

Article

WiGeR: WiFi-Based Gesture Recognition System

Mohammed Abdulaziz Aide Al-qaness^{1,2} and Fangmin Li^{3,*}

¹ Key Laboratory of Fiber Optical Sensing Technology and Information Processing, Ministry of Education, Wuhan University of Technology, Wuhan 430070, China; alqaness@whut.edu.cn

² School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China

³ Department of Mathematics and Computer Science, Changsha University, Changsha 410022, China

* Correspondence: lfm@ccsu.edu.cn; Tel.: +86-731-84261249

Academic Editors: Sisi Zlatanova and Wolfgang Kainz

Received: 11 February 2016; Accepted: 6 June 2016; Published: 14 June 2016

Abstract: Recently, researchers around the world have been striving to develop and modernize human–computer interaction systems by exploiting advances in modern communication systems. The priority in this field involves exploiting radio signals so human–computer interaction will require neither special devices nor vision-based technology. In this context, hand gesture recognition is one of the most important issues in human–computer interfaces. In this paper, we present a novel device-free WiFi-based gesture recognition system (WiGeR) by leveraging the fluctuations in the channel state information (CSI) of WiFi signals caused by hand motions. We extract CSI from any common WiFi router and then filter out the noise to obtain the CSI fluctuation trends generated by hand motions. We design a novel and agile segmentation and windowing algorithm based on wavelet analysis and short-time energy to reveal the specific pattern associated with each hand gesture and detect duration of the hand motion. Furthermore, we design a fast dynamic time warping algorithm to classify our system’s proposed hand gestures. We implement and test our system through experiments involving various scenarios. The results show that WiGeR can classify gestures with high accuracy, even in scenarios where the signal passes through multiple walls.

Keywords: device-free; CSI; WiFi; gesture recognition; human–computer interfaces

1. Introduction

Recently, gesture recognition systems have become increasingly interesting to researchers in the field of human–computer interfaces. Specifically, the promise of device-free gesture recognition systems is the lure that attracts researchers to this new and promising technology. Building interactive systems based on wireless signals (such as ubiquitous WiFi) that do not require installing cameras or sensors will permanently change the computing industry and smart device manufacturing: for example, when manufacturing a smart interactive TV, manufacturers would not need to equip the TV with expensive sensors or vision-based technologies; instead, they could adopt device-free gesture recognition technology. Therefore, this technology has the potential to provide a tremendous advancement in the field of human–computer interaction that will affect both smart home systems and smart device manufacturing.

Traditional gesture recognition systems depend on vision technology such as a Microsoft Kinect [1] or wearable sensors such as Magic rings [2]. For the future, researchers are endeavoring to eliminate such sensors and move straight to the promising technology of device-free sensing systems. Device-free systems were given this name because the object perceived by the computer does not need cameras or sensors in the sensing area or on the perceived object. Previous approaches show that WiFi signal analysis can support localizing humans in indoor environments both when in line-of-sight (LOS) and non-line-of-sight (NLOS), such as through walls and behind closed doors.

WiFi signals can also be used to identify the number of people in a room along with their locations [3]. Some of the previous approaches have been extended to leverage WiFi signals analysis for hand gesture recognition systems, which enable humans to communicate with devices without requiring sensors or vision-based technology [4]. However, WiFi-based gesture recognition systems still lack security and are still limited in their sensing capabilities through multiple walls.

In this paper, we present a device-free hand gesture recognition system that enables humans to interact with home devices and also provides user security and device selection mechanisms. We leverage the Channel State Information (CSI) of wireless signals as a metric for our system. Previous WiFi-based sensing mechanisms leveraged the received signal strength (RSS) from the wireless MAC layer to track human motion. However, the RSS value decreases as distance increases and suffers from multipath fading in complex environments. Moreover, RSS is measured from WiFi signals per-packet level; therefore, RSS is single-valued. In contrast, physical layer CSI is more robust to environmental changes and is measured at the orthogonal frequency-division multiplexing (OFDM) subcarrier level from the received packet [5]. Therefore, we leverage CSI to build a more stable and more robust system.

In general, previous WiFi-based gesture recognition approaches applied a moving window and extracted several features from the time and frequency domains of WiFi signals—such as the median, standard deviation, second central moment, least mean square, entropy, interquartile, signal energy, and so on. Experimentally, we find that some gestures produce values similar to other features that can raise the false detection rate, especially in through-wall scenarios. Increasing the number of features also increases the accuracy but limits the speed of the system. Therefore, our system does not extract such features; instead, we design an efficient segmentation method based on wavelet analysis and Short-Time Energy (STE). The wavelet algorithm reveals the unique pattern associated with each gesture and the STE algorithm dynamically detects the duration of each gesture in the signal time series. Therefore, our system selects these patterns as the features of each gesture. It is worth mentioning that each pattern detected by our system preserves most of the frequency domain and time domain components; therefore, our feature selection method considers even more features than previous approaches, which mainly select a small number of features that may not be sufficient for micro-motions like hand gesture. Thereafter, we adopt a dynamic time warping (DTW) algorithm to compare gesture patterns and to discriminate between the gestures, because each gesture pattern has tens of points. We choose DTW because of its high efficiency in matching the corresponding points of two time-series.

In creating a device-free gesture recognition system, we face three main challenges. The first challenge is the problem of how to classify gestures through multiple walls. From previous approaches, we find that CSI-based gesture classification systems still have limited through-wall sensing capabilities. We solve this problem by building an efficient segmentation method and fast classification algorithm that can recognize gestures accurately through one wall by using only a ubiquitous WiFi router, such as TP-LINK TL-WR842N. Then, to extend hand gesture recognition to scenarios when the user is separated from the receiver by multiple walls, we modify the transmitter hardware (the WiFi router) by installing high throughput antennas in place of the router's original antennas. The second challenge is how to add security to a device-free gesture recognition system. Adding security is a crucial problem and a limiting factor in most previous approaches that, when it is omitted, will lead to an insecure system. For example, without built-in security, a neighbor moving his hand could control his nearby neighbors' devices. The third challenge is avoiding random interactions of user and devices that work through a WiFi-based gesture recognition system (WiGeR). This is also a crucial issue that has the potential to cause chaos in human-computer interfaces. For instance, if a user swipes his hand leftward and his TV changes the channel, his laptop swipes to the next page, and his air-conditioner increases the temperature setting, the result will be chaos. Obviously, such a scenario involves an inappropriate system response. To solve the second and third challenges, our system requires the user to perform a unique gesture that the system can recognize and treat as that user's identifier or authorized gesture.

After a gesture has been detected as an authorized user gesture, WiGeR will enable the user to begin interacting with devices. Moreover, each device also requires the user to perform a unique gesture that can be treated as a device selection gesture. When WiGeR detects a specific device selection gesture, the device will subsequently respond to other user interaction gestures. This simple solution is based on novel and efficient pattern segmentation and classification algorithms. The gestures can be modified through learning by both users and devices.

In this paper, our main contributions are as follows:

- We present a gesture recognition system that enables humans to interact with WiFi-connected devices throughout the entire home, using seven hand gestures as interactive gestures, three air-drawn hand gestures that function in the security scheme as users' authenticated gestures, and three additional air-drawn hand gestures that function as device selection gestures. Unlike previous WiFi-based gesture recognition systems, our system can work in different scenarios—even through multiple walls. Moreover, our system is the first device-free gesture recognition system that provides security for users and devices.
- We design a novel segmentation method based on wavelet analysis and Short-Time Energy (STE). Our novel algorithm intercepts CSI segments and analyzes the variations in CSI caused by hand motions, revealing the unique patterns of gestures. Furthermore, the algorithm can determine both the beginning and the end of the gesture (the gesture duration).
- We design a classification method based on Dynamic Time Warping to classify all proposed hand gestures. Our classification algorithm achieves high accuracy in various scenarios.
- We conduct exhaustive experiments. Each gesture has been tested several hundred times. The experiments provide us with important insights and improve our system classification ability by letting us choose the best parameters—gained through experience—to build a gesture learning category profile that can handle gestures with various users' physical shapes in different positions.

The remainder of this paper is organized as follows. Section 2 provides an overview of some related works. Section 3 presents the system architecture. Section 4 describes the methodology. Section 5 discusses the experimental setup and results. Section 6 concludes this paper and provides suggestions for future research.

2. Related Works

2.1. Device-Based Gesture Recognition Systems

Commercial gesture recognition systems utilize various technologies to identify a wide variety of gestures. These technologies include cameras and computer vision and room-based or built-in sensors such as cameras [6], PointGrab [7], laptops [8], smartphones [9,10], GPS [11], vision-based [12], and accelerometers [13,14], as well as sensors worn on the human body—such as rings [15], armbands [16], and wristbands [17,18].

However, they have drawbacks. For example, vision-based systems require a line of sight between the user and the camera device; interior sensors need special sensors installed; and, of course, wearable devices need to be worn by the user, which can be inconvenient. Moreover, none of these systems are free. In contrast, our system interacts with WiFi-connected devices without any special devices and costs nearly nothing.

2.2. Device-Free Gesture Recognition Systems

The existing works on device-free wireless-based sensing systems can be categorized into three main trends: (1) received signal strength (RSS) based systems; (2) Radio Frequency (RF) based systems, sometimes called Software Defined Radio (SDR); and (3) Channel State Information (CSI) based systems. Here, we review some of the previous approaches to describe the latest developments in this research area.

2.2.1. RSS

Abdelnasser *et al.* [19] proposed a WiFi-based gesture recognition system (WiGest) by leveraging changes in RSS due to hand gestures. The WiGest system can identify several hand gestures and map them to commands to control various application actions. WiGest achieved gesture recognition accuracy rates of 87.5% and 96% using a single access point and three overhead access points, respectively. Nonetheless, RSSI is an inadequate metric because the severe variation in RSSI values causes continual misdetection. Furthermore, WiGest and others RSS-based gesture recognition systems still lack security and cannot operate through walls.

2.2.2. SDR

Recently, various device-free sensing systems have been proposed that use Software Defined Radio with devices such as Universal Software Radio Peripheral (USRP) and Radio-frequency identification (RFID) readers. Pu *et al.* [4] presented the WiSee system, which can recognize nine body gestures to interact with home Wi-Fi-connected devices by leveraging the Doppler shift of wireless signals. Adib *et al.* presented WiVi [3], Witrack [20], and WiTrack02 [21]. These systems can track human movement through walls and classify simple hand gestures. Wang *et al.* [22] presented the RF-IDraw system, which was designed for commercial RFID readers. RF-IDraw enables a virtual touch screen, allowing a user to interact with a device through hand gestures. Kellogg *et al.* [23] designed the AllSee system that can identify human gestures using RFID tags and power harvesting sensors, but AllSee cannot work with WiFi signals; it works only with TV and RFID transmissions.

All these SDR-based systems require special devices whose costs are high and whose installation is burdensome. In contrast, our system requires only an access point (a Wi-Fi router) which are now ubiquitously available almost everywhere. In addition, SDR-based systems suffer from a lack of security. In contrast, our system supports security for both users and devices.

2.2.3. CSI

Channel State Information-based sensing systems have been designed for various purposes, such as localization [24], human motion detection [25], and counting humans [26]. Recently, CSI has been extended to recognize human motions such as fall detection [27], daily activity recognition [28], micro-movement recognition [29], and gesture recognition [30,31]. Nandakumar *et al.* [30] presented a hand gesture recognition system that can identify four hand gestures with an accuracy of 91%, and 89% in LOS and in a backpack, respectively. He *et al.* [31] presented a hand gesture recognition system called WiG. WiG classifies four hand gestures in both LOS and NLOS with an accuracy of 92% and 88%, respectively. Compared to the SDR-based systems, previous CSI-based gesture recognition systems have proposed fewer and simpler hand gestures. In contrast, our system is designed for whole-home use, able to control multiple devices with sufficient interactive gesture recognition capabilities.

In summary, all previous device-free gesture recognition systems still suffer from security issues, and still lack the ability to sense gestures through multiple walls. In this paper, we present the WiGeR system, which overcomes these limitations and outperforms previous device-free gesture recognition systems.

3. System Overview

3.1. Gestures Overview

We design a novel system that enables a human to send messages to devices wirelessly without having to carry or wear any wireless or sensor device to implement a command or relay a message to the device with which they wish to communicate. Our system recognizes seven hand gestures for controlling devices as shown in Figure 1.

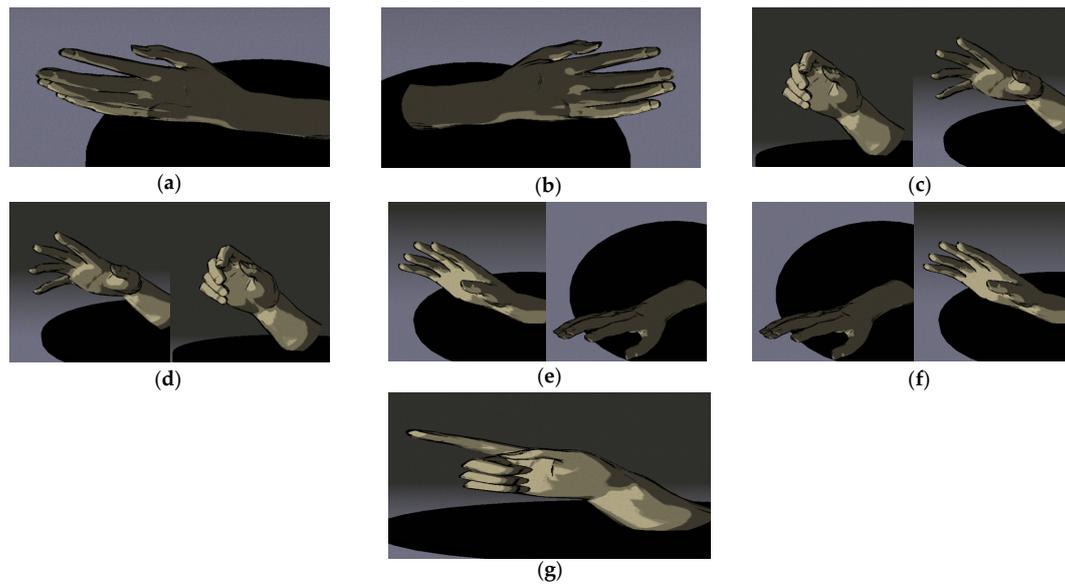


Figure 1. Seven hand gestures for interacting with home devices wirelessly. The gestures are as follows: (a) Swipe leftward: the user moves his hand from right to left; (b) Swipe rightward: the user moves his hand from left to right; (c) Flick: the user first grabs his hand, pushes it forward, and then flicks it; (d) Grab: the user first flicks his hand and then grabs it and pulls it backward; (e) Scroll up: the user raises his hand up from the side of his body; (f) Scroll down: the user moves his hand down; (g) Pointing: the user points forward.

These seven gestures can be adjusted to control several devices with different functions; Table 1 lists the tasks that these gestures can accomplish.

Table 1. Proposed gestures utilized for different devices.

Gesture	TV	Laptop, Smart Device	Air-Conditioner
Leftward	(−) Change channel	Swipe files leftward	(−) Temperature setting
Rightward	(+) Change channel	Swipe files rightward	(+) Temperature setting
Flick	Zoom in	Zoom in	(+) Fan speed
Grab	Zoom out	Zoom out	(−) Fan speed
Scroll up	(+) Volume	Scroll pages/files up	Change mode heat/cold
Scroll down	(−) Volume	Scroll pages/files down	Change mode cold/heat
Point	Open/Close/Enter	Open/Close/Enter	On/Off

As a unique recognition method, users are required to perform some unique gesture for authorization. We asked three users to perform three different gestures by drawing special shapes in the air. Here, we propose three shapes as described in Figure 2a–c. WiGeR can recognize and authenticate a user by classifying the implemented gesture; subsequently, that user can interact with a target device. Each target device also has a special gesture. The user is asked to implement draw-in-air gesture to be able to select a target device to interact with. We propose three shapes to be drawn in the air as illustrated in Figure 2d–f. If WiGeR classifies the performed gestures as a specific device gesture, the device will be able to respond to subsequent commands by that user. This solution protects our interactive system and avoids both outside intrusion and unwanted random user–device interaction.

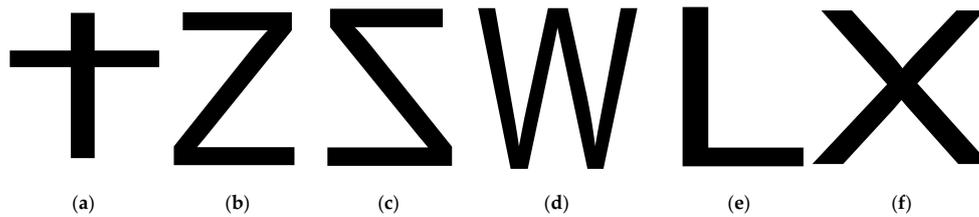


Figure 2. User security gestures and device selection gestures. (a) User 1 security gesture: drawing a cross shape in the air; (b) User 2 security gesture: drawing a Z shape in the air; (c) User 3 security gesture: drawing the inverse of a Z shape in the air; (d) Device 1 selection gesture: the user is asked to draw a W shape in the air; (e) Device 2 selection gesture: the user is asked to draw an L shape in the air; (f) Device 3 selection gesture: the user is asked to draw an X shape in the air.

3.2. System Architecture

Our system consists of the following stages:

1. Preparation: In this stage, WiGeR collects information from the WiFi access point, extracts amplitude information, and then filters out the noise.
2. Pattern segmentation: In this stage, the system differentiates between gestures. We apply a multi-level wavelet decomposition algorithm and the short-time energy algorithm to extract gesture patterns and detect the start and end point of a gesture, respectively, and detect the width of the motion window.
3. Gesture classification: In this stage, the system compares the patterns in each gesture window. The DTW algorithm is applied and accurately classifies the candidate gestures.

Figure 3 shows the WiGeR architecture and workflow. In Section 4, we will explain each stage in detail.

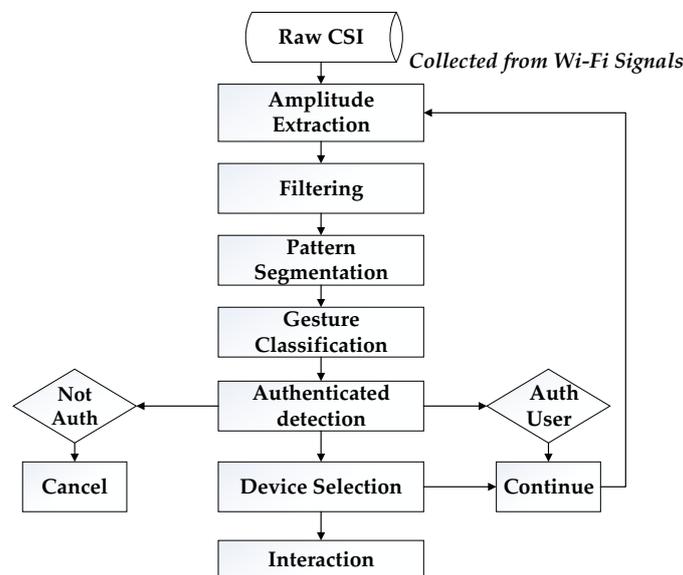


Figure 3. System Architecture and Workflow. WiFi-based gesture recognition system (WiGeR) starts by collecting channel state information (CSI) and then extracts amplitude information, removes noise, detects abnormal patterns, and classifies gestures. If the first gesture is classified as an authenticated user, the system will continue. Otherwise, the process will fail.

4. Methodology

4.1. Preparation

Channel state information portrays how the signals propagate from transmitter to receiver in a wireless communication system and exposes the channel characteristics by amplitude and phase information of every subcarrier. Let H denote the Channel State Information and T and R express the transmitter and receiver, respectively. The relation between CSI, transmitter, and receiver can be expressed as Equation (1):

$$R = HT + N \quad (1)$$

where N is the noise. A WiFi channel in the 2.4 GHz band can be considered as a narrowband flat-fading channel, which can be interpreted as shown in Equation (2) [32]:

$$H_{i,j} = \| H_{i,j}(f_k) \| e^{j\angle H_{i,j}(f_k)} \quad (2)$$

where $\| H_{i,j}(f_k) \|$ is the amplitude, $\angle H_{i,j}(f_k)$ is the phase, i is the stream number, and j is the subcarrier number. In our system, we leverage only the amplitude information of CSI, due to the instability of phase information.

We use a WiFi TP-Link router that has two antennas as the transmitter or access point (AP) and a Lenovo ThinkPad x201 Laptop equipped with an Intel WiFi link IWL 5300 network card as a receiver or detection point (DP). This means that we have a 2*3 MIMO system. According to the OFDM scheme, CSI is divided into six streams, and each stream has 30 subcarriers. Therefore, 180 groups of data can be derived from each received packet as illustrated in Equation (3):

$$\| H_{i,j}(f_k) \| = \begin{bmatrix} \| H_{1,1}(f_k) \| & \dots & \| H_{1,30}(f_k) \| \\ \vdots & \vdots & \vdots \\ \| H_{6,1}(f_k) \| & \dots & \| H_{6,30}(f_k) \| \end{bmatrix} \quad (3)$$

We aggregate the CSI in the 30 subcarriers into one single value H_i , where i is the stream number. Thus, we acquire six streams that can represent CSI values, as shown in Figure 4.

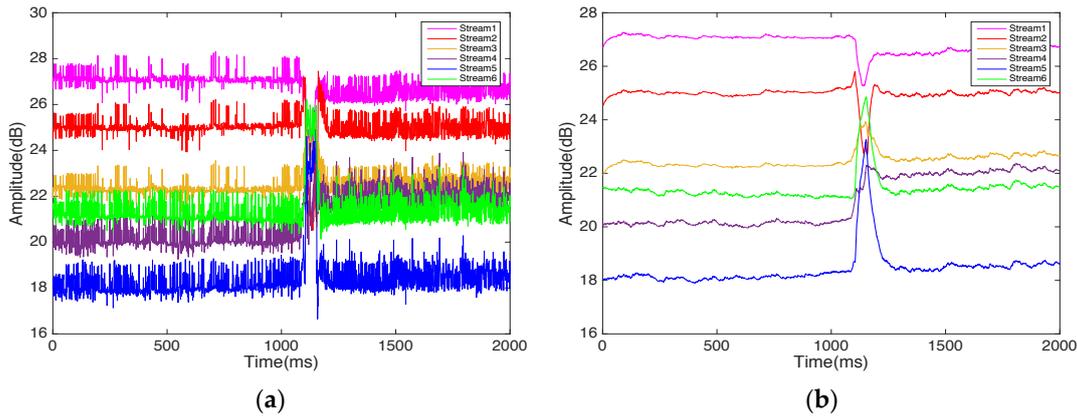


Figure 4. Raw CSI before and after filtering (a) Raw CSI represented by six streams with noise; (b) Raw CSI represented by six streams after applying the Butterworth Filter.

As shown in Figure 4a the real trend of CSI is drowned in the noise. Hence, we design a Butterworth low-pass filter to remove the noise but preserve the real trend of CSI caused by hand motion. As shown in Figure 4b, our filter effectively removes all the noise.

4.2. Pattern Segmentation

Moving objects such as hand gestures in the test environments are associated with a duration for analyzing gesture performance with the raw CSI frequency. The duration of different gestures may be identical, but the frequency of each gesture is different. Therefore, we need to extract the frequency and time components of different hand gestures at different durations to obtain a unique pattern for each hand gesture. To this end, we design an efficient pattern segmentation algorithm based on wavelet analysis and Short Time energy.

First, we design a multi-level wavelet decomposition algorithm to decompose CSI and analyze its frequency and time components to obtain a unique pattern for each gesture. WiGeR calculates the mean value of CSI six streams to obtain a CSI one-dimension vector which is the input of the wavelet analysis algorithm. Then, the multilevel 1-dimension wavelet decomposition is applied as follows:

$$f[CSI] = \sum_k a_{j_0k} \phi_{j_0k}(CSI) + \sum_{j=0}^{\infty} \sum_k \beta_{jk} \psi_{jk}(CSI) \quad (4)$$

where a_{j_0k} and β_{jk} are the approximation coefficients and detailed coefficients, respectively. $\phi_{j_0k}(CSI)$ and $\psi_{jk}(CSI)$ are called the scaling function (father wavelet), and the wavelet function (the mother wavelet). a_{j_0k} and β_{jk} can be calculated as

$$a_{j_0k} = \int f(CSI) \overline{\phi_{j_0k}} dCSI, \quad \beta_{jk} = \int f(CSI) \overline{\psi_{jk}} dCSI \quad (5)$$

Here, we test our classifier with Daubechies wavelet and Symlets wavelet; we find that Symlets of order 4 (*sym4*) up to level 5 ($j = 5$) achieves the best accuracy. Therefore, we use Symlets (*sym4*) as a wavelet type, and we select the detailed coefficients of the level ($j = 5$) as the wavelet analysis result that represents the CSI of the implemented gesture. Figure 5 shows the result of the wavelet decomposition algorithm for three security gestures and three device-selection gestures (drawn in the air). Note that each hand gesture has a unique pattern in the raw CSI. The three patterns in Figure 5a represent the three security gestures and the three patterns in Figure 5b represent the three device-selection gestures.

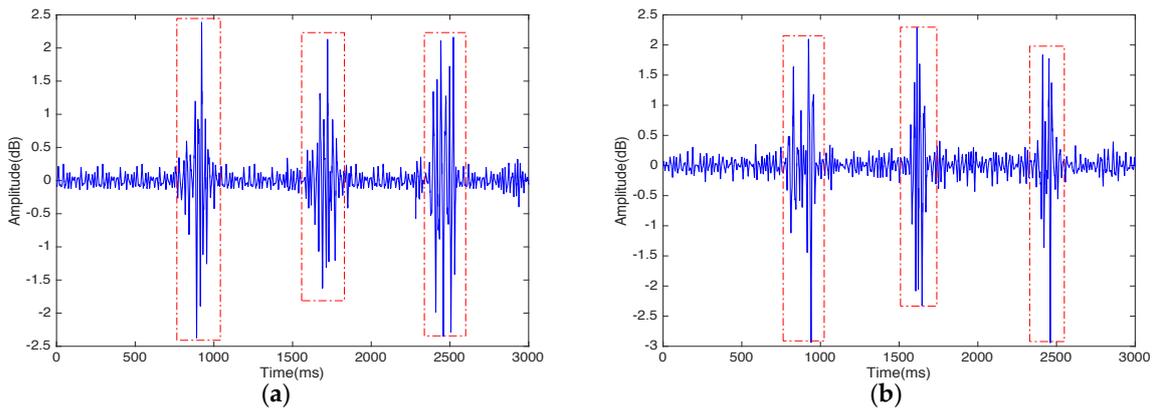


Figure 5. Wavelet analysis results of air-drawn gestures. (a) Wavelet analysis result of security gestures; (b) Wavelet analysis results of device-selection gestures.

Second, we apply short-time energy (STE) algorithm to the results of the wavelet decomposition of CSI. STE is a famous algorithm used in speech signal processing that can classify speech and silent periods. A high STE rate denotes the time during which speech is occurring, whereas a low rate reflects non-speech or silence. In the same manner, the STE rate will vary in consonance with the variations in

CSI. We design our algorithm as suggested in the definition of STE mentioned in [33]. We redefine STE for CSI as follows.

The long-term energy of CSI signal can be expressed as shown in Equation (4):

$$E = \sum_{csi=n-N+1}^n h^2(csi) = h^2(n-N+1) + \dots + h^2(n) \quad (6)$$

where E is the energy of the CSI signal after wavelet decomposition ($h(csi)$). The wavelet decomposition $h(csi)$ can be expressed as shown in Equation (5):

$$h_n(csi) = h(csi)w(n-csi) \quad n-N+1 \leq csi \leq n \quad (7)$$

where $n = 0, 1, 2, \dots, N$ is the window length. The STE of CSI can be determined from the following expression:

$$E_n = \sum_{csi=n-N+1}^n [h(csi)w(n-csi)]^2 \quad (8)$$

where $w()$ is the window function, n is the sample that the analysis window is centered on, and N is the window length. Here, we apply a rectangular window.

In Figure 6, the red line represents STE, which follows the CSI signal and increases as CSI varies because of human hand motion. The width of the time window can be seen clearly.

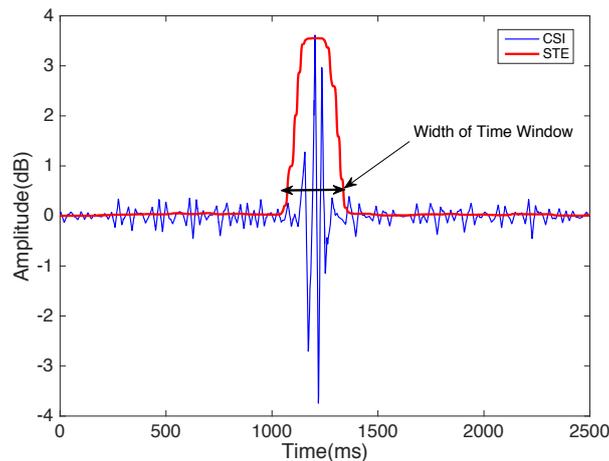


Figure 6. Detecting hand gesture interval by short-time energy with a rectangular window. The short-time energy (STE) rate changes according to the variations in CSI generated by a user's hand motion.

4.3. Classification

We adopt the dynamic time warping (DTW) algorithm to classify gestures for our system. DTW matches the temporal distortions between two models to find the warp path between the corresponding points in two time-series. In general, DTW [34], given two time-series g and \hat{g} with lengths g_m and \hat{g}_n , respectively, can be written as:

$$g = g_1, g_2, \dots, g_i, \dots, g_m \quad \hat{g} = \hat{g}_1, \hat{g}_2, \dots, \hat{g}_j, \dots, \hat{g}_n$$

The warp path W between the g and \hat{g} time-series can be represented as follows:

$$W = W_1, W_2, \dots, W_L \quad \text{where} \quad W_l = (i, j) = \|g_i - \hat{g}_j\|$$

where L is the length of the warp path, l denotes the l^{th} element of the path, i is an index in the time series g , and j is an index in the time series \hat{g} . Therefore, the warp path W can be defined as a Euclidian distance and calculated as shown in Equation (7):

$$W(g_i, \hat{g}_j) = \sqrt{\sum_{l=1}^L (g_i - \hat{g}_j)^2} \quad (9)$$

However, the optimal warp path is considered as the minimum distance warp path of two-time series. Therefore, the distance of the optimal warp path can be calculated as shown in Equation (8):

$$D(g_i, \hat{g}_j) = \min \sum_{l=1}^L W(g_i, \hat{g}_j) \quad (10)$$

where $D(g_i, \hat{g}_j)$ is the distance between the two data point indices of two time series of hand gestures (indices from g and from \hat{g}) in the l^{th} element of the warp path. Our DTW dynamically sets up warp path index points based on the starting and ending points of the hand motion detected by the pattern segmentation algorithm (discussed in Subsection 4.2) and, consequently, speeds up the classification process.

WiGeR can dynamically classify gestures by comparing the warp path distance of gestures based on minimum distances that were determined from exhaustive experiments during the learning phase. These gestures are stored in the system profile and considered as identified gestures. We use the open source machine learning tool box [35]. Figure 7 shows the alignments between two-time series of two hand gestures. DTW dynamically compares the intervals of hand motion detected by the STE algorithm. DTW is a fast and agile algorithm that improves our system's performance. The experimental results verify the validity and efficiency of our segmentation and classification methods.

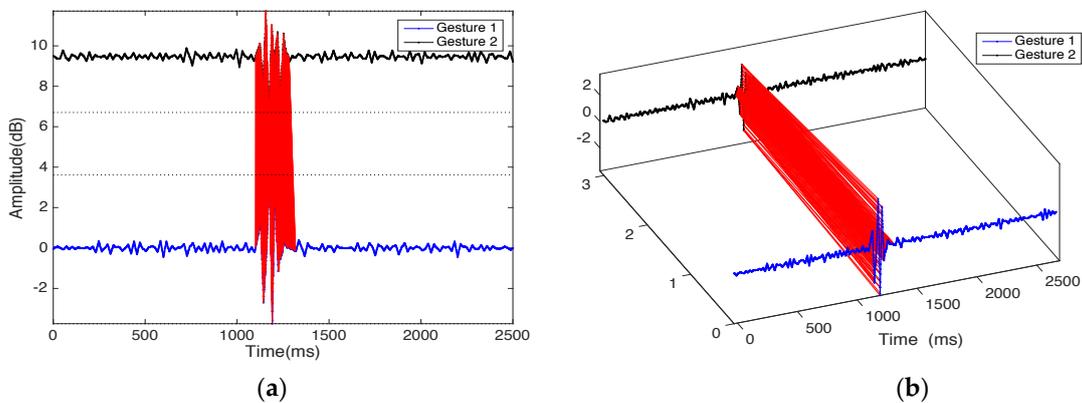


Figure 7. This Figure shows alignments between two-time series of two hand gestures by dynamic time warping (DTW): (a) 2D plot of two gestures time sequences alignments by DTW; (b) 3D plot of DTW alignments of two gestures signal.

5. Experiments and Results

In this section, we describe the implementation of our proposed system and the results we have obtained from more than two months of testing with six volunteers in different scenarios.

5.1. Experimental Set up

We use a commercial WiFi router (TL-WR842N) with two antennas as an access point (AP). For a detection point (DP), we use a Lenovo x201 laptop equipped with an IWL 5300 network card,

running a 32-bit Linux operating system (Ubuntu version 14.04), and with the open source CSI-TOOLS installed [36]. We set the packet transmission rate to 100 packets per second to obtain sufficient information from each hand gesture motion.

We set up our system in a laboratory building using five different scenarios as shown in Figure 8. The testing area is 8 m × 8 m and contains a main lab room and a small interior room. The volunteer users perform gestures for the different scenarios in the main lab room, the interior room, and the adjoining corridor, as depicted in Figure 8. The scenarios are as follows:

1. Scenario 1: The AP, DP, and user are all in the main lab room. The user performs the gestures between the AP and DP, positioned as shown in Figure 8 (label S1).
2. Scenario 2: The AP and the user are in the same room, but the DP is in the interior lab room separated by one wall. The distance between the AP and the DP is approximately 3.5 m, and the distance between the DP and the user is approximately 3 m. The user is approximately 2 m from the AP, as shown in Figure 8 (label S2).
3. Scenario 3: The DP and the user are in the interior lab room, while the AP is in the main lab room separated by one wall. The distance between the AP and the DP is 3.5 m, and the distance between the DP and the user is 2 m. The user is approximately 4 m away from the AP.
4. Scenario 4: The AP and the DP are both in the main lab room. The user is asked to perform gestures in the corridor, approximately 7 m and 5 m away from the AP and the DP, respectively. There is one intervening wall.
5. Scenario 5: The AP and the DP are both in the small interior room, while the user is in the corridor separated from them by two walls at a distance of approximately 8 m.

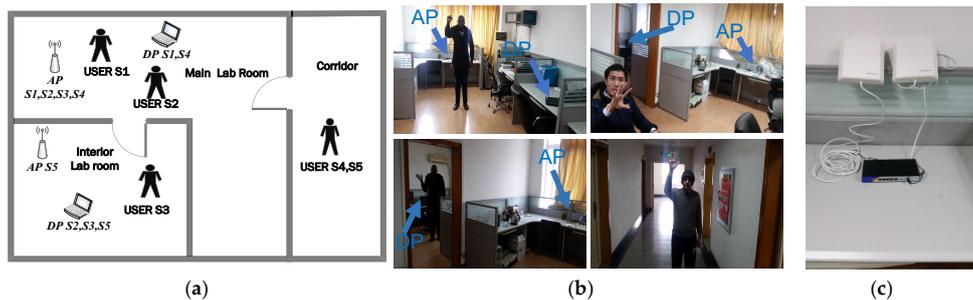


Figure 8. Experiments environment and setting. (a) Floor plan of experimental environment, setting and scenarios. S1 to S5 refer to the user's location relative to the access point (AP) and detection point (DP) in scenarios 1–5; (b) Experimental environment and settings; (c) Two WB-2400D300 antennas.

5.2. Experiment Results

We implement our proposed system in a lab with six volunteer users. For each scenario, three users are asked to perform the proposed gestures individually. We collect a total of 300 samples of each gesture from these three users for each scenario—a total of 100 samples of each gesture from each user collected over three sessions, comprising 30, 30, and 40 samples of each gesture, respectively, as shown in Table 2.

Table 2. Collected samples of each gesture in each scenario.

Session	Samples from 3 Users	Samples-per-User	Training Samples	Testing Samples
Session 1	90	30	60	30
Session 2	90	30	60	30
Session 2	120	40	80	40

We train our classifier for each session using the whole samples of each gesture collected from two different users. Then, we test the samples of each gesture that were collected from another user.

We do the test for the three users. In each test, the training samples do not include samples from the target user, and we leave the target user out for cross validation.

We evaluate the accuracy of the experimental results in terms of average accuracy per gesture for each user and the average accuracy on all the implemented gestures for each user. Generally, accuracy can be computed as shown in Equation (9):

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} \times 100 \tag{11}$$

where TP, TN, FP and FN refer to true positive, true negative, false positive and false negative results, respectively. We show the results of each scenario using a confusion matrix.

Figure 9 shows confusion matrices for different scenarios. Figure 9a shows the accuracy of scenario 1; the average accuracy is 97.28%. The accuracy scores for scenarios 2 and 3 are 91.8% and 95.5%, respectively, as shown in Figure 10b,c. As the figure shows, the gesture recognition accuracy increases when the user is closer to the DP. Figure 9d shows the accuracy of scenario 4, which decreased to 83.85% due to the extended distance and the wall between the user and the AP and DP.

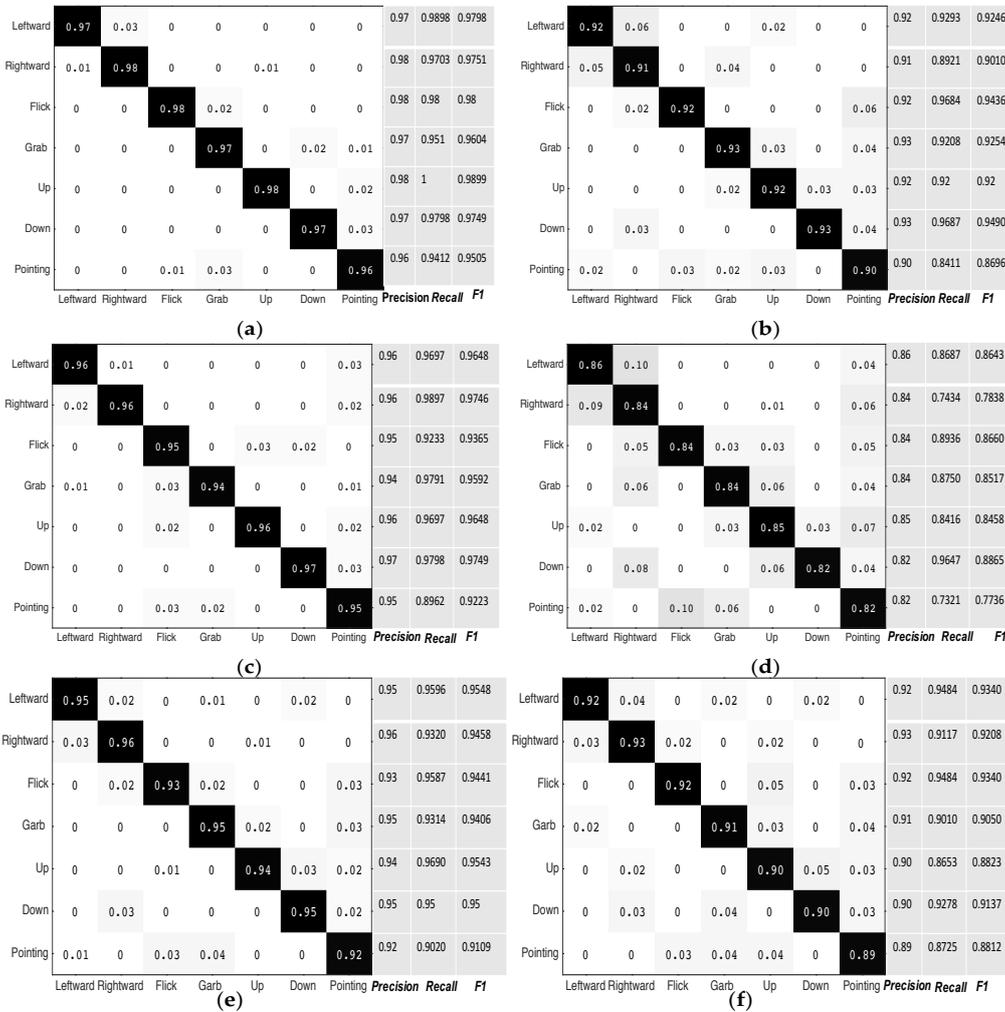


Figure 9. Confusion matrices for seven proposed interaction gestures in different scenarios. (a) Confusion matrix for scenario 1; (b) Confusion matrix for scenario 2; (c) Confusion matrix for scenario 3; (d) Confusion matrix for scenario 4; (e) Confusion matrix for scenario 4 with WB-2400D300 antennas; (f) Confusion matrix for scenario 5.

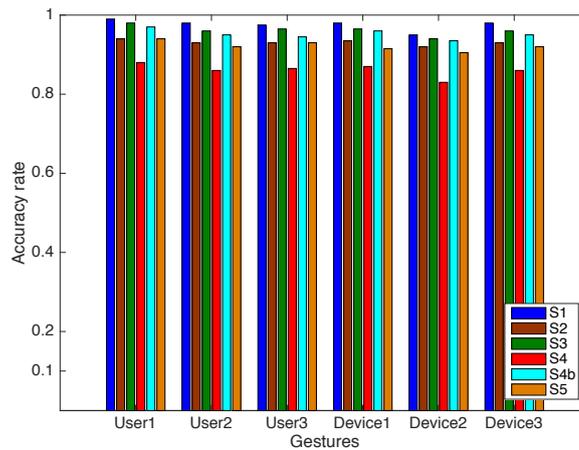


Figure 10. The recognition accuracy rates for user security and device selection gestures in all scenarios. S1–S5 denote scenario 1 to scenario 5, S4b denotes scenario 4 with the added WB-2400D300.

To overcome this problem, we used two WB-2400D300 antennas (Figure 8c). The WB-2400D300 is a commercial antenna available in [37]; it is a 300 M, 16 dB dual-polarized directional antenna that enhances the power of wireless transmission and reception and can be used with a transmitter, a receiver, or a relay station. This type of antenna can be used for long-distance wireless video monitoring with network cameras and supports remote wireless high-definition video transmission. Overall, it is an excellent and high throughput antenna that provides a high transmission rate and good stability at a low cost (the antenna cost the same as the router). We connect two WB-2400D300 antennas to the router and set up the router in the main lab room as the AP. Otherwise, the experimental set up was the same as before for scenarios 4 and 5.

After making this antenna change, the results shown in Figure 9e,f reflect the improved gesture recognition accuracy (94.4% and 91% for scenarios 4 and 5, respectively). Using a more powerful antenna helps solve the problem of recognizing gestures through multiple walls. At the same time, such antennas do not cost more than an ordinary router.

Moreover, we also use precision, recall and F-measure to evaluate experimental results. Precision is the positive predictive value, recall is the sensitivity, and F-measure or F1 score is the weighted average of both the precision and the recall. Precision, recall, and F-measure are calculated as the following expressions:

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (14)$$

From Figure 9, we see the precision, recall, and F1 values of each gesture in each scenario. The achieved results of precision, recall, and F1 confirmed the efficacy of the proposed system.

Overall, the system is more accurate at recognizing the security gestures and device selection gestures than the interaction gestures because performing the air-drawn gestures takes more time than the seven interactive gestures and, consequently, the variations in CSI are clearer. Figure 10 shows the accuracy rates for users and device gestures in all the scenarios.

We experimentally compare WiGeR with WiG [31] to verify the advantages of our approach. We choose four gestures for the comparison: two gestures from the gestures proposed for WiGeR (leftward and rightward swipes) and two gestures from the gestures proposed by WiG (push and pull). We set up the AP and the DP the same as in the LOS and NLOS scenarios described in [31]. We collect

data from one volunteer user in our lab room. For each scenario, 90 samples of each gesture are collected from this single user over 3 sessions. During each session, we collect 30 samples of each gesture. We use the methodology in WiG [31] to classify the gestures. Thereafter, we compare the results of WiG with the results of WiGeR. The comparison results show that WiGeR outperforms WiG in both LOS and NLOS scenarios, as shown in Figure 11.

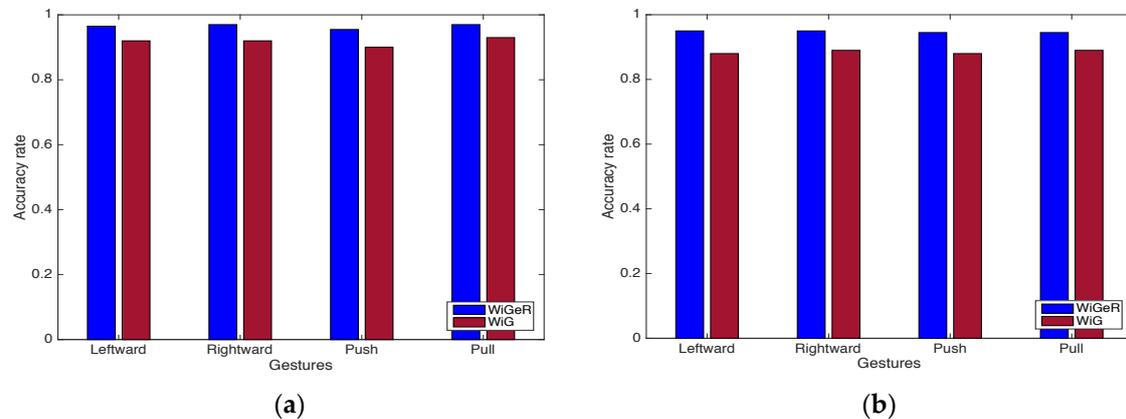


Figure 11. Classification accuracy rate of WiGeR and WiG. (a) Accuracy rate in line-of-sight (LOS) scenarios; (b) Accuracy rate in non-line-of-sight (NLOS) scenarios.

As these results show, WiGeR outperforms previous approaches in hand gesture classification accuracy and in its ability to recognize gestures through multiple walls. Moreover, WiGeR addresses the issues of security and device selection using a simple and fast mechanism.

6. Conclusions and Future Work

In this paper, we present WiGeR, a device-free gesture recognition system that works by leveraging the channel state information of WiFi signals. Our system enables humans to communicate with household appliances connected to a WiFi router. The human user can control a target device using simple hand gestures. We propose seven hand gestures intended to control various functions of several home appliances. To our knowledge, our approach is the first device-free gesture recognition system that includes both user security and device selection capabilities. We conduct more than 300 experiments for each gesture in an enclosed laboratory using complex scenarios and different volunteer users to evaluate our proposed system. The results show that our system achieves a high accuracy rate in different scenarios due to our robust pattern segmentation and classification methods. Improvements in home device hardware receiver antennas can make this system more flexible and reliable.

We plan to extend this system to classify gestures for multiple users with multiple devices all implemented concurrently. In addition, we aim to extend this work to classify the hand gestures of various human activities, which will allow users to communicate with devices in any situation. This work will lead to the development of a comprehensive device-free sensing system that can accurately classify human movements such as walking, falling, standing, lying down, performing gestures, and so on. Furthermore, we are interested in extending this approach to build a virtual keyboard by developing air-drawn gestures for all the potential letters, punctuation marks, and symbols.

Acknowledgments: This work is supported by the National Natural Science Foundation of China under grant Nos. 61373042 and 61502361 and by the Major Natural Science Foundation of Hubei Province, China under grant No. 2014CFB880.

Author Contributions: Fangmin Li proposed the idea of the paper; Fangmin Li and Mohammed Abdulaziz Aide Al-qaness conceived and designed the experiments. Mohammed Abdulaziz Aide Al-qaness performed the experiments, analyzed the data and wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

WiGeR: WiFi-based gesture recognition system

CSI: Channel State Information

AP: Access Point

DP: Detection Point

STE: Short-Time Energy

DTW: Dynamic Time Warping

LOS: Line-of-sight

NLOS: Non-line-of-sight

RSS: Received signal strength

OFDM: Orthogonal frequency-division multiplexing

SDR: Software Defined Radio

TP: True positive

FP: False positive

TN: True negative

FN: False negative

USRP: Universal Software Radio Peripheral

RFID: Radio-frequency identification

References

1. Microsoft Kinect. Available online: <http://www.microsoft.com/en-us/kinectforwindows> (accessed on 27 November 2012).
2. Jing, L.; Zhou, Y.; Cheng, Z.; Huang, T. Magic ring: A finger-worn device for multiple appliances control using static finger gestures. *Sensors* **2012**, *12*, 5775–5790.
3. Adib, F.; Katabi, D. See through walls with WiFi! In Proceedings of the the ACM SIGCOMM 2013 conference on SIGCOMM, Hong Kong, China, 12–16 August 2013; Volume 43, pp. 75–86.
4. Pu, Q.; Gupta, S.; Gollakota, S.; Patel, S. Whole-home gesture recognition using wireless signals. In Proceedings of the 19th Annual International Conference on Mobile Computing and Networking, Miami, FL, USA, 30 September–4 October 2013; pp. 27–38.
5. Yang, Z.; Zhou, Z.; Liu, Y. From RSSI to CSI: Indoor Accuracy rate localization via channel response. *ACM Comput. Surv.* **2014**, *46*, 1–33.
6. Kim, K.; Kim, J.; Choi, J.; Kim, J.; Lee, S. Depth camera-based 3D hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions. *Sensors* **2015**, *15*, 1022–1046.
7. PointGrab. Available online: <http://www.pointgrab.com/> (accessed on 6 April 2015).
8. Block, R. Toshiba Qosmio G55 Features SpursEngine, Visual Gesture Controls. Available online: <http://www.engadget.com/2008/06/14/toshiba-qosmio-g55-features-spursengine-visual-gesture-controls/> (accessed on 14 June 2008).
9. Fisher, M. Sweet Moves: Gestures and Motion-Based Controls on the Galaxy S III. Available online: <http://pocketnow.com/2012/06/28/sweet-moves-gestures-and-motion-based-controls-on-the-galaxy-s-iii> (accessed on 28 June 2012).
10. Gupta, S.; Morris, D.; Patel, S.; Tan, D. Soundwave: Using the doppler effect to sense gestures. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; pp. 1911–1914.

11. Santos, A. Pioneer's Latest Raku Navi GPS Units Take Commands from Hand Gestures. Available online: <http://www.engadget.com/2012/10/07/pioneer-raku-navi-gps-hand-gesture-controlled/> (accessed on 7 October 2012).
12. Palacios, J.M.; Sagüés, C.; Montijano, E.; Llorente, S. Human-computer interaction based on hand gestures using RGB-D sensors. *Sensors* **2013**, *13*, 11842–11860.
13. Wu, J.; Pan, G.; Zhang, D.; Qi, G.; Li, S. Gesture recognition with a 3-D accelerometer. In *Ubiquitous Intelligence and Computing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 25–38.
14. Liu, J.; Zhong, L.; Wickramasuriya, J.; Vasudevan, V. uWave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive Mob. Comput.* **2009**, *5*, 657–675.
15. Nirjon, S.; Gummeson, J.; Gelb, D.; Kim, K.H. TypingRing: A wearable ring platform for text input. In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, Florence, Italy, 19–22 May 2015; pp. 227–239.
16. Myo. Available online: <https://www.thalmic.com/en/myo/> (accessed on 6 April 2014).
17. Cohn, G.; Morris, D.; Patel, S.; Tan, D. Humantenna: Using the body as an antenna for real-time whole-body interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; pp. 1901–1910.
18. Kim, D.; Hilliges, O.; Izadi, S.; Butler, A.D.; Chen, J.; Oikonomidis, I.; Olivier, P. Digits: Freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, Cambridge, MA, USA, 7–10 October 2012; pp. 167–176.
19. Abdelnasser, H.; Youssef, M.; Harras, K.A. WiGest: A ubiquitous WiFi-based gesture recognition system. In Proceedings of 2015 IEEE Conference on Computer Communications (INFOCOM), Kowloon, Hong Kong, China, 26 April–1 May 2015; pp. 1472–1480.
20. Adib, F.; Kabelac, Z.; Katabi, D.; Miller, R.C. 3D tracking via body radio reflections. In Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation, Seattle, WA, USA, 2–4 April 2014; pp. 317–329.
21. Adib, F.; Kabelac, Z.; Katabi, D. Multi-person localization via RF body reflections. In Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation, Oakland, CA, USA, 4–6 May 2015; pp. 279–292.
22. Wang, J.; Vasisht, D.; Katabi, D. RF-IDraw: Virtual touch screen in the air using RF signals. In Proceedings of the SIGCOMM'14, Chicago, IL, USA, 17–22 August 2014; pp. 235–246.
23. Kellogg, B.; Talla, V.; Gollakota, S. Bringing gesture recognition to all devices. In Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation, Seattle, WA, USA, 2–4 April 2014; pp. 303–316.
24. Chapre, Y.; Ignjatovic, A.; Seneviratne, A.; Jha, S. CSI-MIMO: Indoor Wi-Fi fingerprinting system. In Proceedings of the IEEE 39th Conference on Local Computer Networks (LCN), Edmonton, AB, Canada, 8–11 September 2014; pp. 202–209.
25. Qian, K.; Wu, C.; Yang, Z.; Liu, Y.; Zhou, Z. PADS: Passive detection of moving targets with dynamic speed using PHY layer information. In Proceedings of THE IEEE ICPADS'14, Hsinchu, Taiwan, 16–19 December 2014; pp. 1–8.
26. Xi, W.; Zhao, J.; Li, X.Y.; Zhao, K.; Tang, S.; Liu, X.; Jiang, Z. Electronic frog eye: Counting crowd using WiFi. In Proceedings of the IEEE INFOCOM, Toronto, ON, Canada, 27 April–2 May 2014; pp. 361–369.
27. Han, C.; Wu, K.; Wang, Y.; Ni, L.M. Wifall: Device-free fall detection by wireless networks. In Proceedings of the IEEE INFOCOM, Toronto, ON, Canada, 27 April–2 May 2014; pp. 271–279.
28. Wang, Y.; Liu, J.; Chen, Y.; Gruteser, M.; Yang, J.; Liu, H. E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures. In Proceedings of the 20th annual International Conference on Mobile Computing and Networking, Maui, HI, USA, 7–11 September 2014; pp. 617–628.
29. Wang, G.; Zou, Y.; Zhou, Z.; Wu, K.; Ni, L.M. We can hear you with Wi-Fi! In Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, Maui, HI, USA, 7–11 September 2014; pp. 593–604.
30. Nandakumar, R.; Kellogg, B.; Gollakota, S. Wi-Fi Gesture Recognition on Existing Devices. Available online: <http://arxiv.org/abs/1411.5394> (accessed on 8 June 2016).

31. He, W.; Wu, K.; Zou, Y.; Ming, Z. WiG: WiFi-based gesture recognition system. In Proceedings of the 24th International Conference on Computer Communication and Networks (ICCCN), Las Vegas, NV, USA, 3–6 August 2015; pp. 1–7.
32. Goldsmith, A. *Wireless Communications*; Cambridge University Press: Cambridge, UK, 2005.
33. Rabiner, L.R.; Schafer, R.W. Introduction to digital speech processing. *Found. Trends Signal Process.* **2007**, *1*, 1–194.
34. Salvador, S.; Chan, P. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **2007**, *11*, 561–580.
35. Jang, J.S.R. Machine Learning Toolbox Software. Available online: <http://mirlab.org/jang/matlab/toolbox/machineLearning> (accessed on 23 December 2014).
36. Halperin, D.; Hu, W.; Sheth, A.; Wetherall, D. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM Comput. Commun. Rev.* **2011**, *41*, 53–53.
37. WB-2400D300 Antenna. Available online: https://item.taobao.com/item.htm?spm=a1z09.2.0.0.NXy0fP&id=39662476575&_u=d1kraq3mf31b (accessed on 9 June 2014).



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).