Wi-Fi Gesture Recognition Technology Based on Time-Frequency Features (Workshop)

Zengshan Tian, Mengtian Ren(✉), Qing Jiang, and Xiaoya Zhang

School of Communication and Information Engineering,
Chongqing University of Posts and Telecommunications,
Chongqing 400065, China
rmt812@126.com

Abstract. With the rapid development of artificial intelligence, gesture recognition has become the focus of many countries for research. Gesture recognition using Wi-Fi signals has become the mainstream of gesture recognition because it does not require additional equipment and lighting conditions. Firstly, how to extract useful gesture signals in a complex indoor environment. In this paper, after de-noising the signal by Discrete Wavelet Transform (DWT) technology, Principal Component Analysis (PCA) is used to eliminate the problem of signal redundancy between multiple CSI subcarriers, further to remove noise. Secondly, the frequency domain features of the gesture signal are constructed by performing Short-Time Fourier Transform (STFT) on the denoised CSI amplitude signal. Then, the time domain features are combined with the frequency domain features, and the features are trained and classified using the Support Vector Machine (SVM) classification method to complete the training and recognition of gesture. The experimental results show that this paper can effectively identify gestures in complex indoor environments.

Keywords: Dynamic gesture recognition · Discrete Wavelet Transform · Principal Component Analysis · Time-frequency domain features

1 First Section

With the rapid development of science and technology in the 21st century and the popularity of computers, human-computer interaction technology has become the object of attention and research in many countries [1]. Gesture as one of the basic features of vision, it is the simplest way to interact with nature. And it plays an important role in many fields such as smart home and auxiliary car control system. Therefore, gesture recognition has gradually become an important research direction of human-computer interaction.

At present, the previous works are mainly based on wearable sensors [2], computer vision [3], and radiofrequency signals [4]. Among them, the gesture

H. Gao et al. (Eds.): ChinaCom 2019, LNICST 313, pp. 402–413, 2020.
https://doi.org/10.1007/978-3-030-41117-6_33

recognition system based on wearable devices and computer vision started earlier, has been very mature so far and achieved gratifying precision. However, the wearable sensors system requires the user to wear a sensing device such as a data glove or an armband to obtain response parameters, it is inconvenient for the user; the computer vision system cannot be used in dark or smoke environments. Compared with them, the Wi-Fi signal-based gesture recognition system does not require a piece of wearable equipment and the condition of light. It only needs to use software upgrades and updates, so it gradually becomes the mainstream of gesture recognition.

In the Wi-Fi based gesture recognition system, most systems currently use Received Signal Strength Indications (RSSI) to capture a signal. However, RSSI has poor stability under uncertain noise and indoor multipath conditions, therefore RSSI does not provide sufficient reliability. As the pursuit of reliability becomes higher and higher, Channel State Information (CSI) gradually appears in everyone's field of vision. In this paper, we realized a gesture recognition system based on CSI which is collected from commercial wireless devices. It does not require additional sensors, is resilient to changes within the environment and can operate in non-line-of-sight scenarios. The basic idea is to leverage the amplitude of CSI to complete the gesture recognition. There are several challenges, however, that need to be addressed to realize our system including handling the noisy and extracting the feature of different gestures. To address these challenges, the main contributions of this paper are as follows:

1. How to extract gesture signals from complex indoor environments? First, the least variance method is used to segment the gesture. Then, we use the Discrete Wavelet Transform-Principal Component Analysis (DTW-PCA) to extract useful gesture signals in a complex environment.
2. How to extract features of different gestures for classification? We propose a method that combines time domain and frequency domain feature information to complete classification.

The rest of this paper is organized as follows. We first summarize the related work in Sect. 2, followed by Sect. 3, which is a brief introduction to the gesture recognition system. Then in Sect. 4 we illustrate the detailed system design and methodology. We show the experimental results and evaluation of the performance in Sect. 5. Finally, we will conclude this work and list our future work in Sect. 6.

2 Related Work

In this section, we introduce the state-of-the-artwork of gesture recognition systems, which are related to our work. These systems are mainly divided into the following two categories.

(1) Device-Based Gesture Systems: These technologies include wearable sensors and computer vision: As mentioned earlier, the wearable based methods usually need gloves or external sensors attached to a user for gesture capture.

These methods can capture the gestures more precisely, but the process is invasive, as the user needs to wear sensors around him. For example, Wang et al. [5] describe a system using data glove for gesture recognition, which can measure the change of hand joints' angles and the motion state of the hand. And PhonePoint Pen [6] recognizes handwriting by holding a mobile phone in hands; The other type is camera-based systems, the camera based method uses the camera to capture user gesture behavior, which relies on high-resolution video or images and cannot be used in non-light conditions for example darkness. [7] proposes a camera-based hand gesture that can achieve a higher average recognition rate and better distinguishes the confusion gesture. The Xbox Kinect [8] and Leap Motion [9] also can be the typical successful examples of applications.

(2) Device-Free Gesture Systems: Nowadays, Kellogg et al. [4] can use Radio-frequency identification (RFID) to identify gestures and it can achieve high precision, but the system only works with RFID transmissions whose costs are high. In the device-free systems, the wireless signal-based method has a low expanse and is easy to deploy. In 2015, Abdelnasser et al. proposed WiGest [10], which performs gesture recognition by analyzing the rising and falling edges of RSSI signal changes. The accuracy is 87.5% in the case of a single access point when there are three access points. the accuracy rate is 96%. But the RSSI is lack of stable in complex indoor environments. The system WiGeR [11] leverages the fluctuation of the amplitude of CSI caused by the gesture to complete the recognition. But the pattern recognition such as dynamic time warping (DTW) takes a lot of time so that it is not practical.

All of these systems contain some weaknesses which make it hard to be further popularized, though most of them can achieve an impressive estimation accuracy.

3 System Overview

In this section, we first briefly introduce the background knowledge of CSI value which is the foundation of our system, then give an overview of the system.

3.1 Channel State Information

The indoor environment is complex and changeable, and there are interferences of multipath effects, such as static reflections on walls, sofas, tables and so on. The model diagram is shown in Fig. 1.

According to Fig. 1, the CSI data received by the receiver is a superposition of multiple path signals, which can be expressed as

$$\begin{aligned} H\left(f_i, t\right) &= \left|H\left(f_i, t\right)\right| \times \arg\left(H\left(f_i, t\right)\right) \\ &= e^{-j2\pi ft} \sum_{k=1}^{N} a_k\left(f_i, t\right) e^{-j2\pi f_i \tau_k(t)} + n \end{aligned} \tag{1}$$

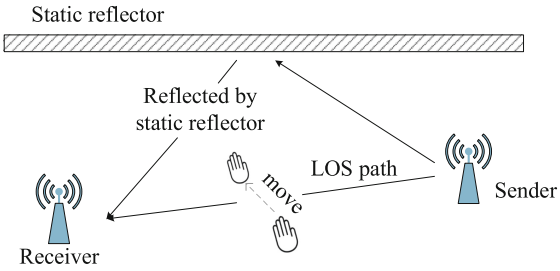

Fig. 1. Indoor environment diagram.

where $|H(f_i,t)|$ and $\arg(H(f_i,t))$ are the amplitude and phase of the i subcarrier at the t moment, f_i is the frequency of the subcarrier, $e^{-j2\pi ft}$ is phase shift due to phase error, N is the number of multipath, $a_k(f_i,t)$ is the propagation attenuation of the first multipath, $e^{-j2\pi f_i \tau_k(t)}$ is the phase shift due to propagation delay, n is noise vector.

The literature [12] derives the square formula of the CSI amplitude according to formula (1) as follows

$$\begin{aligned}|H(f_i,t)|^2 = &\sum_{k\in P_d} 2\left|H_s(f_i)\,a_k(f_i,t)\right|\cos\left(\tfrac{2\pi\nu_k t}{\lambda}+\tfrac{2\pi d_k(0)}{\lambda}+\varphi_{sk}\right)\\ &+\left|H_s(f_i)\right|^2+\sum_{k\in P_d}\left|a_k(f_i,t)\right|^2\\ &+\sum_{\substack{k,l\in P_d\\ k\neq l}}\left|a_k(f_i,t)\,a_l(f_i,t)\right|\cos\left(\tfrac{2\pi(\nu_k-\nu_l)t}{\lambda}+\tfrac{2\pi(d_k(0)-d_l(0))}{\lambda}+\varphi_{kl}\right)\end{aligned} \tag{2}$$

It can be obtained from formula (2) that the square of the CSI amplitude is composed of some constant and cosine functions, which can accurately reflect the jitter of the target motion. Therefore, this paper considers the feature information of CSI amplitude to complete the training and recognition of the gesture.

3.2 Overview

As shown in Fig. 2, our gesture recognition system consists of three modules: CSI data preprocessing, feature extraction of gesture and gesture recognition. In the CSI data preprocessing, we use the DWT to remove noise, and PCA is also used to reduce the data dimension as well as remove noise further. After CSI data preprocessing, we combine time domain and frequency domain features information to extract the features such as event duration, interquartile range, spectral entropy and so on. At last, these features be used for classification in gesture recognition.

4 System Design

In this section, we elaborate on the methodology of our system relied on three blocks as mentioned before.

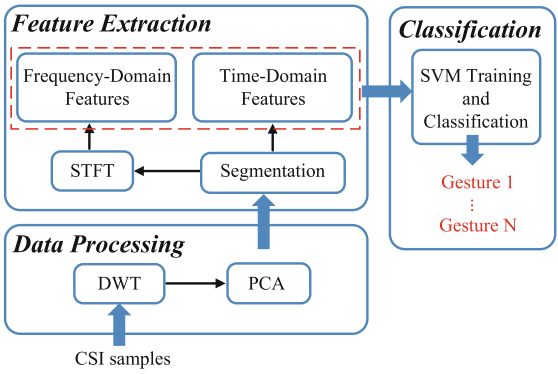

Fig. 2. Overview of system.

4.1 Data Preprocessing

As can be seen from Fig. 1, the CSI data received by the receiver is a superposition of multiple path signals, and there are interferences from static reflections such as walls, tables and chairs. So how to extract the signal changes caused by the user's gesture movement in a complex indoor environment is a huge challenge.

In response to this problem, first, this paper uses the Discrete Wavelet Transform (DWT) technique to effectively remove the noise and smooth the CSI data. The core thought is three steps: decomposition, noise removal and reconstruction. Firstly, DWT decomposes the signal into detail coefficients $\left\{\beta^1 f \beta^2 f \cdots f \beta^J\right\}$ (with $J = 5$) and approximation coefficients α^J. Then we apply hard thresholding to denoise the 1–3 layer detail coefficients. Finally, we combine the approximation coefficient and the denoised detail coefficient to reconstruct the CSI signal.

Second, since the CSI signal has 30 subcarriers, the information carried by the 30 subcarriers is redundant and the number of all subcarriers increases the computational complexity. Therefore, we use the Principal Component Analysis (PCA) to effectively reduce the data, and further achieve the denoising. PCA is a mathematical transformation method that converts a given set of related variables into another set of unrelated variables by a linear transformation. These new variables are arranged in descending order of variance. In the mathematical transformation, the total variance of the variables is kept constant, so that the first variable has the largest variance, which is called the first principal component, and the variance of the second variable is the second largest, and is not related to the first variable, and is called the second principal component, and so on. According to the literature [13], the first principal component after the dimension reduction contains almost all the information of the gesture motion signal.

Figure 3 shows the result of noise filtering. It shows the DWT-PCA method can extract the signal from the complex indoor environment. Figure 4 compares

the first three principal components for the same activity when de-noising is applied before PCA. It also shows that a jority of the gesture activity induced variation is concentrated in the first and second principal components, but in the third principal component (and onwards) the noise level begins to have a higher influence [13].

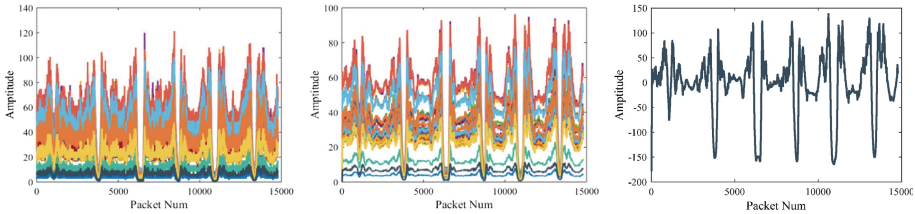

Fig. 3. Channel State Information (CSI) preprocessing.

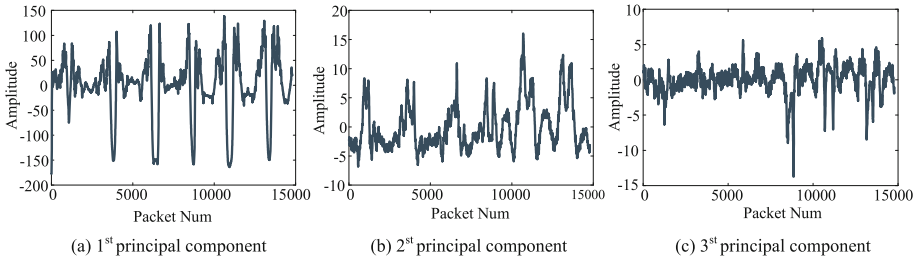

Fig. 4. Comparison of the resultant first three PCs after wavelet de-noising and then PCA. First and second PCs are less noisy, yet in the third PC, the noise level has a higher influence.

4.2 Segment Algorithm

The CSI data is a continuous-time series. To complete feature extraction and classification of gestures, segmentation of the received time series is required. To solve this problem, we divide the continuous-time series by the smallest variance segment algorithm.

In the segmented algorithm, we use two windows on the CSI time series to segment it. Firstly, The first window corresponds to the beginning of the time series, and the beginning of the second window corresponds to the end of the first window; Secondly, the width W is initialized to $W_{\min}$, the first window is $P_{frist} = (P_1, P_2, \cdots, P_w)$ and the second window is $P_{\sec ond} = (p_{w+1}, p_{w+2}, \cdots, p_{2w})$; Thirdly, we calculate the variance difference between the two windows; Then,

we make $w = w+1$ and repeat the second step to calculate the variance difference of the two windows until the width of the window reaches $W_{\max}$. We can get $W_{\max} - W_{\min} + 1$ pairs of (w, d). The window width W corresponds to the minimum d is the final window width. Finally, repeat the previous steps unless there is not enough time series. Figure 5 shows the result of gesture segmentation, we can see from Fig. 5 the algorithm can segment the gesture more accurately. The results of 5 kinds of gesture segmentation are shown in Fig. 6.

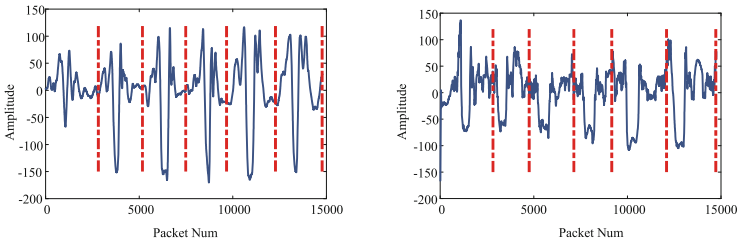

Fig. 5. The segmentation graph of Wave gesture.

4.3 Time Frequency Analysis

After CSI data collection and preprocessing, a transformation to the time-frequency domain is necessary to perform the feature extraction. For the time-frequency analysis, there are various linear and non-linear techniques. Among them, the non-linear methods tend to distort the frequency components generated by movement. Therefore, we selected Short-Time Fourier Transform (STFT) for our time-frequency analysis.

In STFT, the time resolution and frequency resolution are inversely proportional, so it is necessary to find an optimal window size to obtain satisfactory time resolution and frequency resolution. Since each person completes a gesture action for 1 to 4 s, the time is short and requires sufficient time resolution. For this problem, we opt for an FFT window size of 256 samples at a sampling rate of 1000 pkts/s. We chose the overlap size of two windows to be 250 samples. So, the selected parameters provide us with a frequency resolution of $\frac{sample\ rate}{FFTsize} \approx 4$ Hz and a time resolution of $\frac{window-overlap}{sample\ rate} \approx 0.006$ s. Figure 6 shows the STFT of five gestures.

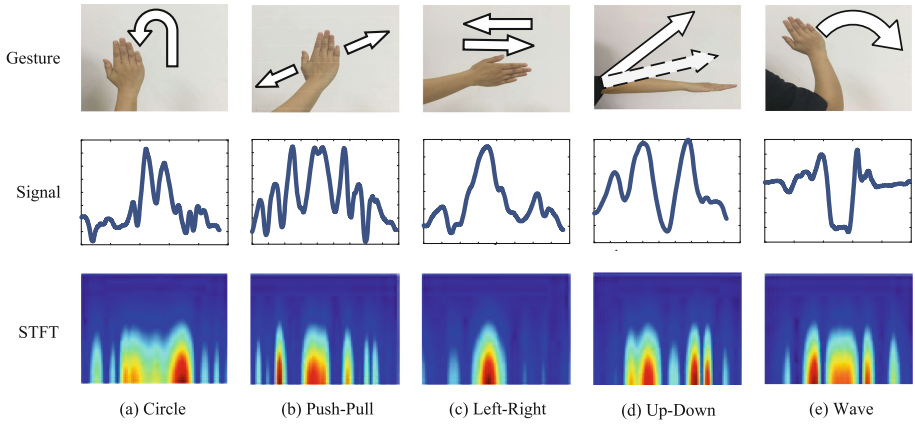

Fig. 6. Time domain signal diagram and spectrum diagram corresponding to the five gestures.

4.4 Feature Extraction

Although the time domain signal and the frequency domain signal can well reflect the characteristics of different gesture motions, most of the pattern recognition methods are not feasible due to the inconsistency of time series. The time-series-inconsistent recognition method, for example, Hidden Markov Model has a large time expenditure and is not suitable for practical application scenarios. Therefore, we propose to select useful features for training and classification. The selected features are shown in Table 1.

Table 1. Selected eigenvalue

Time domain	Frequency domain
Standard deviation	Spectral entropy of 1–10 Hz
Interquartile range	Spectral entropy of 10–20 Hz
Event duration	Spectral entropy of 20–30 Hz

The literature [14] uses the mean, standard deviation, and interquartile range of time-domain features to identify the gesture but due to the small amplitude of gestures, these statistical feature values are not representative. In response to this problem, we choose the standard deviation, interquartile range and duration of each gesture motion are used as the time-domain feature. The standard deviation can well reflect the degree of dispersion of the gesture signal based on the mean, which is more representative than the mean; The interquartile range reflects the degree of dispersion of 50% of the data in the middle of the signal, and is not affected by the extreme value; Since the time of each action is different, the duration of each action obtained by segmentation is also used as the feature.

For frequency-domain features, conventional algorithms use feature such as mean and standard deviation of frequency domain information as criteria for gesture classification. In this paper, we propose spectral entropy in different frequency ranges as the feature. This is a normalized feature and measures the textural properties of a gesture (randomness in the distribution of energy in a spectrogram). The calculation formula is as follows

$$\begin{cases} H = -\sum_{i=k_l}^{k_u} p\left(n_i\right) \ln p\left(n_i\right) \\ p\left(n_i\right) = \frac{\hat{p}(n_i)}{\sum \hat{p}(n_i)} \\ \hat{p}\left(n_i\right) = |S|^2 \end{cases} \tag{3}$$

where $\hat{p}\left(n_i\right)$ is the of spectrum amplitude, $p\left(n_i\right)$ is the normalized power spectral density. k_l and k_u are lower and upper frequency bounds. H is the spectral entropy in the corresponding frequency range.

5 Experimental Evaluation

In this section, we will describe the relevant experimental setup and analyze the result of our experiments.

5.1 Lab Environment

The proposed gesture system contains a transmitter and a receiver both equipped with the Intel 5300 wireless NIC and CSI toolkit, and the parameters setting is shown in Table 2. All the experiments are conducted in a typical indoor environment with the size for 70 square meters, surrounded by meeting tables, chairs and other furniture. During the experiments, five gestures shown in Fig. 6 are designed to verify the effectiveness of the proposed system.

Table 2. Parameter setting

Parameters	Transmitting AP	Receiving AP
Mode	Injection	Monitor
Channel number	Default = 149 5.745 GHz	
Bandwidth	Default = 40 MHz	
Number of subcarriers	30	
Index of subcarriers	$[-58, -54, \ldots, 54, 58]$	
Transmit power	15 dBm	

5.2 Result

Classification Accuracy. This paper first verifies the effectiveness of the algorithm. Five groups of 100 gestures each with 500 groups to build classifiers and test samples, and used 3 times cross-validation and SVM classifier to calculate classification accuracy. Figure 7 is the confusion matrix diagram.

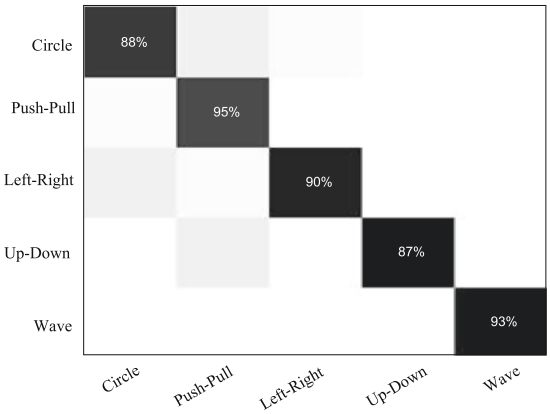

Fig. 7. Confusion matrix diagram at 1000 sampling rate.

The results of Fig. 8 show that the classification accuracy of the five gestures in the method of extracting time domain features and frequency domain features is {88%, 95%, 90%, 87%, 93%}, and the overall average accuracy of the five gestures is 90.6%. Among them, the gestures Push-Pull and Wave have higher recognition accuracy, while the Circle and Up-Down gesture recognition accuracy is relatively low, because the Push-Pull and Wave gestures have a greater impact on the direct link, and the features are more obvious.

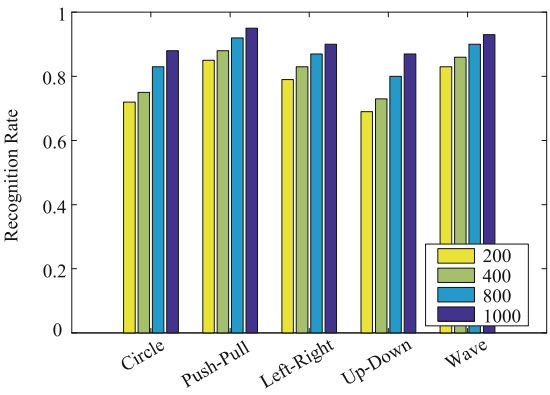

Fig. 8. Recognition accuracy under different sampling rates.

Impact of Sample Rate. As can be seen from Fig. 8, the increase of the sampling rate improves the recognition rate of the system. If the sampling rate is increased to 2000 packet/s, the recognition accuracy will be higher, but it will also reduce the performance of the device and the operating efficiency of the system, so the sampling rate of this paper is 1000 Hz.

6 Conclusion

In this paper, we propose a device-free gesture recognition system based on channel state information. First, the DWT-PCA combined denoising method is adopted for extracting useful gesture signals in complex environments, after denoising by DWT, PCA solves the problem of signal redundancy between multiple CSI subcarriers to further remove noise; Secondly, the frequency domain characteristics of the gesture signal are constructed by using the STFT of the processed CSI amplitude signal; Then, the time domain features are combined with the frequency domain features, and the features are trained and classified using the SVM classification method. The experimental results show that this paper can effectively recognize gestures in complex indoor environments. The average recognition rate reached 90.6%.

References

1. Zhou, Y., Jiang, G., Lin, Y.: A novel finger and hand pose estimation technique for real-time hand gesture recognition. Pattern Recogn. **49**, 102–114 (2016)
2. Sturman, D.J., Zeltzer, D.: A survey of glove-based input. IEEE Comput. Graphics Appl. **14**(1), 30–39 (1994)
3. He, Y., Yang, J., Shao, Z., Li, Y.: Salient feature point selection for real time RGB-D hand gesture recognition. In: 2017 IEEE International Conference on Real-time Computing and Robotics (RCAR), pp. 103–108. IEEE (2017)
4. Kellogg, B., Talla, V., Gollakota, S.: Bringing gesture recognition to all devices. In: 11th fUSENIXg Symposium on Networked Systems Design and Implementation (fNSDIg 2014), pp. 303–316 (2014)
5. Wang, X., Sun, G., Han, D., Zhang, T.: Data glove gesture recognition based on an improved neural network. In: Proceedings of the 29th Chinese Control Conference, pp. 2434–2437. IEEE (2010)
6. Agrawal, S., Constandache, I., Gaonkar, S., Roy Choudhury, R., Caves, K., DeRuyter, F.: Using mobile phones to write in air. In: Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, pp. 15–28. ACM (2011)
7. Wu, X., Mao, X., Chen, L., Xue, Y.: Trajectory-based view-invariant hand gesture recognition by fusing shape and orientation. IET Comput. Vision **9**(6), 797–805 (2015)
8. Microsoft kinect. http://www.microsoft.com/en-us/kinectforwindows
9. Leap motion. https://www.leapmotion.com
10. Abdelnasser, H., Youssef, M., Harras, K.A.: WiGest: a ubiquitous wifi-based gesture recognition system. In: 2015 IEEE Conference on Computer Communications (INFOCOM), pp. 1472–1480. IEEE (2015)

11. Al-qaness, M., Li, F.: WiGeR: WiFi-based gesture recognition system. ISPRS Int. J. Geo-Inf. **5**(6), 92 (2016)
12. 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, pp. 65–76. ACM (2015)
13. Palipana, S., Rojas, D., Agrawal, P., Pesch, D.: FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **1**(4), 155 (2018)
14. He, W., Wu, K., Zou, Y., Ming, Z.: WiG: WiFi-based gesture recognition system. In: 2015 24th International Conference on Computer Communication and Networks (ICCCN), pp. 1–7. IEEE (2015)