* is used for similarity calculation by aligning CSI pattern to the gesture profiles while correcting for difference in finger movement speed. The one with the profile in the library that has the highest and also sufficient similarity with the testing CSI pattern is then identified as the recognized gesture.

To construct gesture profile, our system could utilize either a supervised or semisupervised approach. For example, one user could perform each finger gesture several times offline and then label the corresponding extracted CSI pattern in the profile library. The system can also continuously monitoring user’s gestures and identify multiple similar instances of CSI pattern without a matching profile. The user then could provide feedback to label such CSI pattern and deposit it to the profile library for subsequent gesture recognition. Moreover, the system could also use the semisupervised approach to update the CSI pattern when the gesture evolves to a slightly different version.

3.4 Environmental Noise Removal

In this subsection, we present the details of two techniques the system used to mitigate the interferences from the surrounding environment: multipath mitigation and wavelet based denoising.

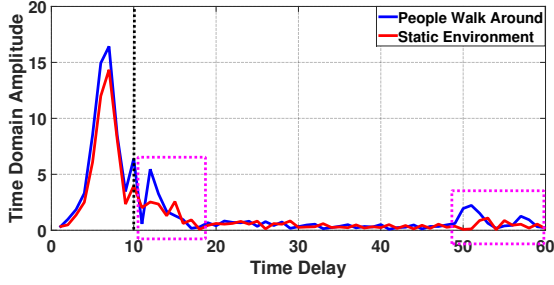


Figure 5: Power delay profile with and without environment change.

3.4.1 Multi-Path Mitigation

Multipath mitigation aims to remove the signal components that arrive at the receiver through longer multipath propagation. As the environment changes such as a table/chair is moved to a different place or a person is walking around will reflect the wireless signal and also create additional multipath, these reflected signals via created multipath will distort the CSI pattern of a finger gesture. Removing these reflected signal components could mitigate such interferences and hence makes the system robust to the environment changes. In particular, the signal reflection via multipath usually has longer propagation delays before arriving at the receiver. By transferring the frequency domain CSI into time-domain power delay profile, we could remove the signal components that have longer delay to mitigate the effect of the changed multipath.

Given the CSI measured at each subcarrier in frequency domain, we can obtain the power delay profile by applying the n -point Inverse Fast Fourier Transform (IFFT) [35]. The commonly used power delay profile is described as:

$$h(t) = \sum_{i=1}^N a_i e^{-j\theta_i} \delta(t - t_i) \quad (1)$$

where i denotes the sequence number of total N multipath channel, a_i , θ_i and t_i are the amplitude, phase angle and signal propagation time delay of i th path, and $\delta(t)$ is the Dirac delta function. Power delay profile gives the intensity of a signal received through a multi-path channel as a function of time delay.

Figure 5 shows the power delay profile with 60-point IFFT for the same gesture (i.e., circle left) under the scenarios with and without people moving around in the environment. We observe that the signal components in these two dash windows have obvious difference due to the environment change. We thus remove the signal components with large time delay (i.e., the part on the right side of the dash line in Figure 5) to retain the CSI pattern of the finger gesture while mitigating the effect of the changed multipath environment. After removing the signal components with larger delay, we apply an FFT transformation to convert the trimmed profile to frequency domain CSI. Previous study shows general indoor environment has the maximum delay less than 500 ns [15]. We use this value as a baseline for removing the signal components with longer delay, shown as the dash line in Figure 5.

3.4.2 Wavelet Based Denoising

Wavelet based denoising is used to further remove the noises presented in the collected CSI measurements. These

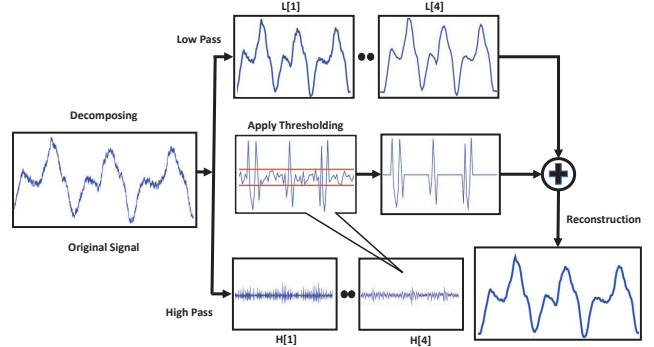


Figure 6: Illustration of Wavelet based denoising.

interferences could come from various sources such as the nearby electric devices and WiFi devices' inner noise. It is based on the Discrete Wavelet Transform (DWT) which analyzes the signal in both time and frequency domain and doesn't make any assumption about the nature of the signal. The DWT decomposes signals into approximation coefficients and detail coefficients. While the approximation coefficients describe the shape/trend of the signal which retain large scale characteristic of the CSI pattern, the detail coefficients capture the low-scale components which represent both high frequency noise and the fine details of the CSI pattern. As we are interested in removing the high frequency noise components while trying to keep sufficient details of CSI pattern for differentiating similar gestures, a dynamic thresholding is applied to the detail coefficients to remove the noisy components.

In particular, the wavelet based denoising includes three steps: decomposition, thresholding, and reconstruction. As shown in Figure 6, we first run the DWT based signal decomposition recursively by four levels with Symlet wavelet filter [26]. The DWT then yields both approximation coefficients α^J (with $J = 4$) and a sequence of detailed coefficients $\beta^1, \beta^2, \dots, \beta^J$. Each level of DWT coefficients are computed based on the following equations:

$$\alpha_k^{(J)} = \langle x_n, g_{n-2^J k}^{(J)} \rangle_n = \sum_{n \in \mathbb{Z}} x_n g_{n-2^J k}^{(J)}, \quad J \in \mathbb{Z} \quad (2)$$

$$\beta_k^{(\ell)} = \langle x_n, h_{n-2^\ell k}^{(\ell)} \rangle_n = \sum_{n \in \mathbb{Z}} x_n h_{n-2^\ell k}^{(\ell)}, \quad \ell \in \{1, 2, \dots, J\} \quad (3)$$

where x_n is the n^{th} input point, $\langle \cdot \rangle$ is the dot product operation, and wavelet basis represents by two sets of discrete orthogonal functions g 's and h 's.

We then apply dynamic thresholding to each level of detail coefficients $\beta^1, \beta^2, \dots, \beta^J$ to remove their noisy components. Finally, by combining all the resulting coefficients (i.e., the approximation coefficients and the detail coefficients after noisy removal), we reconstruct the final denoised CSI measurements with the inverse DWT. The inverse DWT is given by following formula:

$$x_n = \sum_{k \in \mathbb{Z}} \alpha_k^{(J)} g_{n-2^J k}^{(J)} + \sum_{\ell=1}^J \sum_{k \in \mathbb{Z}} \beta_k^{(\ell)} h_{n-2^\ell k}^{(\ell)} \quad (4)$$

The reconstructed measurements enable us to remove the noise components while keeping the detailed patterns of the CSI. This could facilitate accurate gesture recognition, especially for those gestures with similar shape of CSI patterns.

3.5 Gesture Pattern Extraction

We next detail the gesture pattern extraction component which is used to identify the principal components of CSI patterns and to choose critical subcarriers for accurate gesture recognition.

3.5.1 Principal Component Identification

The principal component identification borrows the idea of the intrinsic gesture behavior of the user in signature verification [25]. In particular, the CSI measurements of each finger gesture could be divided into several gesture components. Due to the individual diversity and gesture inconsistency, only part of these components are invariant across the same set of finger gestures that one user performed. We refer such components as principal components which capture the intrinsic gesture behavior of the user. Our system thus extracts these principal components to facilitate gesture recognition for improving the resilience to individual diversity and gesture inconsistency.

To identify the principal components of the CSI pattern, we examine and compare multiple instances of the same gesture. In particular, our method takes two instances of CSI measurements and compares them to find the best alignment by calculating a cost matrix and discovering the lowest cost route. The resulted lowest cost route could be represented by a coupling sequence in which the direct matching samples denote the components without significant distortion between two instances. We thus incorporate these direct matching samples into a weight vector to represent the principal components of the finger gesture. We run this process repeatedly between different pairs of instances that are available during the profile construction phase, and then average over the resulting weight vectors to obtain the principal components of each finger gesture.

Following shows the details of the principal component identification algorithm. After environmental noise removal, we first interpolated CSI measurements of each gesture instance to a fixed length L . We then assume $\{c_i, 1 \leq i \leq N\}$ is a set of interpolated CSI measurements with the fixed length L extracted from N gesture instances. The weight vector derived from a pair of instances c_i and c_j can be described as: $w_l^{c_i, c_j}$ where $i \neq j$ and $1 \leq l \leq L$. We then use the coupling sequence which is the alignment between c_i and c_j to estimate the weight value. All the direct matching samples in the coupling sequence are considered as the principal component candidates which represent the consistent gesture segments between two CSI instances. We simply use 1 as the weight if it is a principal component, and assign a weight 0 otherwise. At last, we generalize the principal components by averaging the weight vectors over each pair of CSI instances. Each averaged weight value ranges from 0 to 1 indicating the consistency of the corresponding segment of the CSI measurements. And a larger weight value means the corresponding segment is more stable when performing finger gestures. Our system thus values the segments with higher averaged weights more significantly during the gesture identification procedure as they represent the intrinsic gesture behavior.

Figure 7 shows one example on the process of the principal component identification using two gesture instances. We first calculate the cost matrix between these two instances, as shown in Figure 7(b). Based on the coupling sequence shown in the cost matrix, we identify these direct matching

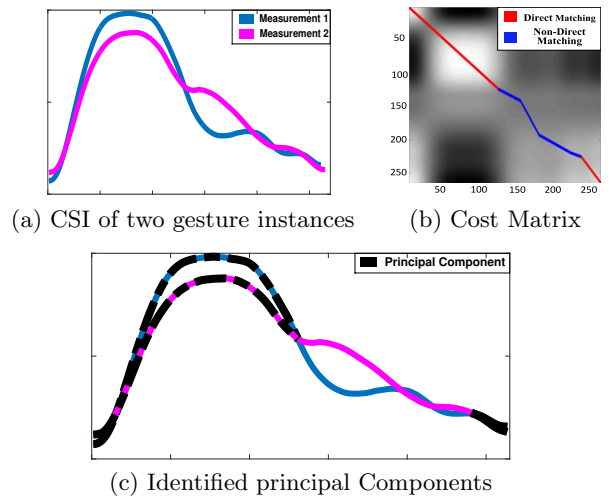


Figure 7: Illustration of principal component identification steps.

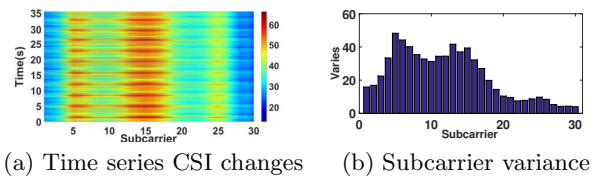


Figure 8: One example of subcarrier sensitivity to the finger gesture of Circle Left.

samples as the candidates of the principal components of the gesture. We then map these direct matching samples back to the CSI measurements of those two instances. The identified principal components in these two instances are highlighted as black color, as shown in Figure 7(c). During the gesture recognition phase, the principal components will be assigned with higher weights while the rest will be assigned with lower weights. The principal components identification thus could effectively capture the intrinsic gesture behaviors and improve the robustness of the system.

3.5.2 Critical Subcarrier Selection

Due to the frequency diversity, different subcarriers have different sensitivity to the subtle movements of finger gestures [17]. Figure 8 illustrates an example of time series CSI changes for 30 subcarriers when performing circle left gesture. We observe that the subcarriers with smaller indices are more sensitive to the circle left gesture, while the CSI from the higher subcarrier indices presents less changes. It is thus desire to assign more weights to these subcarriers with higher sensitivity for gesture recognition. Specifically, we calculate the variance of the CSI in a moving window in time series to quantify the sensitivity of the subcarriers to the finger gesture. We then filter out these subcarriers with small variance and use the variance values as the weights for the remain subcarriers.

3.6 Gesture Identification

As one user may perform gestures with different speeds and multiple subcarriers could be used for gesture identification, we utilize Multi-Dimensional Dynamic Time Warping (MD-DTW) [31] to align the CSI pattern to the gesture profiles in the library. MD-DTW allows us to overcome the

different speed problem by focusing on shifts in the CSI pattern such as the peaks and valleys, as evidenced by prior work [31]. It thus provides a robust metric for measuring the similarity between the testing CSI pattern and the gesture profiles among multiple subcarriers. In particular, the similarity is quantified by the Euclidean distance of the optimal warping path between the CSI pattern and the gesture profile. It is shown as

$$d(q_i, c_j) = \sum_{m=1}^M (q_{i,m} - c_{j,m})^2 \quad (5)$$

where $Q = q_1, q_2, \dots, q_m$ and $C = c_1, c_2, \dots, c_m$ are two CSI patterns where M is the number of chosen subcarriers. During gesture identification, our system uses MD-DTW to calculate the similarity between the testing CSI pattern and each of the known gesture profile in the profile library. The one with the highest and sufficient similarity to the CSI pattern is identified as recognized gesture. The CSI pattern with insufficient similarity to any known gesture profile is identified as an unknown gesture, which could then trigger user to label it in the profile library.

4. SYSTEM IMPLEMENTATION

4.1 Data Preprocessing

Human gesticulation is usually moving at a lower pace with the range between 0.3Hz and 4.5Hz [32]. And some of the noises have much higher frequency compare to finger gesture movements. We thus can apply a band-pass filter on the raw CSI measurements to remove these out-of-band noises. We choose Butterworth filter for band pass filtering as it has a maximally flat frequency response in the pass band which will not cause large distortion over finger gesture signal. In our system, we choose a 2-order Butterworth bandpass filter with cut-off frequency at 0.2Hz and 5Hz. After band-pass filtering, our system performs data normalization to scale different CSI traces in the same range. Data normalization could effectively overcome the problem of different transmission power of WiFi devices.

4.2 Segmentation

The system requires a user to have a short static interval between gestures to serve as the sentinel signal. Our system then can identify the movement of a gesture by detecting the static interval. In detail, we accumulate the amplitude differential between adjacent time points within each sliding window. The accumulated value is then compared to a empirical threshold for determining the sentinel signal for CSI trace segmentation.

5. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our WiFi-nger system using a commodity WiFi device in both office and home environments with multiple participants under various conditions.

5.1 Experimental Setup

5.1.1 Device and Network

We conduct experiments with a single WiFi device (i.e., Dell LATITUDE E5540 Laptop) connected to a commercial wireless Access Point (LINKSYS E2500 N600 Wire-

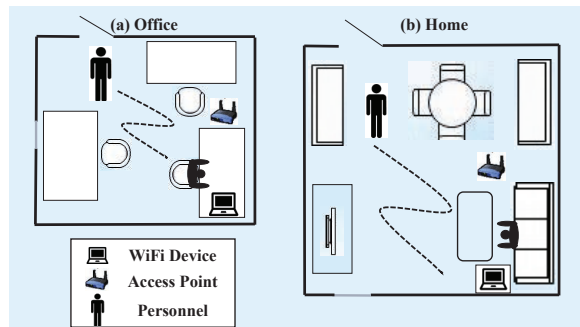


Figure 9: Illustration of experimental setup.

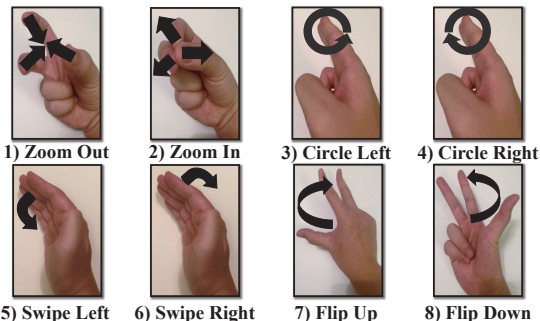


Figure 10: Illustration of eight finger gestures.

less Router) in an 802.11n WiFi network. The laptop runs Ubuntu 10.04 LTS and is equipped with an Intel WiFi Link 5300 for extracting CSI measurements [13]. The package transmission rate is set to 20 pkts/s. We will discuss the impact of packet rate on overall recognition accuracy in Section 5.6. For each packet, we extract CSI for 30 subcarrier groups, which are evenly distributed in the 56 subcarriers of a 20MHz channel.

5.1.2 Environments and Finger Gestures

We conduct experiments in both an office and an apartment environments with five participants. The experimental setup in these two environments are shown in Figure 9. The office has the size of about 9 ft by 9 ft with three tables and chairs, and some electronic devices inside, whereas the apartment is about 16 ft by 13 ft with regular living room furniture setup, such as dining table, book shelf, sofa, and TV. The office environment represents a more compact space filled with furniture, while the apartment setup describes a typical home environment with larger space. When the participant is performing the finger gesture, she/he is sitting on the sofa in the apartment environment and sitting in front of the table in office environment respectively. The AP and the laptop are placed at two sides of the sofa and table, as shown in Figure 9.

We evaluate the performance of our system with eight commonly used finger gestures including swipe left, swipe right, zoom in, room out, circle left, circle right, flip up, and flip down, as shown in Figure 10. These gestures are also widely used in current human-computer interaction systems such as Microsoft Kinect or Leap Motion. Each participant performs one gesture fifty times in office and apartment environments respectively. We use ten instances of each finger gesture to extract the CSI pattern for building the gesture profile. To test the robustness of our system to environment changes, we experiment with both furniture move and



(a) Home environment



(b) Office environment

Figure 11: Confusion matrix of finger gesture recognition under different environments.

people walking around scenarios. In particular, when the participant is performing finger gesture, a second person is randomly walking around within the environment to create interference. Examples of the walking trajectories are shown in dash curve in Figure 9. For the furniture change, we move the chairs and tables from one place to another inside the room.

5.1.3 Metrics

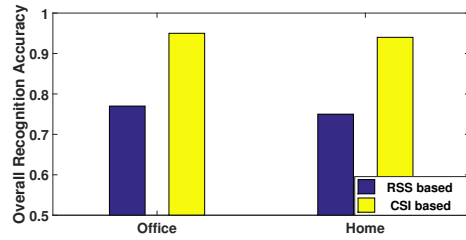
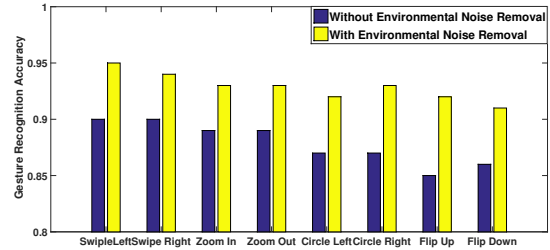
We use both confusion matrix and recognition accuracy to evaluate the performance of our system.

Confusion Matrix. Each column represents the finger gesture that was classified by our system and each row shows finger gestures performed the user. Each cell in the confusion matrix represents the percentage of finger gesture in the row that was classified as the gestures in the column.

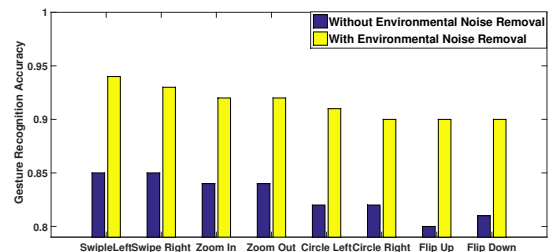
Recognition Accuracy. The percentage of the finger gestures correctly classified by our system.

5.2 Overall Performance

Figure 11 shows the confusion matrix of finger gesture recognition under both home and office environments. We observe that in both environments, our system achieves an overall recognition accuracy over 93% with the standard deviation at about 1.5%. By comparing the details of each finger gesture recognition in these two environments, we find that the recognition accuracy distribution are similar. In both environments, the swipe left and right have the highest recognition accuracy, whereas the flip up and down have the lowest accuracy. In particular, the swipe left achieves 96% and 95% accuracy in home and office environments respectively. This is possibly because of the relative larger finger movements involved in swipe left and right. Consequently, more finger movement details could be captured by CSI for differentiating from other similar finger gestures. The above results show that our system could provide high accuracy in recognizing finger gestures by using only a single WiFi de-

**Figure 12: Performance comparison when using CSI and RSS.**

(a) Furniture change



(b) People walking around

Figure 13: Recognition accuracy under the environment changes.

vice. The recognition accuracy could be further improved, for example, by using multiple available devices or the WiFi device equipped with multiple antennas.

We also compare the performance of using CSI to that of using RSS for finger gesture recognition. As RSS is the more sensitive to the physical movements when the transmitter and receiver are closer due to the log distance propagation, we place the WiFi device and the AP very close to each other (i.e., 3 ft) and compare the performance of CSI-based and RSS-based recognition in the same setup. Figure 12 illustrates the performance comparison of the overall recognition accuracy with each finger gesture tested for fifty times in each of these two environments. We observe that under the same setup, the CSI based method could achieve around 95% accuracy in both environments, whereas the RSS based method has only 76% recognition accuracy. It indicates that the detailed CSI could provide more fine-grained information than that of RSS, and can result in much better gesture recognition accuracy.

5.3 Impact of Environment Change

We next evaluate the robustness of our system to the environment changes. Specifically, we introduce environment changes including furniture change and people walking around that described in the experimental setup when the participant is doing finger gestures. We then compare the performance of our system with and without using en-

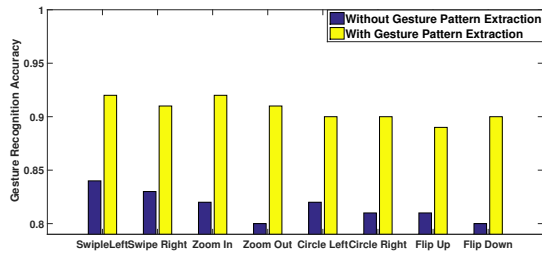


Figure 14: Recognition accuracy under individual diversity.

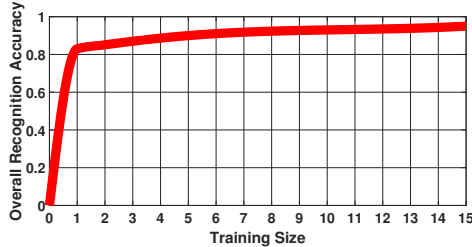


Figure 15: System performance under different training size.

environmental noise removal technique. Figure 13 depicts the performance comparison for each gesture recognition in the home environment. We find that the environmental noise removal technique improves the performance significantly for each of the finger gesture under both furniture change and people walking around scenarios. Moreover, by comparing Figure 13 with Figure 11(a), we observe that the performance doesn't have obvious degradation due to the use of environmental noise removal technique. In addition, we find that people walking around has larger impact on the CSI measurements, as indicated by the performance under the case without using environmental noise removal. This study demonstrates that our system can effectively mitigate the impact from the surrounding objects or people and is robust to the environment changes.

5.4 Impact of Individual Diversity

We further test the resilience of our system to individual diversity by applying the gesture profile built from one participant to another participant. We compare the performance of our system to the one without using the gesture pattern extraction method. Figure 14 presents the performance comparison for each finger gesture under individual diversity. We observe that without using gesture pattern extraction method, the performance degrades dramatically due to individual diversity and gesture inconsistency. Our system, with the gesture pattern extraction, provides much higher recognition accuracy than that of without using gesture pattern extraction method. For example, our system could improve the recognition accuracy by over 10% for most of the finger gestures. These results show that by incorporating the pattern extraction method, our system is resilient to individual diversity. Our system, once setup, could be used by multiple users without user-specific calibration.

5.5 Impact of Training Size

When building the profile for each finger gesture, our system requires to extract the CSI patterns from multiple gesture instances. We refer such number as the training size. Figure 15 studies the impact of training size to the per-

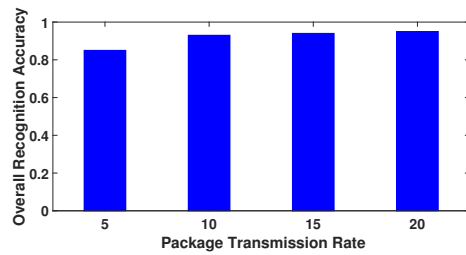


Figure 16: System performance under different packet rate.

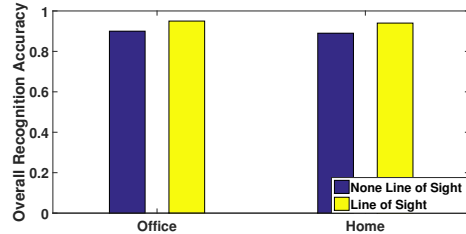


Figure 17: System performance under both NLOS and LOS scenarios.

formance of our system. Overall, we observe that our system could achieve considerable accuracy with a few tracing instances. In particular, with only one training size, our system achieves more than 80% of the recognition accuracy with one training instance. And the accuracy is improved to over 90% with five training instances. This result shows that our system could provide accurate gesture recognition with only a few training instances and hence doesn't incur high overhead on building the gesture profile, especially when the built profile from one user could be used by others.

5.6 Impact of Packet Rate

As a higher packet transmission rate results in more CSI measurements to capture the finger gestures, we are thus interested in whether existing WiFi traffic is sufficient to provide accurate gesture recognition. We experiment with four packet transmission rates, 5 pkts/s, 10 pkts/s, 15 pkts/s, and 20 pkts/s. The results are shown in Figure 16. We observe that a higher transmission rate results in a better recognition accuracy. Moreover, with 10 pkt/s transmission rate, our system is able to achieve more than 90% recognition accuracy. This demonstrates that our system could work with very low packet transmission rate. As the commercial AP sends beacon signals at 10 pkts/s, our system thus can reuse existing WiFi beacon signals for accurate gesture recognition.

5.7 Impact of NLOS

We study the impact of NLOS by placing the WiFi device and the AP in two connected rooms with the door closed. When the door is open, there exists LOS between the AP and the WiFi device. Figure 17 presents the performance comparison under the NLOS and LOS scenarios in both office and home environments. Results show that NLOS slightly degrades the system performance. Still, NLOS scenario has the overall recognition accuracy at around 90% in both office and home environments. It demonstrates that the proposed system could even work under the NLOS scenario. This allows us to deploy the proposed system to a wider range of application domains.

6. CONCLUSION

In this paper, we exploit the prevalence of WiFi devices and networks and design a system called WiFinger to perform fine-grained finger gesture recognition by utilizing the detailed CSI available in commodity WiFi devices. We find that CSI can capture the subtle movements of finger gestures. Our system benefits from such observation and examines the unique pattern exhibits in the detailed CSI for gesture recognition. To address the challenge of signal dynamic due to the environment changes, we devise environment noise removal mechanism to filter out the environmental noise while keeping the CSI pattern resulted from the finger gesture. Moreover, we propose to capture the intrinsic gesture behavior and to select critical subcarriers for accurate gesture recognition. Extensive experiments in both home and office environments demonstrate that WiFinger is effective in distinguishing a number of finger gestures, and that it can achieve over 93% recognition accuracy. In addition, we show that our system can work with WiFi beacon signals and provide considerable recognition accuracy under NLOS scenario.

7. ACKNOWLEDGEMENTS

This work is supported in part by the National Science Foundation Grants CNS-1514238, CNS-1505175 and CNS-1464092.

8. REFERENCES

- [1] Kinect. <https://dev.windows.com/en-us/kinect>.
- [2] Android Wear. <https://www.android.com/wear/>.
- [3] Google Project Soli. <https://www.google.com/atap/project-soli/>.
- [4] Leap Motion. <https://www.leapmotion.com/>.
- [5] Nintendo WiiU. <http://www.nintendo.com/wiiu>.
- [6] Study: 61 percent of U.S. Households Now Have WiFi. <http://techcrunch.com/2012/04/05/>.
- [7] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller. 3d tracking via body radio reflections. In *NSDI*, 2014.
- [8] F. Adib and D. Katabi. See through walls with wifi! In *ACM SIGCOMM*, 2013.
- [9] S. Agrawal, I. Constandache, S. Gaonkar, R. Roy Choudhury, K. Caves, and F. DeRuyter. Using mobile phones to write in air. In *ACM MobiSys*, 2011.
- [10] K. Ali, A. X. Liu, W. Wang, and M. Shahzad. Keystroke recognition using wifi signals. In *ACM MobiCom*, 2015.
- [11] B. Chen, V. Yenamandra, and K. Srinivasan. Tracking keystrokes using wireless signals. In *ACM MobiSys*, 2015.
- [12] J. Gummeson, B. Priyantha, and J. Liu. An energy harvesting wearable ring platform for gestureinput on surfaces. In *ACM MobiSys*, 2014.
- [13] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM CCR*, 2011.
- [14] J. A. Jacko and C. Stephanidis. *Human-computer interaction: theory and practice*. CRC Press, 2003.
- [15] Y. Jin, W.-S. Soh, and W.-C. Wong. Indoor localization with channel impulse response based fingerprint and nonparametric regression. *IEEE Transactions on Wireless Communications*, 2010.
- [16] B. Kellogg, V. Talla, and S. Gollakota. Bringing gesture recognition to all devices. In *NSDI*, 2014.
- [17] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *ACM MobiHoc*, 2015.
- [18] P. Melgarejo, X. Zhang, et al. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *ACM UbiComp*, 2014.
- [19] R. Nandakumar, B. Kellogg, and S. Gollakota. Wi-fi gesture recognition on existing devices. *arXiv preprint arXiv:1411.5394*, 2014.
- [20] A. Nelson, J. Schmandt, et al. Wearable multi-sensor gesture recognition for paralysis patients. In *IEEE SENSORS*, 2013.
- [21] S. Nirjon, J. Gummeson, D. Gelb, and K.-H. Kim. Typingring: A wearable ring platform for text input. In *ACM MobiSys*, 2015.
- [22] A. Parate, M.-C. Chiu, et al. Risq: Recognizing smoking gestures with inertial sensors on a wristband. In *ACM MobiSys*, 2014.
- [23] Q. Pu, S. Gupta, S. Gollakota, and S. Patel. Whole-home gesture recognition using wireless signals. In *ACM MobiCom*, 2013.
- [24] J. M. Rehg and T. Kanade. Visual tracking of high dof articulated structures: an application to human hand tracking. In *Springer Computer Vision-ECCV*. 1994.
- [25] Y. Ren, C. Wang, Y. Chen, M. C. Chuah, and J. Yang. Critical segment based real-time e-signature for securing mobile transactions. In *IEEE CNS*, 2015.
- [26] S. Sardy, P. Tseng, and A. Bruce. Robust wavelet denoising. *IEEE Trans. on Signal Processing*, 2001.
- [27] T. Starner and A. Pentland. Real-time american sign language recognition from video using hidden markov models. In *Springer Motion-Based Recognition*. 1997.
- [28] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim. Widraw: Enabling hands-free drawing in the air on commodity wifi devices. In *ACM MobiCom*, 2015.
- [29] J. Wang, D. Vasisht, and D. Katabi. Rf-idraw: virtual touch screen in the air using rf signals. In *ACM SIGCOMM*, 2014.
- [30] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu. Understanding and modeling of wifi signal based human activity recognition. In *ACM MobiCom*, 2015.
- [31] 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 *ACM MobiCom*, 2014.
- [32] Y. Xiong and F. Quek. Hand motion gesture frequency properties and multimodal discourse analysis. *International Journal of Computer Vision*, 2006.
- [33] C. Xu, P. H. Pathak, and P. Mohapatra. Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch. In *ACM HotMobile*, 2015.
- [34] J. Yang, Y. Ge, H. Xiong, Y. Chen, and H. Liu. Performing joint learning for passive intrusion detection in pervasive wireless environments. In *IEEE INFOCOM*, 2010.
- [35] X. Zheng, C. Wang, Y. Chen, and J. Yang. Accurate rogue access point localization leveraging fine-grained channel information. In *IEEE CNS*, 2014.