


Towards Location Independent Gesture Recognition with Commodity WiFi Devices

Yong Lu , Shaohe Lv and Xiaodong Wang *

Science and Technology on Parallel and Distributed Processing Laboratory, National University of Defense Technology, Changsha 410073, China; yonglu@nudt.edu.cn (Y.L.); shaohelv@nudt.edu.cn (S.L.)

* Correspondence: xdwang@nudt.edu.cn

Received: 10 September 2019; Accepted: 18 September 2019; Published: 20 September 2019



Abstract: Recently, WiFi-based gesture recognition has attracted increasing attention. Due to the sensitivity of WiFi signals to environments, an activity recognition model trained at a specific place can hardly work well for other places. To tackle this challenge, we propose WiHand, a location independent gesture recognition system based on commodity WiFi devices. Leveraging the low rank and sparse decomposition, WiHand separates gesture signal from background information, thus making it resilient to location variation. Extensive evaluations showed that WiHand can achieve an average accuracy of 93% for various locations. In addition, WiHand works well under through the wall scenario.

Keywords: gesture recognition; WiFi sensing; channel state information

1. Introduction

With the rapid development of Human–Computer Interaction (HCI) technologies, gesture recognition is gaining increasing attentions. During the past years, gesture recognition has been widely deployed in many areas including tracking and detection [1–3], pattern understanding [4,5], health fitting [6] and other industrial or academic areas [7]. Unlike conventional HCI methods such as keyboards, mice or touch screens, gesture recognition can deal with human–computer interaction remotely with no additional devices. Specifically, gesture recognition is very suitable for future applications such as smart home and virtual reality. Numerous applications [8–12] have been proposed for smart home and virtual reality. It is more convenient and relaxing for people to interact with machines using gestures, thus making gestures a great potential for the future HCI technology.

During the past decades, many efforts have been devoted to the development of gesture recognition. Vision-based methods [13–19] try to deploy image processing technologies to process gesture image sequences captured by the video. Despite its high accuracy, vision-based methods are limited to the light conditions and computation costs. Inferred-based technologies can achieve high precision recognition accuracy without the restriction of the light condition. Applications using the inferred technologies, for example, the Microsoft Kinect [20] and Leap motion [21], both achieve great success. Sensor-based technologies deploy various kinds of special sensors such as accelerometers to capture an individual’s gesture. For example, TEXIVE [22] employs intelligent sensors to detect driving activities. This technology can recognize fine-grained activities but its installation is inconvenient and the cost is high.

Recently, WiFi-based gesture recognition has attracted a lot of attention for its ubiquitousness and effectiveness. In 2013, Pu proposed Wisee [23] which uses the doppler shift resulted from gesture as the feature to recognize nine gestures. However, Wisee uses customized software defined radio devices which cannot be deployed directly on existing WiFi devices. WiGest [24] uses the received signal strength (RSS) value to detect five gestures, which are defined as five different operations of

the computer. Since RSS only reports one value a time, it can hardly achieve fine-grained recognition. In addition, its detection range is limited. WiG [25], WiFinger (UbiComp 2016) [26], and WiFinger (MobiHoc 2016) [27] all focus on utilizing the channel state information (CSI) feature to recognize gesture. They use different features to recognize the gestures. However, their detection range is limited. Mudra [28] takes advantages of the multiple antennas and proceeds signal elimination, thus making it possible to sense subtle actions without training. However, Mudra uses two antennas and the distance between two antennas should be larger than 10 cm, thus making it hard to deploy on commodity devices, especially on mobile devices.

Compared to conventional gesture recognition methods, WiFi-based gesture recognition has many advantages. *Device-free*: There is no need to carry any devices or sensors when dealing with gesture recognition base on WiFi signals. Once WiFi signals exist, you can utilize it to recognize gestures. *None-line-of-sight (NLOS) recognition*: As WiFi signals can propagate through walls and obstacles, it is obvious that WiFi based gesture recognition can still be effective even under NLOS condition. *No restrictions on light conditions*: Unlike vision-based methods, WiFi-based gesture recognition can be effective under different light conditions even in dark environment. *Easy to deploy*: According to Cisco's latest report [29], there will be 542.6 million WiFi access points in 2021. We can deploy WiFi based gesture recognition easily on existing WiFi devices.

While state-of-the-art systems obtain reasonable performance in their given scenario, there is a key limitation considering the existing gesture recognition systems based on commercial WiFi devices. As we all know, although different methods are used in these systems, they follow similar procedures: collecting the wireless signal data between the transmitter and receiver and analyzing the information it contains, then we can infer the actions of the person. However, the wireless signal we collected not only contains the action information but also the unique features of the environment where the action is performed. As a matter of fact, the action recognition model trained on a specific environment can hardly perform well on actions collected in a different environment.

To address this challenge, we propose WiHand, a location independent gesture recognition system based on commodity WiFi devices. To resist the influence of environment variation, we try to separate the gesture signal from the background information, thus minimizing the influence of the environment. The key component of WiHand is the low rank and sparse decomposition (LRSD) algorithm. The most remarkable characteristic of this algorithm is that it can separate the background information and the noise. It has been widely used in the image processing area [30,31]. We have to clarify that our methods are quite different from those in [32,33]. Jiang et al. [32] used deep learning based methods to extract environment independent features. Virmani and Shahzad [33] used a translation function that can generate virtual samples for different positions to realize position agnostic recognition. They are both labor intensive and energy consuming, thus making it difficult to deploy on mobile devices. WiHand uses the LRSD algorithm to extract location independent gesture signal. With general classifiers, we can accomplish the recognition task.

To implement WiHand, we encountered several technical challenges. The first one is that the signal we collected is continuous, in which contains the gesture segments. We have to detect the boundary of the gesture signal and extract it for the followup procedures. The second one is that the subcarriers of the WiFi channel encounter different fading schemes in different environments. As we can learn from Section 3, there are 168 subcarriers in WiHand system. There is no need to input all of them into the classifier, since they all represent the same gesture. What we have to do is to choose the most affected subcarrier for recognition during environment variation. The third challenge is to make WiHand resilient to the environment change. We try to separate the gesture signal from the background information to make it stable under various environments.

There are three main contributions in this paper. Firstly, we propose a deviation based gesture signal extraction algorithm utilizing the deviation changes of different subcarriers. Second, we propose an adaptive subcarrier selection algorithm, which can automatically choose the most affected subcarriers for sensing. Third, We propose WiHand, a location independent gesture recognition system

based on commodity WiFi signals. Taking advantage of the low rank and sparse decomposition algorithm, we separate the gesture signal from the background information.

We built a proof-of-concept prototype of WiHand on commodity WiFi devices including two desktop computers equipped with TP-Link TL-WDN 4800 wireless NIC. Extensive evaluations showed that WiHand can achieve fine-grained gesture recognition under various environments without retraining. Even under through the wall scenario, WiHand shows satisfying performance.

The rest of this paper is organized as follows. Section 2 introduces some gesture recognition related technologies. Section 3 introduces some preliminaries of WiFi sensing and makes detailed instructions on our problem. Section 4 describes the design details of the WiHand system. Section 5 presents the extensive evaluations of WiHand under various scenarios. We conclude the paper in Section 6.

2. Related Work

Radio-based Activity Recognition. Radio-based activity recognition is a recently developed recognition technology relying on the development of signal processing technology. Based on the extraction of the characteristic pattern of the multi-path superimposed wireless signals generated by specific activities, activity recognition can be achieved through identifying and interpreting these patterns. Many studies focus on the recognition of activities. Scholars of Rutgers University proposed E-eyes [34], which tries to conduct the recognition of a continuous sequence of actions. WiZ [35] was proposed by Dina Katabi at MIT. It uses the multi-antenna processing technology to enable the concurrent recognition of two or more users' actions. Sigg conducted a research of multi-user concurrent actions recognition technology based on K-nearest neighboring method [36]. Zhu [37] proposed a novel scheme for robust device-free through the wall detection of a moving human with commodity devices.

Gesture Features Extraction Methods. Most of the existing gesture recognition approaches share the same scheme of extracting gesture features and training models for recognition. Sigg [36,38] used statistic features for classification, i.e., mean value, variance, kurtosis, skewness and so on. Besides, some works [39] try to utilize convolutional neural network (CNN) network to extract hyper-level feature from the original signal. In addition, there are many different feature extraction methods in areas of gesture tracking and understanding fields [40–46].

3. Basic Idea

The key insight of WiFi based sensing is to analyze the signal fluctuations caused by human actions. However, WiFi signals are not only affected by the human body, but also the surrounding environments while propagating across space. In consequence, the model trained by data collected in a specific environment can hardly work well in a new environment. In this section, we show how the environments affect the signal and introduce our basic idea.

3.1. Channel State Information

We take advantage of the most widely used Channel State Information (CSI) to collect gesture data. CSI indicates the channel link states in wireless multiple input multiple output (MIMO) systems. We can get CSI data from commodity devices using customized hardware drivers. CSI is sensitive to the variations of the channel link. Compared to other signals, for example, USRP or RSS, CSI is fine-grained and has a relatively small size. By analyzing the CSI value, we can infer the condition of the link and further infer the corresponding actions. The CSI can be formulated as Equation (1) in frequency domain. It is also called Channel Frequency Response (CFR).

$$H(f; t) = \sum_n^N a_n(t) e^{-j2\pi f \tau_n(t)} \quad (1)$$

where $a_n(t)$ is the amplitude attenuation factor, $\tau_n(t)$ is the propagation delay and f is the carrier frequency. CSI captures the characteristic of the surrounding environment including the furniture and human movements. In WiHand, we have one transmitting antenna and three receiving antennas. Each has 56 subcarriers. Then, the CSI we receive is a three-dimensional matrix. For each package, we can get a CSI matrix, as shown below:

$$H_{i,j} = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} & \dots & h_{1,56} \\ h_{2,1} & h_{2,2} & h_{2,3} & \dots & h_{2,56} \\ h_{3,1} & h_{3,2} & h_{3,3} & \dots & h_{3,56} \end{bmatrix} \quad (2)$$

3.2. Problem Statement

The signal collected at the receiver side contains information from the environment. Change of location means the change of corresponding surroundings, which makes great difference in the propagation of signals. The effect of reflection, diffraction, multipath and so on will be changed during the variation of location, thus making the received signal different from the original one. As shown in Figure 1a, changes of location can result in totally different waveform for the same gesture.

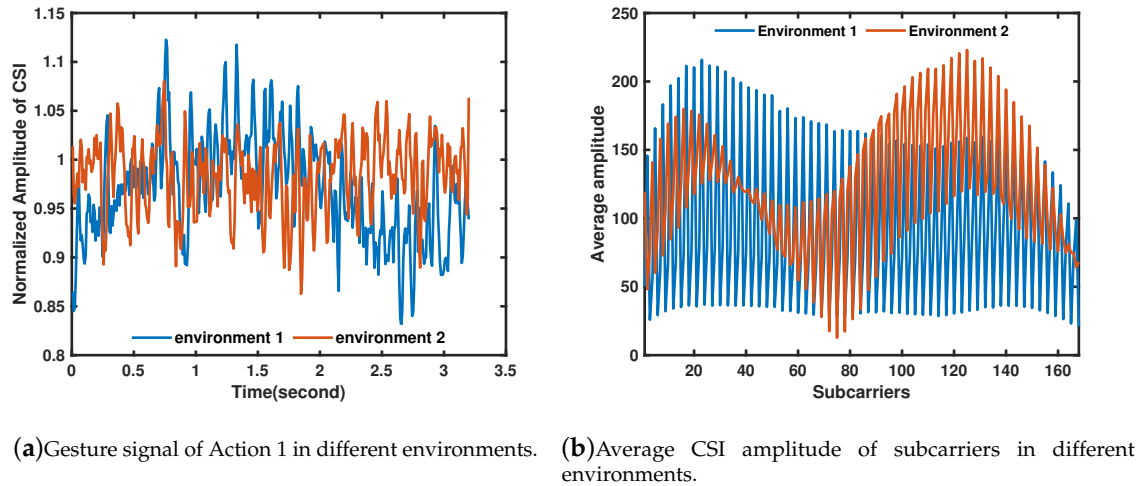


Figure 1. The influence of different environments on the CSI.

To realize location independent recognition, we first have to understand how the environment affects the gesture signal. Overall, the received signal may encounter changes from two aspects. The one is that the received signal may encounter different fading schemes on different subcarriers. As shown in Figure 1b, the average amplitude of CSI on each subcarriers changes with the location variation. That is the extent of how each subcarrier is affected changes with the variation of location. The other one is that the superposition of new multipath or reflection path may result in total different waveform, which is very different from the original location. As shown in Figure 1a, the same action may result in different waveforms under different environments. To achieve location independent recognition, we need to first select the subcarrier which is affected severely by the gestures and then extract features which are less affected by the environment.

4. Design Details of WiHand

4.1. Overview

WiHand harnesses the commodity WiFi devices to realize location independent fine-grained hand gesture recognition. The five gestures we chose are shown in Figure 2. They are all hand gestures that

are repeatedly used in many scenarios. We performed the gestures between the transmitter and the receiver and recorded the corresponding CSI data streams.

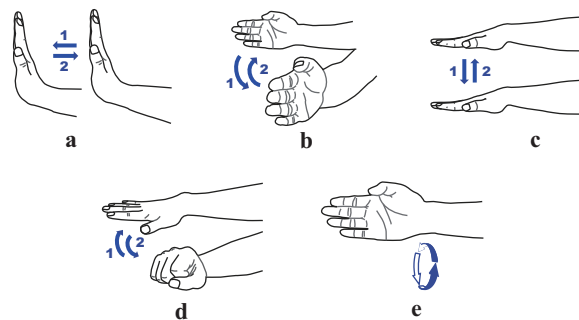


Figure 2. The five hand gestures: (a) pushing and pulling (PP), firstly pushing your hand to the front and then pull it back; (b) waving left and right (WLR), firstly waving your hand to the left and then to the right; (c) waving up and down (WUD), firstly waving your hand up and then down; (d) stretching each finger (SF), pushing your hand and at the same time stretching each finger; and (e) circling clockwise (CC), using your hand to draw circle with a clockwise manner.

We can further get an overview of the WiHand system in Figure 3. According to the signal processing procedure, we can divide the WiHand system into four parts.

- **Preprocessing.** The CSI data collected from the commodity devices contain random noise. We should filter out the signals that are out of the frequency range of hand gestures. In addition, the time interval of the packages is not strictly equal. We should use interpolation to make the CSI stream more related to the gestures.
- **Gesture Detection.** The CSI streams are continuous; we should detect if there is a gesture being acted and at the same time extract the gesture related segments generating corresponding gesture profiles. With the standard deviation based algorithm, we can easily detect the appearance of gesture signal.
- **Feature Extraction.** To recognize the gestures, we need to first use the binned entropy based subcarrier selection algorithm to find the most affected subcarrier. Then, we take advantage of the low rank and sparse decomposition (LRSD) algorithm to extract location independent gestures signal from the original signal. Then, we extract features from the signal using the histogram.
- **Classification.** We use a one class Support Vector Machine (SVM) classifier to recognize the gestures. Firstly, collected gesture data are used to train a SVM machine. Then, the SVM machine is used to recognize the newly arrived gestures.

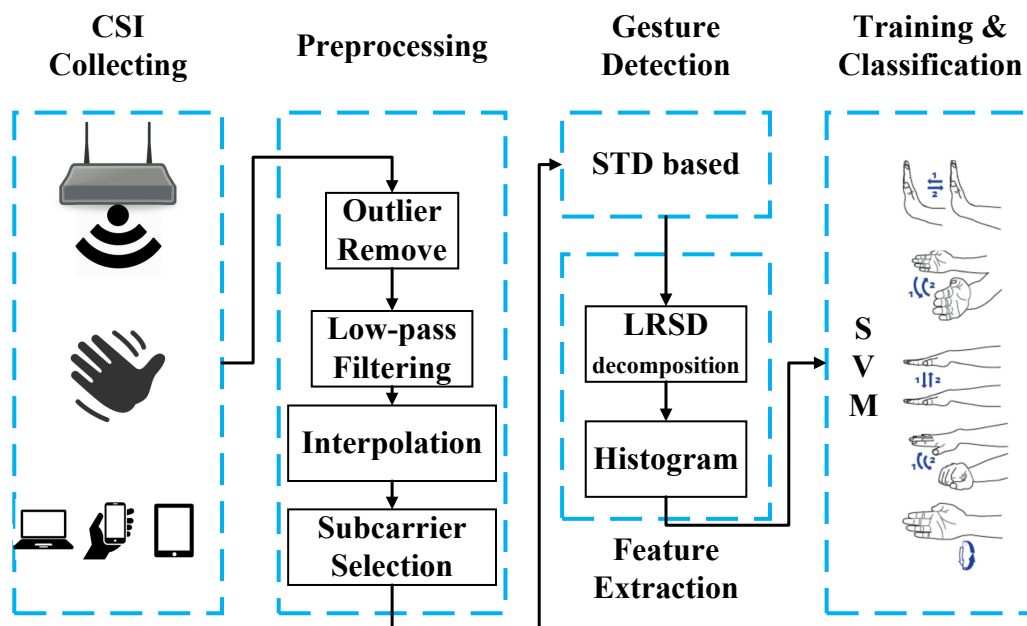


Figure 3. Outline of the WiHand system.

4.2. Preprocessing

Raw CSI streams contain noise and outliers which may seriously affect the recognition results. By preprocessing, we try to eliminate all the noises and outliers that are not related to gestures.

There are mainly two kinds of noise in the CSI stream. One is caused by hardware errors during transmitting and receiving, such as cyclic shift diversity (CSD), sampling time offset (STO), sampling frequency offset (SFO) and beamforming. This kind of noise mainly causes phase offsets and introduces high frequency components to the CSI. Firstly, we use the amplitude of the CSI stream as the gesture signal and eliminate the phase offsets. Secondly, we apply the low-pass filter to eliminate high frequency components. The detailed instruction of how the low-pass filter is designed can be found in Section 4.2.2.

The other noise is caused by the furniture or other moving persons. Here, we only consider interferences from the other rooms. The scenario of multiple persons in the same room is out of the scope of this paper. According to Zeng et al. [47], we can calculate the delay profile of multi-paths. Firstly, we convert the CSI to channel impulse response (CIR) and get the power delay profile (PDP) of multi-paths. Then, we remove the multi-path components with delay more than 0.5 millisecond. This threshold is chosen based on previous research of [48]. Afterwards, we can get the CSI without interferences by converting the CIR back to CSI with fast Fourier transform (FFT).

Apart from the above processing procedures, there are still several steps to perform to remove the noise of the CSI stream. Next, we look deep into the process of preprocessing and give details instructions on each of the step.

4.2.1. Outliers Removal

The original CSI data contain many spikes which are caused by many reasons. The spikes heavily affect the features extracted from the raw CSI signals. Thus, we have to remove the spikes from the raw CSI signals.

We use the Hampel identifier algorithm proposed in [49] to remove all the abnormal points. Here, we define all the points that out of the range of Equation (3) as outliers.

$$[\mu - \gamma \times \sigma, \mu + \gamma \times \sigma] \quad (3)$$

where μ is the mean value of the signal, σ is the standard variance, and γ is the removing factor. The value of γ can be various in different scenarios. According to Li et al. [26], we set the value of γ to be 3, which is the most widely used one in similar works.

As shown in Figure 4b, the original CSI streams contain many spikes. With outlier removal, we can remove all the outliers and get a relatively clear CSI stream.

4.2.2. Low-Pass Filtering

The frequency of the hand gestures is relatively low. The collected CSI contains high frequency components which affects the accuracy of detection. We need to remove such noise components with low-pass filter. Considering the frequency range of the movements, there is currently no uniform standard. Table 1 shows the frequency range adopted in other research works. Referring to these values, with experiments, we chose the Butterworth low-pass filter to remove frequency higher than 60 Hz, which could get the best performance.

Table 1. Frequency range of common activities.

System	Activities	Frequency Range (Hz)
CARM [50]	body movements	0.15~300
WiSee [23]	hand gestures	8~134
WiFinger [27]	hand gestures	0.2~5
WiFinger [26]	ASL gestures	1~60
WiWho [47]	gaits	0.3~2
WiFi-ID [51]	walking behavior	20 ~80
UniBreathe [52]	respiration	0.1~0.5
WiHear [53]	mouth	2~5
Zeng [54]	shopping behavior	0.3~2

With low-pass filtering, we can eliminate much of the noise and get CSI streams that are more related to the gestures. The result of low-pass filtering is shown in Figure 4c. It is obvious that, with low-pass filtering, we can get a smoother CSI stream.

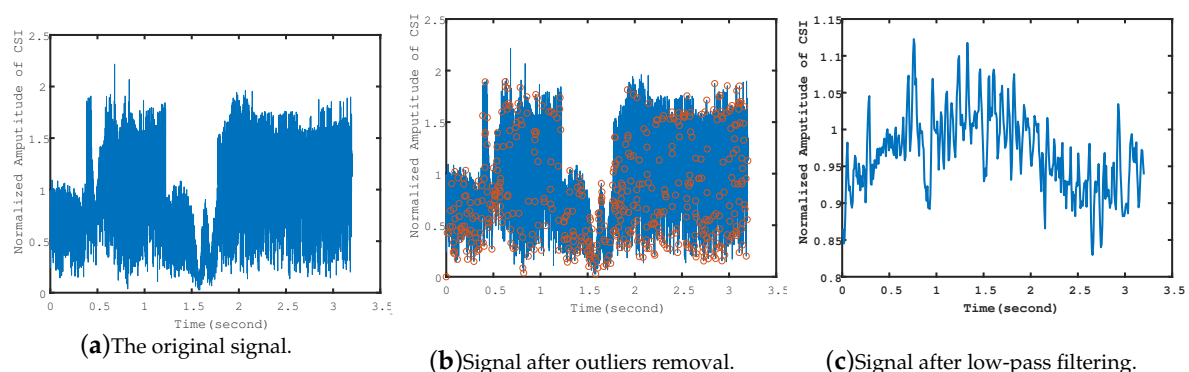


Figure 4. The results of preprocessing.

4.2.3. Interpolation

According to our observations, the CSI packages we received are not with equal time interval. Transmitting delay, receiving delay and decoding delay all heavily affect the CSI time. Besides, due to

package loss, the length of CSI segments varies with time. Figure 5 shows the cumulative distribution function (CDF) of the time interval between CSI packages. In the figure, we can see that about 90% of the intervals are equal. The rest are with randomized value. That is, for the same gestures, we may receive CSI with different length.

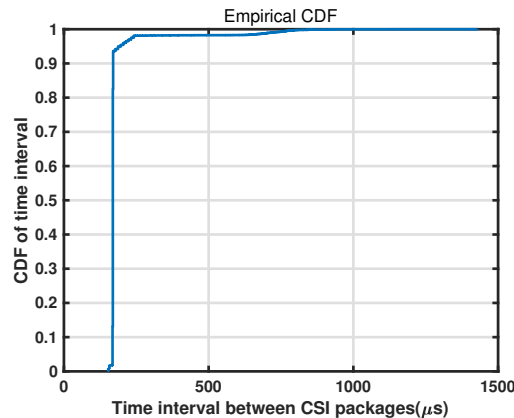


Figure 5. The CDF of time intervals for the received CSI packages.

We need to make the interval equal to make the CSI streams represent hand gestures well. To realize this procedure, we add a time stamp to every transmitted CSI package. We propose a time-dependent interpolation (TDI) algorithm which can smartly interpolate points according to the time stamp and the preset sampling rates.

The TDI algorithm is shown in Algorithm 1. For every received packet, we first calculate the time interval between the current packet and the previous one. We set the window size of the detection to be 0.2 s. Based on the intervals between packages, we can evaluate the actual sampling rate. Then, we make the resampling according to the standard sampling rate. The resampling procedure contains the interpolation and down-sampling operations. The practical coefficients are determined by the real time interval and the sampling rate.

Algorithm 1: Time-dependent Interpolation (TDI) Algorithm.

Data: Input: $sampleRate = 2500$

```

1 Set the sampling rate;
2 while 1 do
3    $CSI = readCSI();$ 
4   Reading CSI Matrix from the drivers;
5    $csiTime(i) = CSI(i).timeStamp;$ 
6    $Interval = mean(diff(csiTime));$ 
7    $newCSI = RESAMPLE(CSI, sampleRate, Interval);$ 
8 end
```

4.3. Deviation-Based Gesture Boundary Detection

After preprocessing, we already get smoothed CSI that contains the gesture information. However, as the CSI is continuously collected, we need to further segment the CSI streams to extract the segments that contain exactly the gesture information. According to our observations, the main evidence of gesture appearance is that the CSI signal becomes fluctuant.

We tried to calculate the average standard deviation of all the subcarriers to figure out where the standard deviation can reflect the emergence of gestures. Figure 6 shows that the standard deviation changes with the subcarrier CSI. The top figure is the preprocessed CSI streams of all 168 subcarriers. The bottom figure is the average standard deviation of all the 168 CSI streams. We can learn that

the standard deviation between the subcarriers are steady when there are no gestures. Once we do gestures, the standard variance starts fluctuating.

According to this observation, we firstly calculate the standard deviation of all the subcarriers and get the average standard deviation. Then, we set a threshold of the standard deviation. With a threshold, we can figure out when the gestures start and end. However, the problem is that there are values under the threshold within the gesture signal. We further calculate the gradient of the standard deviation and set a threshold to extract the gesture signal.

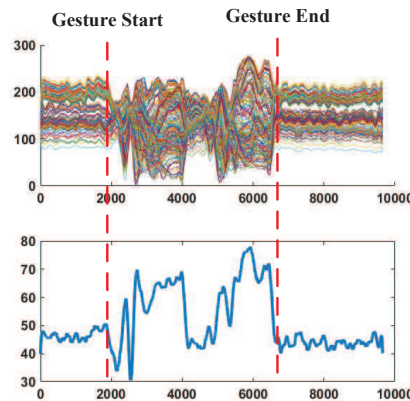


Figure 6. The average standard deviation of the gesture signal.

4.4. Binned Entropy Based Subcarrier Selection

Due to the location variation, subcarrier may encounter different fading situation. That is to say, the extent of how the gestures affect different subcarrier changes with the location variation. We need to select subcarriers whose signal are highly related to the gestures. Firstly, we need to define a metric which can measure the extent of the impact of gestures on the subcarriers.

We propose a binned entropy based adaptive subcarrier selection algorithm. The binned entropy is defined as follows:

$$\text{binned_entropy}(X) = - \sum_{k=0}^{\min(\text{maxbin}, \text{len}(X))} p_k \ln(p_k) \quad (4)$$

where p_k represents the probability that the value of X_T is within the k th bin. maxbin means the number of bins. $\text{len}(X_T) = T$ stands for the length of series X_T .

The value of binned entropy represents the information contained in the signal series. If the value is large, the signal contains lots of information. Otherwise, it contains less information. In our scenario, the value of binned entropy stands for the importance of the subcarrier. We adaptively calculate the binned entropy of each subcarrier, and then choose the subcarriers with larger entropy.

Firstly, we calculate the binned entropy of each subcarrier. The function is defined as Equation (4). It can measure the affection of gestures to each subcarrier very well. When the binned entropy is large, the subcarrier is severely affected by the gestures. Through this algorithm, we can select the subcarriers that are most affected by the gestures, thus making it possible to improve the detection accuracy.

4.5. Feature Extraction and Classification

With the above steps, we already get smooth CSI segments that contain gesture information. Most of the existing works try to extract features directly from such CSI segments using statistics features or frequency domain features [55–61].

However, for the location variant scenario, one has to resist to the influence resulted from the location change. As mentioned in the Introduction, there are mainly two kinds of consequences caused by the location change. One is the subcarrier fading mode change, the other is signal waveform change. For the subcarrier fading variant, we propose an adaptive subcarrier selection algorithm which can

always choosing the most affected subcarriers from all 168 subcarriers. For the waveform change, we use low rank and sparse decomposition algorithm to separate the gesture signal from the background one, thus making it possible to get relatively stable waveform for each gesture. Below, we introduce how to separate gesture signal from the background information in detail.

4.6. Gesture Feature Extraction

We obtain from the above procedures a smooth CSI segments containing gesture information. In the conventional case, we can extract features from the segments directly. Under the scenario of location variation, if we extract features directly from the processed segments, it may perform well within one specific place. However, it cannot perform well in a different place. It is the information of the environment that contains in the signal affects the performance of recognition under various locations. To resist the location variation influence, we propose to separate the gesture signal from the background information.

Assume X is the original signal; we can separate X into two matrix A and E , i.e., $X = A + E$. Here, A is a matrix with low rank, while E is a matrix with sparse values. This is called the low rank and sparse decomposition (LRSD). To realize the decomposition, we need to find the A and E .

We can formulate this into Equation (5).

$$\min \|A\|_* + \lambda \|E\|_1 \quad (5)$$

$$\text{s.t.} \quad X = A + E \quad (6)$$

where $\|\cdot\|_1$ is L1 norm, which leads to being sparse. $\|\cdot\|_*$ is the nuclear norm, which is the L1 norm of the singular values. Minimizing this value will result in sparse singular values, which means low rank.

We use the augmented Lagrangian algorithm to solve the above problem. The augmented Lagrangian formulation of Equation (5) is shown as follows.

$$L_\mu(A, E, Y) = \|A\|_* + \lambda \|E\|_1 + \langle X - A - E, Y \rangle + \frac{\mu}{2} \|X - A - E\|_F^2 \quad (7)$$

The iteration values of A and E are shown as below.

$$A^* = \operatorname{argmin}_A L_\mu(A, E, Y) \quad (8)$$

$$= \operatorname{argmin}_A \|A\|_* + \langle X - A - E, Y \rangle + \frac{\mu}{2} \|X - A - E\|_F^2 \quad (9)$$

$$= D_{\mu^{-1}}\{X - E + \mu^{-1}Y\} \quad (10)$$

$$E^* = \operatorname{argmin}_E L_\mu(A, E, Y) \quad (11)$$

$$= \operatorname{argmin}_E \lambda \|E\|_1 + \langle X - A - E, Y \rangle + \frac{\mu}{2} \|X - A - E\|_F^2 \quad (12)$$

$$= S_{\lambda\mu^{-1}}\{X - A + \mu^{-1}Y\} \quad (13)$$

With the LRSD algorithm, we can separate the original signal into two parts. The low rank part represent the background information. The sparse part stands for the noise.

$$CSI_{original} = CSI_{background} + CSI_{noise} \quad (14)$$

As shown in the above equation, the gesture is contained in the CSI_{noise} part as, in the original WiFi signal, the gesture is a kind of noise. Further, we deploy PCA algorithm to get the pure gesture signal.

4.6.1. Feature Extraction

The last step is to extract features from the extracted gesture signals. Here, we choose the histogram of the gesture CSI profile as the feature. The histogram is easy to compute and it has strong ability to resist noise. To calculate the histogram of the gesture, we need to set the number of features. According to the CSI amplitude, we calculate the number of points in each segments.

Figure 7 shows the extracted features of gesture PP from the processed gesture signal. Figure 7a is the gesture signal after all the processing. Figure 7b is the features we finally get. It is a vector of the probability of CSI value distribution.

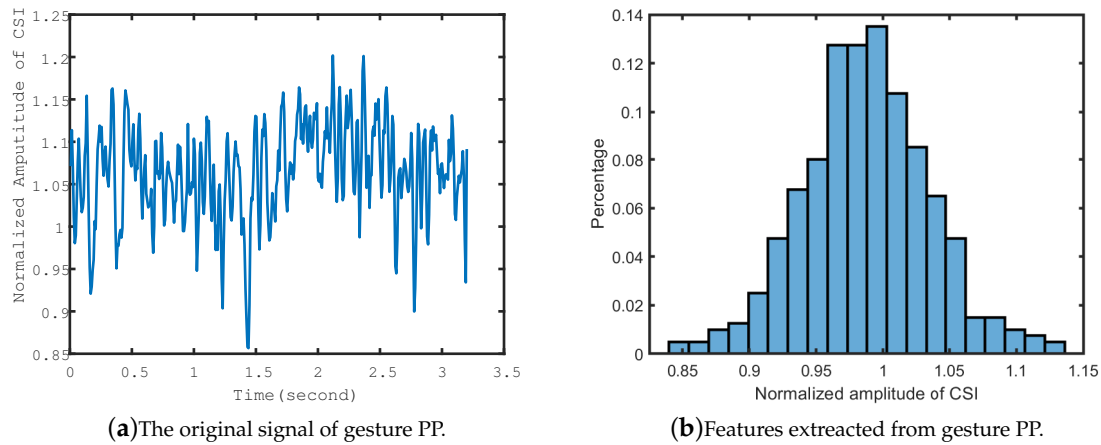


Figure 7. The features of gesture PP, extracted from the original signal.

4.6.2. Classification

As for classification, we use one class support vector machine (SVM) based methods to deal with the recognition of different gestures. The SVM model is shown as follows:

$$\min \frac{\|\omega\|^2}{2} \quad (15)$$

subject to $y_i(\omega^T \varphi(x_i) + b) \geq 1, i = 1, 2, \dots, n$. Here, x_i is the i th sample in the hyperplane. y_i is the value of the i th sample. $\varphi(x)$ is the kernel function. (ω, b) represents the hyperplane.

Then, we can get the Lagrangian function of it:

$$L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^n \alpha_i (y_i (\omega^T \varphi(x_i) + b) - 1) \quad (16)$$

Then, we can calculate the partial derivatives of equation

$$L(\omega, b, \alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \varphi(x_i^T) \varphi(x_j) \quad (17)$$

Finally, we get an optimization problem about α . We can calculate ω, b to construct the hyperplane and then separate the data by:

$$f(x) = \text{sign}(\omega^T \varphi(x) + b) \quad (18)$$

For the training phase, we input the collected gestures feature profile into an SVM machine. For the classification phase, we input the newly collected CSI into the trained SVM machine and get the classification results. We choose the radial basis function (RBF) kernel in our algorithm. We use the open source machine learning library libsvm [62] for classification.

4.7. Discussions

When Wi-Fi signals propagate in the space, it can be affected by many factors. The nature of dynamics makes it very hard to achieve persistent high detection accuracy with Wi-Fi signals. Despite the numerous methods proposed by researchers, there remains some challenges considering WiFi based gesture recognition.

Multi-person scenario. When more than one person is in the same room, actions from different persons will interfere with each other, thus decreasing the accuracy or even resulting in recognition failure. However, considering the deployment environment of WiFi, there will always be more than one person inside the room. How to separate different person or to detect multiple person simultaneously is still a challenge.

Interpersonal difference. Although we can define the pattern of different gestures, people perform them differently. In addition, due to interpersonal differences, for example, the elderly may perform gestures more slowly, we may encounter accuracy reduction when facing many people.

5. Performance Evaluation

We extensively evaluated WiHand. By different scenarios, we tried to find the factors that affect the recognition results and to see whether WiHand can resist the location variation.

5.1. Experimental Setup

We used two computers to act as AP and client. Each is with a TP-LINK TL-WDN 4800 wireless network interface card (NIC). This NIC was equipped with the Atheros AR9380 chip. It is compatible with the ath9k driver. Thus, we installed the customized driver from [63] to extract the CSI streams. The transmission was with one Transmitting antenna and three receiving antennas. We could extract CSI streams of all 56 subcarriers from the customized drivers. For every packet, we could get a CSI matrix with size of $3 \times 1 \times 56$.

As shown in Figure 8, the transmitter and receiver were put on each side of the room close to the wall. The size of the room is $3 \text{ m} \times 4.5 \text{ m}$. We performed all gestures at the different locations shown in Figure 8. The exact coordinates of the nine testing positions are shown in Table 2. The corner around Point 6 was set as the origin of the coordinate.

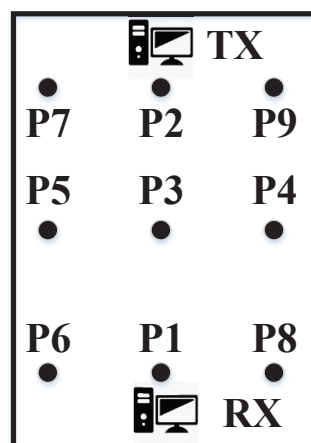


Figure 8. The layout of the transmitter, receiver and the nine testing positions.

Table 2. Location of the testing points. The units of the coordinate are meters.

P1	P2	P3	P4	P5	P6	P7	P8	P9
(1.5, 0.5)	(1.5, 4)	(1.5, 2.25)	(2.5, 2.25)	(0.5, 2.25)	(0.5, 0.5)	(0.5, 4)	(2.5, 0.5)	(2.5, 4)

As for software in our system, we ran the Linux 2014 LTS edition to collect and analyze data. All evaluations were based on the MATLAB 2018b platform.

We collected gesture data from 26 volunteers, among whom six are female and the others are male. Two of the males are teachers aged 35 and 43. The others are students aged from 20 to 28. We trained all the volunteers for all the gestures before the data collection to make them familiar with the gestures. Each user made the five gestures repeatedly for 100 times at each of the nine locations and the through the wall scenario. Then, we used the collected data to test the system. Unless otherwise specified, we repeated the detection for 100 times and calculated the average accuracy.

5.2. Performance of Gesture Detection

As for the gesture detection rate, we wanted to make sure our algorithm could detect the gesture signal and extract it correctly. We chose 10 people and asked them to perform all five gestures continuously for 50 times. Then, we calculated the detection rate of gestures for each person.

Figure 9 shows the detection rate of gesture detection. WiHand could detect the gesture of different persons with an average accuracy of 91%. We also found that the detection rate differed among the volunteers. That is because the performance varied among the volunteers. Despite the same motion pattern, everyone had their personal characteristics considering the frequency, range and quality of the movements.

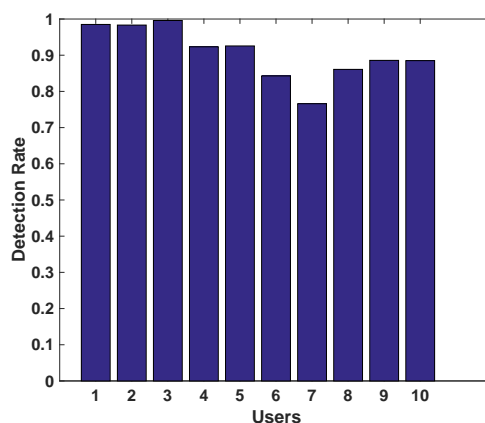


Figure 9. The performance of gesture detection.

5.3. Performance of Gesture Recognition Accuracy at Fixed Location

We performed all the gestures at a fixed location and evaluated the performance of WiHand. For fixed location, we mean the training data and test data were all collected from the same location. Here, we chose ten volunteers to perform all five gestures 100 times repeatedly. Then, we randomly chose 60% of the data as training data and 40% as the test data. To make the result more representative, we applied cross validation of 10 times on the training.

Figure 10 shows the average detection rate considering each action. We found that WiHand showed better performance than the traditional recognition algorithm. Especially for specific gestures such as Gestures 3 and 5, WiHand greatly outperformed the traditional algorithm.

Figure 11 shows the average detection rate of each person. From the results, we found that WiHand effectively improved the detection accuracy. The average accuracy of WiHand for every actions was 97.3%. The average accuracy for every person was 93.3%.

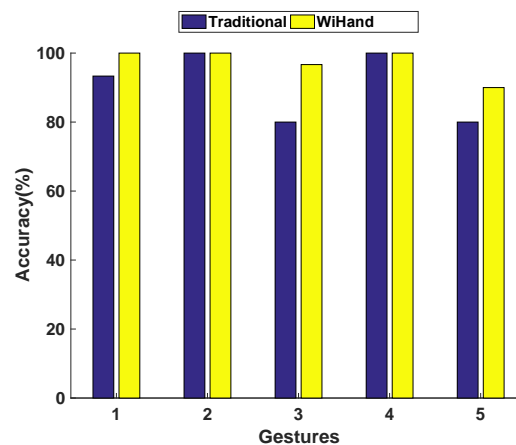


Figure 10. The average accuracy of detection for each gesture.

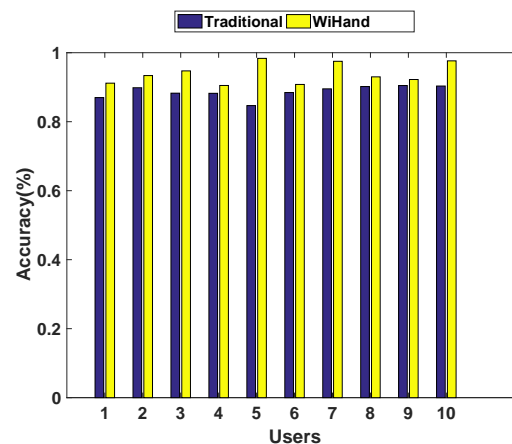


Figure 11. The average accuracy of gesture detection for each person.

5.4. Performance of Gesture Recognition at Various Locations

To test the performance of WiHand, we collected the gesture data of all volunteers at nine different locations. We used data from Position P1 to train the classifier and the rest for test. At each location, every volunteer performed the five gestures 100 times repeatedly. An illustration of all nine locations are shown in Figure 8. Their exact coordinates are shown in Table 2.

To get a comparative comprehension of the performance of WiHand, we also added some state-of-the-art classification methods from related works. One is the hidden Markov model based classification [55,61,64–67]. The other is the CNN based classification. We built a five-layer CNN network with have two convolutional layers, two sub-sampling layers and one fully-connected layer. For the first convolutional layer, the size of a convolutional kernel was 3×3 . We applied six different kernels to generate six feature maps. Considering the second convolutional layer, there were two kernels and the kernel size was still 3×3 . There were, in total, 12 feature maps as the output of this layer. There was only one kernel in the sub-sampling layer and the size was 2×2 . The input of the CNN network was a matrix of size 10×10 , which represented 10 subcarriers selected by our algorithm and each subcarrier had 10 features.

The performance of all the algorithms are shown in Figure 12. As shown in the figure, the performance of all the three classifiers with the LRSD algorithm outperformed the recognition from the original data. Besides, the performance of CNN was better than the others.

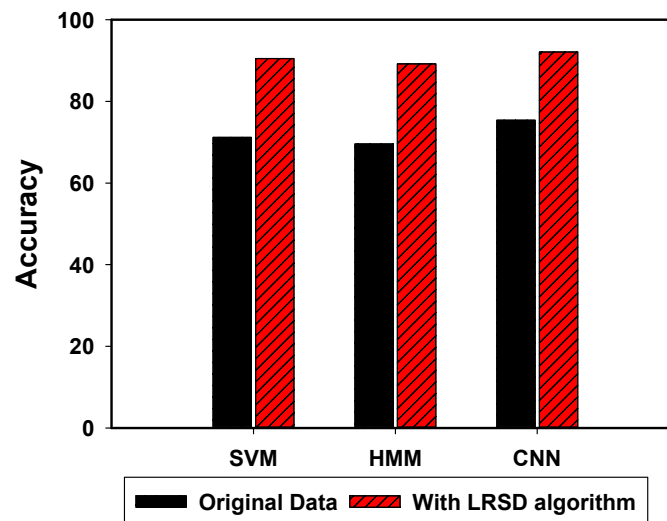


Figure 12. The performance of classifiers with different locations.

We also evaluated the performance of WiHand at different locations. The training data were collected from position P1. The results are shown in Figure 13. The average detection accuracy of WiHand at different locations is 93%. The results show that WiHand can resist the influence of location variation very well. In addition, we observed that the detection rate on the LOS path is a little higher than that on the side path. That is because the signal strength on the LOS path is higher than the other place.

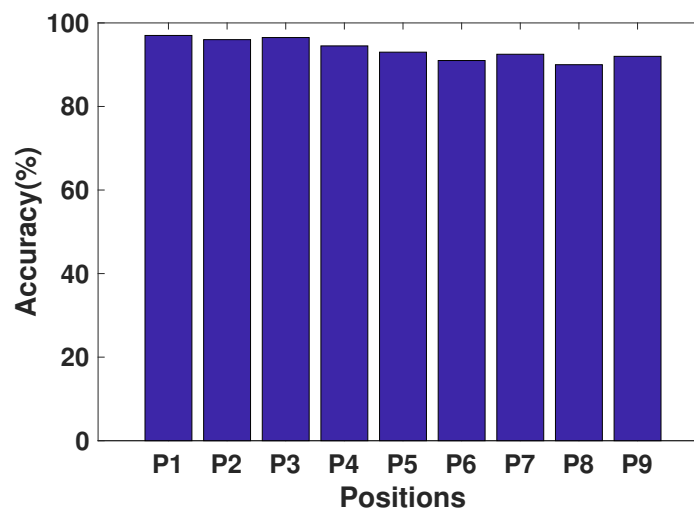


Figure 13. The performance of WiHand under different locations.

5.5. Through Wall Detection

To evaluate the performance of WiHand under none-line-of-sight (NLOS), i.e., through the wall scenario, we moved the RX to behind a concrete wall of 20 cm. Gestures were performed behind the wall. The result is shown in Figure 14a. Despite the decrease in accuracy under NLOS scenario, WiHand could still achieve accuracy as high as 91%. In addition, we tested the performance of different classifiers under through the wall scenario. With the LRSD algorithm, all classifiers achieved a good performance under through the wall scenario.

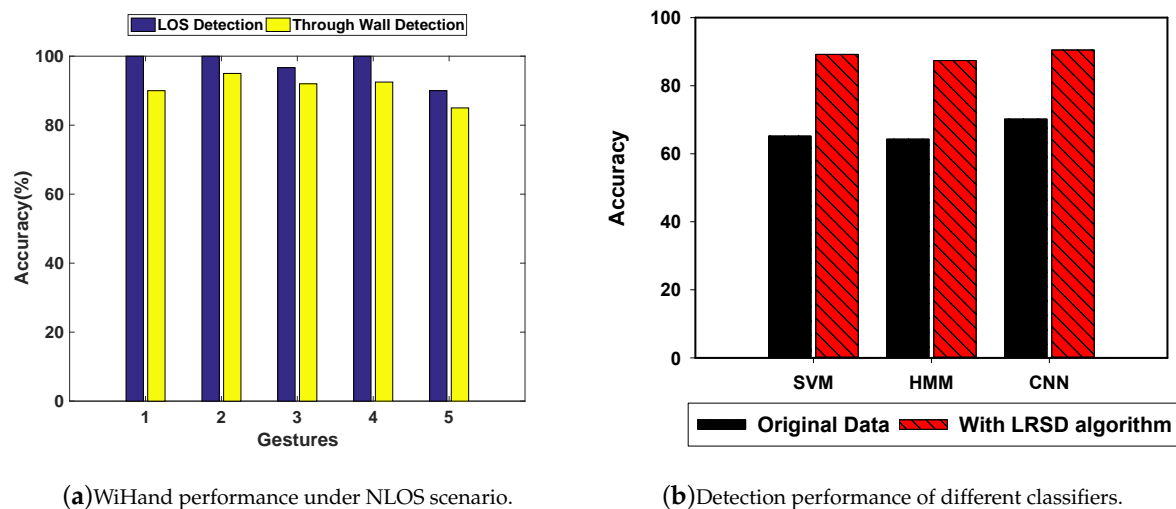


Figure 14. Detection performance under through the wall scenario.

6. Conclusions and Future Work

This paper proposes WiHand, a location independent gesture recognition system based on commodity WiFi devices. The proposed binned entropy based subcarrier selection algorithm can always find the mostly affected subcarriers. With the LRSD decomposition, WiHand separates gesture signals from background information. The extensive evaluations showed that WiHand can achieve an average accuracy of 93% across different locations and 91% for through the wall scenario.

As for future work, we will continue to collect data on more gestures and more persons under different scenarios.

Author Contributions: Y.L. and S.L. conceived and designed the method; X.W. guided the students to complete the research; Y.L. performed the simulation and experiment tests; S.L. and X.W. helped in the simulation and experiment tests; and Y.L. wrote the paper.

Funding: This research was funded by the National Science Foundation of China under grant number 61572513 and 61472434.

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

References

1. Uddin, M.T.; Uddiny, M.A. Human activity recognition from wearable sensors using extremely randomized trees. In Proceedings of the 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, 21–23 May 2015; pp. 1–6.
2. Jalal, A.; Zeb, M.A. Security and QoS optimization for distributed real time environment. In Proceedings of the 7th IEEE International Conference on Computer and Information Technology (CIT 2007), Fukushima, Japan, 16–19 October 2007; pp. 369–374.
3. Jalal, A.; Kamal, S. Real-time life logging via a depth silhouette-based human activity recognition system for smart home services. In Proceedings of the 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, Korea, 26–29 August 2014; pp. 74–80.
4. Ahad, M.A.R.; Kobashi, S.; Tavares, J.M.R. Advancements of image processing and vision in healthcare. *J. Healthc. Eng.* **2018**, *2018*. [[CrossRef](#)] [[PubMed](#)]
5. Jalal, A.; Quaid, M.A.K.; Kim, K. A Wrist Worn Acceleration Based Human Motion Analysis and Classification for Ambient Smart Home System. *J. Electr. Eng. Technol.* **2019**, *14*, 1–7. [[CrossRef](#)]
6. Jalal, A.; Nadeem, A.; Bobasu, S. Human Body Parts Estimation and Detection for Physical Sports Movements. In Proceedings of the 2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE), Islamabad, Pakistan, 6–7 March 2019; pp. 104–109.

7. Jalal, A.; Mahmood, M. Students' behavior mining in e-learning environment using cognitive processes with information technologies. *Educ. Inf. Technol.* **2019**, *24*, 2797–2821. [\[CrossRef\]](#)
8. Wan, Q.; Li, Y.; Li, C.; Pal, R. Gesture recognition for smart home applications using portable radar sensors. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 6414–6417, doi:10.1109/EMBC.2014.6945096. [\[CrossRef\]](#)
9. Patsadu, O.; Nukoolkit, C.; Watanapa, B. Human gesture recognition using Kinect camera. In Proceedings of the 2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE), Bangkok, Thailand, 30 May–1 June 2012; pp. 28–32. doi:10.1109/JCSSE.2012.6261920. [\[CrossRef\]](#)
10. Rahman, A.M.; Hossain, M.A.; Parra, J.; El Saddik, A. Motion-path Based Gesture Interaction with Smart Home Services. In Proceedings of the 17th ACM International Conference on Multimedia, Beijing, China, 19–24 October 2009; ACM: New York, NY, USA, 2009; pp. 761–764, doi:10.1145/1631272.1631407. [\[CrossRef\]](#)
11. Jalal, A.; Uddin, M.Z.; Kim, T. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. *IEEE Trans. Consum. Electron.* **2012**, *58*, 863–871. doi:10.1109/TCE.2012.6311329. [\[CrossRef\]](#)
12. Dinh, D.L.; Kim, J.T.; Kim, T.S. Hand Gesture Recognition and Interface via a Depth Imaging Sensor for Smart Home Appliances. *Energy Procedia* **2014**, *62*, 576–582. [\[CrossRef\]](#)
13. Wu, Y.; Huang, T.S. Vision-based gesture recognition: A review. In *International Gesture Workshop*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 103–115.
14. Chen, I.K.; Chi, C.Y.; Hsu, S.L.; Chen, L.G. A real-time system for object detection and location reminding with rgb-d camera. In Proceedings of the 2014 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 10–13 January 2014; pp. 412–413.
15. Jalal, A.; Kim, S.; Yun, B. Assembled algorithm in the real-time H. 263 codec for advanced performance. In Proceedings of the 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry, Busan, Korea, 23–25 June 2005; pp. 295–298.
16. Fonseca, L.M.G.; Namikawa, L.M.; Castejon, E.F. Digital Image Processing in Remote Sensing. In Proceedings of the 2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI-TUTORIALS'09, Rio de Janeiro, Brazil, 11–14 October 2009; IEEE Computer Society: Washington, DC, USA, 2009; pp. 59–71, doi:10.1109/SIBGRAPI-Tutorials.2009.13, [\[CrossRef\]](#)
17. Jalal, A.; Kim, S. The mechanism of edge detection using the block matching criteria for the motion estimation. In Proceedings of the Korea HCI Society Conference, Las Vegas, NV, USA, 22–27 July 2005; pp. 484–489.
18. Jalal, A.; Uddin, M.Z.; Kim, J.T.; Kim, T.S. Daily Human Activity Recognition Using Depth Silhouettes and R Transformation for Smart Home. In Proceedings of the International Conference on Smart Homes and Health Telematics, Montreal, QC, Canada, 20–22 June 2011; pp. 25–32.
19. Procházka, A.; Kolinova, M.; Fiala, J.; Hampl, P.; Hlavaty, K. Satellite image processing and air pollution detection. In Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (Cat. No. 00CH37100), Istanbul, Turkey, 5–9 June 2000; Volume 4, pp. 2282–2285.
20. Microsoft X-box Kinect. Available online: <http://www.xbox.com> (accessed on 18 September 2019).
21. Leap Motion. Available online: <https://www.leapmotion.com> (accessed on 18 September 2019)
22. Bo, C.; Jian, X.; Li, X.Y.; Mao, X.; Wang, Y.; Li, F. You're driving and texting: detecting drivers using personal smart phones by leveraging inertial sensors. In Proceedings of the 19th Annual International Conference on Mobile Computing & Networking, Miami, FL, USA, 30 September–4 October 2013; pp. 199–202.
23. 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 & Networking, Miami, FL, USA, 30 September–4 October 2013; ACM: New York, NY, USA, 2013; pp. 27–38. doi:10.1145/2500423.2500436. [\[CrossRef\]](#)
24. Abdelnasser, H.; Youssef, M.; Harras, K.A. WiGest: A ubiquitous WiFi-based gesture recognition system. In Proceedings of the 2015 IEEE Conference on Computer Communications, INFOCOM, Hong Kong, China, 26 April–1 May 2015; pp. 1472–1480. doi:10.1109/INFOCOM.2015.7218525. [\[CrossRef\]](#)
25. 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 2015, Las Vegas, NV, USA, 3–6 August 2015; pp. 1–7. doi:10.1109/ICCCN.2015.7288485. [\[CrossRef\]](#)

26. Li, H.; Yang, W.; Wang, J.; Xu, Y.; Huang, L. WiFinger: Talk to Your Smart Devices with Finger-grained Gesture. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; ACM: New York, NY, USA, 2016; pp. 250–261, doi:10.1145/2971648.2971738. [CrossRef]
27. Tan, S.; Yang, J. WiFinger: Leveraging Commodity WiFi for Fine-grained Finger Gesture Recognition. In Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc'16, Paderborn, Germany, 5–8 July 2016; ACM: New York, NY, USA, 2016; pp. 201–210. doi:10.1145/2942358.2942393. [CrossRef]
28. Zhang, O.; Srinivasan, K. Mudra: User-friendly Fine-grained Gesture Recognition Using WiFi Signals. In Proceedings of the 12th International Conference on Emerging Networking EXperiments and Technologies, CoNEXT'16, Irvine, CA, USA, 12–15 December 2016; ACM: New York, NY, USA, 2016; pp. 83–96. doi:10.1145/2999572.2999582. [CrossRef]
29. CISCO. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021*; Cisco VNI Forecast, New York, NY, USA, 2016.
30. Peng, Y.; Ganesh, A.; Wright, J.; Xu, W.; Ma, Y. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2233–2246. [CrossRef]
31. Liu, X.; Zhao, G.; Yao, J.; Qi, C. Background subtraction based on low-rank and structured sparse decomposition. *IEEE Trans. Image Process.* **2015**, *24*, 2502–2514. [CrossRef] [PubMed]
32. Jiang, W.; Miao, C.; Ma, F.; Yao, S.; Wang, Y.; Yuan, Y.; Xue, H.; Song, C.; Ma, X.; Koutsoukolas, D.; et al. Towards Environment Independent Device Free Human Activity Recognition. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, MobiCom'18, New Delhi, India, 29 October–2 November 2018; ACM: New York, NY, USA, 2018; pp. 289–304. doi:10.1145/3241539.3241548. [CrossRef]
33. Virmani, A.; Shahzad, M. Position and Orientation Agnostic Gesture Recognition Using WiFi. In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, Niagara Falls, MobiSys'17, New York, NY, USA, 19–23 June 2017; ACM: New York, NY, USA, 2017; pp. 252–264, doi:10.1145/3081333.3081340. [CrossRef]
34. 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, MobiCom'14, Maui, HI, USA, 7–11 September 2014; ACM: New York, NY, USA, 2014; pp. 617–628, doi:10.1145/2639108.2639143. [CrossRef]
35. Adib, F.; Kabelac, Z.; Katabi, D. *Multi-Person Motion Tracking via RF Body Reflections*; CSAIL Technical Reports. Available online: <http://hdl.handle.net/1721.1/86299> (accessed on 18 September 2019).
36. Sigg, S.; Shi, S.; Ji, Y. RF-Based Device-free Recognition of Simultaneously Conducted Activities. In Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp'13 Adjunct, Zurich, Switzerland, 8–12 September 2013; ACM: New York, NY, USA, 2013; pp. 531–540, doi:10.1145/2494091.2497611. [CrossRef]
37. Zhu, H.; Xiao, F.; Sun, L.; Wang, R.; Yang, P. R-TTWD: Robust device-free through the wall detection of moving human with WiFi. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1090–1103. [CrossRef]
38. Sigg, S.; Shi, S.; Büsching, F.; Ji, Y.; Wolf, L.C. Leveraging RF-channel fluctuation for activity recognition: Active and passive systems, continuous and RSSI-based signal features. In Proceedings of the 11th International Conference on Advances in Mobile Computing & Multimedia, MoMM'13, Vienna, Austria, 2–4 December 2013; p. 43, doi:10.1145/2536853.2536873. [CrossRef]
39. Lv, S.; Lu, Y.; Dong, M.; Wang, X.; Dou, Y.; Zhuang, W. Qualitative action recognition by wireless radio signals in human–machine systems. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 789–800. [CrossRef]
40. Rathore, M.M.U.; Ahmad, A.; Paul, A.; Wu, J. Real-time continuous feature extraction in large size satellite images. *J. Syst. Archit.* **2016**, *64*, 122–132. [CrossRef]
41. Jalal, A.; Kim, Y.; Kamal, S.; Farooq, A.; Kim, D. Human daily activity recognition with joints plus body features representation using Kinect sensor. In Proceedings of the 2015 International Conference on Informatics, Electronics & Vision (ICIEV), Fukuoka, Japan, 15–18 June 2015; pp. 1–6.

42. Jalal, A.; Kamal, S.; Kim, D. Individual detection-tracking-recognition using depth activity images. In Proceedings of the 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Goyang, Korea, 28–30 October 2015; pp. 450–455.
43. Yoshimoto, H.; Date, N.; Yonemoto, S. Vision-based real-time motion capture system using multiple cameras. In Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI2003, Tokyo, Japan, 1 August 2003; pp. 247–251.
44. Kamal, S.; Jalal, A.; Kim, D. Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM. *J. Electr. Eng. Technol.* **2016**, *11*, 1921–1926. [[CrossRef](#)]
45. Jalal, A.; Kamal, S.; Kim, D. Facial Expression recognition using 1D transform features and Hidden Markov Model. *J. Electr. Eng. Technol.* **2017**, *12*, 1657–1662.
46. Huang, Q.; Yang, J.; Qiao, Y. Person re-identification across multi-camera system based on local descriptors. In Proceedings of the 2012 Sixth International Conference on Distributed Smart Cameras (ICDSC), Hong Kong, China, 30 October–2 November 2012; pp. 1–6.
47. Zeng, Y.; Pathak, P.H.; Mohapatra, P. WiWho: WiFi-Based Person Identification in Smart Spaces. In Proceedings of the 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Vienna, Austria, 11–14 April 2016; pp. 1–12, doi:10.1109/IPSN.2016.7460727. [[CrossRef](#)]
48. Jin, Y.; Soh, W.S.; Wong, W.C. Indoor Localization with Channel Impulse Response Based Fingerprint and Nonparametric Regression. *Trans. Wireless. Comm.* **2010**, *9*, 1120–1127. doi:10.1109/TWC.2010.03.090197. [[CrossRef](#)]
49. Davies, L.; Gather, U. The identification of multiple outliers. *J. Am. Stat. Assoc.* **1993**, *88*, 782–792. [[CrossRef](#)]
50. Wang, W.; Liu, A.X.; Shahzad, M.; Ling, K.; Lu, S. Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, MobiCom 2015, Paris, France, 7–11 September 2015; pp. 65–76, doi:10.1145/2789168.2790093. [[CrossRef](#)]
51. Zhang, J.; Wei, B.; Hu, W.; Kanhere, S.S. WiFi-ID: Human Identification Using WiFi Signal. In Proceedings of the 2016 International Conference on Distributed Computing in Sensor Systems (DCOSS), Washington, DC, USA, 26–28 May 2016; pp. 75–82, doi:10.1109/DCOSS.2016.30. [[CrossRef](#)]
52. Abdelnasser, H.; Harras, K.A.; Youssef, M. UbiBreathe: A Ubiquitous non-Invasive WiFi-based Breathing Estimator. In Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc 2015, Hangzhou, China, 22–25 June 2015; pp. 277–286, doi:10.1145/2746285.2755969. [[CrossRef](#)]
53. 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, MobiCom’14, Maui, HI, USA, 7–11 September 2014; ACM: New York, NY, USA, 2014; pp. 593–604, doi:10.1145/2639108.2639112.
54. Zeng, Y.; Pathak, P.H.; Mohapatra, P. Analyzing Shopper’s Behavior Through WiFi Signals. In Proceedings of the 2nd Workshop on Workshop on Physical Analytics, WPA’15, Florence, Italy, 22 May 2015; ACM: New York, NY, USA, 2015; pp. 13–18. doi:10.1145/2753497.2753508. [[CrossRef](#)]
55. Piyathilaka, L.; Kodagoda, S. Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features. In Proceedings of the 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), Melbourne, Australia, 19–21 June 2013; pp. 567–572.
56. Jalal, A.; Kamal, S.; Kim, D. A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems. *Int. J. Interact. Multimed. Artif. Intell.* **2017**, *4*, 54–62. doi: 10.9781/ijimai.2017.447
57. Mahmood, M.; Jalal, A.; Evans, H.A. Facial expression recognition in image sequences using 1D transform and gabor wavelet transform. In Proceedings of the 2018 International Conference on Applied and Engineering Mathematics (ICAEM), Taxila, Pakistan, 4–5 September 2018; pp. 1–6.
58. Jalal, A.; Kamal, S. Improved Behavior Monitoring and Classification Using Cues Parameters Extraction from Camera Array Images. *Int. J. Interact. Multimed. Artif. Intell.* **2019**, *5*, 71–78. [[CrossRef](#)]
59. Jalal, A.; Quaid, M.A.; Siddiqui, M. A Triaxial acceleration-based human motion detection for ambient smart home system. In Proceedings of the 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 8–12 January 2019; pp. 353–358.

60. Jalal, A.; Mahmood, M.; Hasan, A.S. Multi-features descriptors for human activity tracking and recognition in Indoor-outdoor environments. In Proceedings of the 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 8–12 January 2019; pp. 371–376.
61. Wu, H.; Pan, W.; Xiong, X.; Xu, S. Human activity recognition based on the combined svm&hmm. In Proceedings of the 2014 IEEE International Conference on Information and Automation (ICIA), Hailar, China, 28–30 July 2014; pp. 219–224.
62. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 27. [[CrossRef](#)]
63. Xie, Y.; Li, Z.; Li, M. Precise Power Delay Profiling with Commodity WiFi. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, MobiCom 2015, Paris, France, 7–11 September 2015; pp. 53–64. doi:10.1145/2789168.2790124. [[CrossRef](#)]
64. Jalal, A.; Kamal, S.; Kim, D. A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors* **2014**, *14*, 11735–11759. [[CrossRef](#)] [[PubMed](#)]
65. Jalal, A.; Kamal, S.; Kim, D. Shape and motion features approach for activity tracking and recognition from kinect video camera. In Proceedings of the 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, Gwangju, Korea, 24–27 March 2015; pp. 445–450.
66. Jalal, A.; Quaid, M.A.K.; Hasan, A.S. Wearable Sensor-Based Human Behavior Understanding and Recognition in Daily Life for Smart Environments. In Proceedings of the 2018 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 17–19 December 2018; pp. 105–110.
67. Mahmood, M.; Jalal, A.; Siddiqui, M. Robust Spatio-Temporal Features for Human Interaction Recognition Via Artificial Neural Network. In Proceedings of the 2018 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 17–19 December 2018; pp. 218–223.



© 2019 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/>).