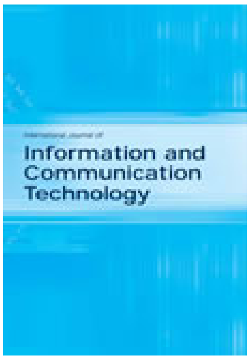


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Algorithm for interference filtering of Wi-Fi gesture recognition

Wei Han, Jing Yu

DOI: 10.1504/IJICT.2022.10045471

Article History:

Received:	26 August 2021
Accepted:	05 November 2021
Published online:	05 December 2023

Algorithm for interference filtering of Wi-Fi gesture recognition

Wei Han* and Jing Yu

School of Telecommunication Engineering,
Beijing Polytechnic,
Beijing, 100176, China
Email: hanwen@bpi.edu.cn
Email: yujing@bpi.edu.cn
*Corresponding author

Abstract: Nowadays with the continuous improvement of computer technology, human-computer interaction becomes more and more important in people's life. Among them, gesture, as an intuitive human language, has become an important way of human-computer interaction. Meanwhile, an important branch of mobile network, WLAN has gradually developed into an irreplaceable technology in indoor communication. This paper is based on the relevant knowledge of the wireless channel state information (CSI), put forward under the environment of Wi-Fi, with the help of fertility carrier amplitude information provided by the CSI for fine-grained gesture recognition. Because of interference, a gesture recognition system based on Wi-Fi accuracy and robustness is used to ascend; therefore, this paper proposes a Wi-Fi gesture recognition interference filter algorithm using Butterworth low-pass filter and principal component analysis (PCA), combining the CSI of raw data denoising processing and filtering CSI noise in the raw data. On the basis of the recognition and machine learning, the results verify the robustness and accuracy of the algorithm.

Keywords: channel state information; CSI; noise and interference filtering; gesture; recognition; machine learning.

Reference to this paper should be made as follows: Han, W. and Yu, J. (2024) 'Algorithm for interference filtering of Wi-Fi gesture recognition', *Int. J. Information and Communication Technology*, Vol. 24, No. 1, pp.1–20.

Biographical notes: Wei Han is an Associate Professor at Beijing Polytechnic and has Master degree in Electrical Automation by Southwest Jiaotong University. Her current research interests focus on electronic information engineering technology, big data technology and application, artificial intelligence application technology.

Jing Yu is a Professor at Beijing Polytechnic and has PhD degree in artificial intelligence pattern recognition by Beijing Jiaotong University. His current research interests focus on image processing, intelligent systems, machine learning, emotion recognition, multi-source information fusion, and natural language processing.

1 Introduction

Wireless local area network (WLAN) can be used not only as the communication medium, but also to perceive the ambient changes. In order to improve the validity and reliability of communications, the Wi-Fi system has employed the channel measurement mechanism, or the CSI or RSSI for description of the information channel. The Wi-Fi technology based on the 802.11n protocol adopts multiple input multiple output (MIMO) and orthogonal frequency division multiplexing (OFDM) technologies. It has 56 subcarriers, and uses multiple receiving and sending antennas. Meanwhile, a series of signal processing algorithms is used to analyse the CSI. The human movement or environmental changes can be analysed in accordance with the impact of the human body or environmental changes on the information channel. Therefore, CSI data measured by the Wi-Fi system can be used for perception applications. Under the indoor Wi-Fi environment, wireless signals are sent from the emitting terminal to the receiving terminal to form multipath overlapped signals (Siamak et al., 2017; Wang et al., 2016). If the channel state changes in the indoor can be learned on a real-time basis, then some artificial interferences (such as hand gestures and movements of humans) can be added to extract data characteristics of relevant parameters from the changing channel state information (CSI) (Behera et al., 2020; Wei et al., 2015). In this way, the human activities can be recognised. Currently, the wireless non-contact sensing and recognition technology has been a general research interest. For example, the WLAN positioning technology based on wireless frequency proposed by the University of Maryland (Youssef et al., 2003), the non-invasive human body movement static state detection obtained by Liu Yunhao's team based on CSI (Wu et al., 2015), the human behaviour recognition system built by Siamak et al. (2017) based on CSI data flow and deep learning model (Halperin et al., 2011), WiFinger (Sheng and Jie, 2016), RASID and PADS, the gesture recognition system realised based on CSI and multidimensional dynamic time domain scanning algorithm detect the human activities indoor via the intensity of the receiving signals and CSI (Kosba et al., 2012), respectively. Gesture is an extremely simple, natural and convenient means of human-machine interaction, which can play an efficient role in motion sensing games, gesture recognition or intelligent home living systems. Currently, the commonly-seen gesture recognition is realised via the computer visual technologies or various sensors. The gesture recognition technology based on the vision, because of its involvement of privacy, is vulnerable to the influence of the environment and rays of light, and might be inapplicable in some occasions. The gesture recognition technology based on the sensor is inconvenient in carrying and using, which is not beneficial for large-scale use and deployment.

Earlier research findings concerning Wi-Fi (Seifeldin et al., 2013) study attenuation characteristics of Wi-Fi signals and use these characteristics for the human positioning. There are also some scholars interpreting the human behavioural information based on research into reflected signals from humans (Kleisouris et al., 2010). In Zhou et al. (2018) perceives and recognises human activities in light of the received signal strength indicator (RSSI) acquired by the Wi-Fi technology. Though the RSSI is easy to acquire, the coarse granularity of the RSSI has decided its limited application to identify obvious activities at a low identification accuracy. Due to limitations of the RSSI, attempts have been made to explore the CSI with a finer granularity to perceive human activities. Recent years have witnessed some scholars' use of CSI to perceive the human walking direction, respiration, heartbeat, and other movements as well as daily activities. In Liu et al. (2015)

detects the falling behaviours by learning the designated CSI models. In Wang et al. (2017) detects the walking activities via the CSI mobile variance, because the CSI amplitude significantly changes with time upon walking. In Wang et al. (2014) seeks interior positioning by taking a look into the correlation between the CSI and the communication distance. In Yang et al. (2014) analyses the CIS data changes and the number of the mobile population to successfully perceive the crowd.

Currently, the wireless sensing technology based on Wi-Fi displays the multipath communication state of the wireless signals (Nan et al., 2018), including perpendicular incidence and reflection, between the receiving terminal and the sending terminal through the CSI data that are collected, and analyse the correlation between the communication state and the ambient changes. However, due to the existence of interference, the accuracy and robustness of the Wi-Fi-based gesture recognition system are calling for further improvement.

However, there are noises existing in CSI raw data. Even after being processed by the existing interference filtering methods, CSI data are still found with many residual noises. If these data are directly applied to the follow-up Wi-Fi gesture identification system, the system accuracy will be impaired. According to the existing interference filtering methods, a low-pass filter is used to remove the high-frequency noises, or an outlier is used to eliminate the abnormal value. WiFinger uses the outlier and a low-pass filter to eliminate the abnormal value and high-frequency noises (Bi et al., 2018). WiGeR uses a Butterworth low-pass filter to remove the high-frequency noise components in CSI data (Xie et al., 2015). Researchers from colleges such as Stevens Institute of Technology, while detecting vital signs of humans when sleeping, adopted a Hampel filter and a moving average filter to remove the abnormal value and noises from CIS raw data (Ming and Bors, 2008). WiWho uses the Butterworth low-path filter to retain signals within 0.3 hz. WiDraw uses a low-pass filter to realise denoising of the raw data (Fazli and Moeini, 2016). However, noise components still remain in CSI data after the low-pass filter and abnormal value processing.

In gesture identification, Widar3.0 system introduces the deep neural network based on the body coordinate speed information, which can achieve a high cross-field identification precision under different environments through one training only.

2 Related technology

2.1 CSI perception model

CSI belongs to information of the physical layer, which is used to achieve equilibrium of information channels in the Wi-Fi system, and it describes the impulse response of information channels. Yue et al. (2019) and Anmol et al. (2011) from Tsinghua University introduces basic perception models based on the RSSI and CSI:

The channel impulse response (CIR) model is used to describe the multipath communication information channel. Under the interior Wi-Fi communication environment, the information channel is a typical multipath information channel, which can be described by the multipath information channel model based on the linear hypothesis. The information CIR can be written as below:

$$h(\tau)=\sum_{i=1}^{N} a_i e^{j\theta_i}\delta\left(t-\tau_i\right) \tag{1}$$

where, denotes the number of paths; a_i, θ_i and τ_i denote the corresponding amplitude attenuation, phase deviation and communication time delay of the i path, respectively.

The Fourier transform of the CIR is the channel frequency response (CFR), which describes the multipath information channel from the perspective of the frequency selective fading, and it contains the information the same to that in CIR. The CRF model can be written as below:

$$H(f)=a(f)e^{j\theta(f)} \tag{2}$$

where, $a(f)$ denotes the amplitude-frequency response, while $\theta(f)$ denotes the phase-frequency response.

The CSI data adopted by the Wi-Fi network card for the commercial use are the down-sampling of CFR at every subcarrier frequency. The CSI data collected from every receiving antenna can be written as below:

$$H(i,k)=a_i(f_k)e^{j\theta_i(f_k)} \tag{3}$$

where, i denotes the number of the antenna pair; k denotes the number of subcarriers; f_k denotes the frequency of the subcarrier whose serial number is k; $H(i, k)$ denotes the CSI value measured by the k subcarrier from the i antenna, and the value is denoted by a real number.

Because of the significant multipath effect under the interior Wi-Fi communication environment, the CSI data collected from the network card are overlapped effects of various reflection paths, including the direct communication path, human body reflection, wall reflection, human body and wall secondary reflection, furniture reflection, etc. How to extract and analyse dynamic multipath component characteristics reflected by the human body from the chaotic multipath signals has been one of primary challenges facing research of the kind.

Figure 1 Schematic diagram of multiple receiving antennas

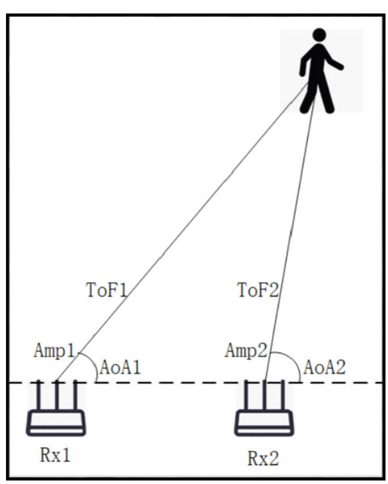


Figure 2 Multi-carrier schematic diagram

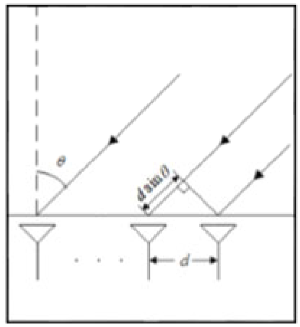
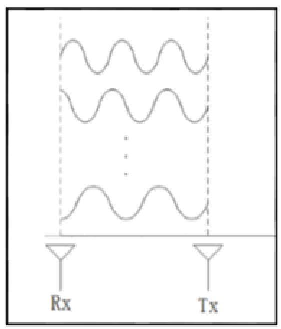


Figure 3 Schematic diagram of multiple receiving devices



Every independent path from the multipath signals can exert different degrees of impact on the CSI data array at different dimensions, including different receiving and sending device pairs, different antennas and different carriers. The direct communication path or the AoA information of the reflective wave from the same object is different between receiving and sending devices located at different spatial positions. The same movement tracks of the human body can also cause different changes of the phase shift introduced to the human body reflection path. As shown in Figure 1, signals either from the sending source or from the reflection source can reflect different AoA, ToF, Amp and other characteristics on different receiving devices. As shown in Figure 2, the phase difference between different antennas of the same sending and receiving device pair from the same path will change with the signal, AoA. As one observes in Figure 3, the phase difference between different subcarriers can reflect the phase shift difference caused by the spatial communication at different frequencies. Therefore, the multidimensional CSI data array can be analysed to realise the extraction and analysis of the dynamic multipath signal components reflected by the human body.

2.2 Acquisition of CSI data

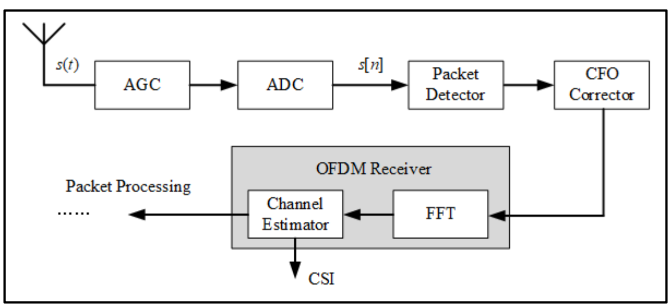
Halperin et al. (2011) developed the CSI tool applicable to the Intel 5300 NIC. By modifying the network card drive, the CSI tool can activate a debugging model to report the CSI data to the core of the operation system, thus allowing users to process the CSI data. Data collection can be realised through the installation of the Intel 5300 NIC and modification of the Linux system PC of core. Meanwhile, this CSI tool can operate under the CSI measurement Monitor model between the network cards, and the CSI measurement AP model between the network cards and routers.

Atheros CSI Tool developed by Wei et al. (2014) can help the 802. 11n Qualcomm Atheros wireless network card to collect all the 56 subcarriers in a similar way. The quantitative bit depth of the real part and virtual is both the 10-bit CSI data.

2.3 CSI noises

The real network card, while receiving and processing signals, will introduce multiple kinds of noises. So, the amplitude and phase from the CSI data collected by us from the network card tentatively contain various noises, which poses a tremendous challenge to our data processing. The CSI computing flow in the network card as proposed by Wu et al. (2013) is shown in Figure 4. The fundamental frequency signal $s(t)$ after going through the mixer and filter is converted into the digital signal $s[n]$ through the automatic gain control (AGC) and analogue-to-digital conversion (ADC). The digital signal is then processed by the packet detector to detect the arrival of the packet lead code. When the arrival of the packet is detected, the CFO corrector will calibrate the central frequency, and then send the signals to the OFDM receiver for modulation and calculation, through which the CSI data can be obtained.

Figure 4 Calculation flow schematic diagram of CSI data in the wireless network card



Source: Xie et al. (2015)

The amplitude noises in the CSI data are mainly caused by AGC. Since the adjustment granularity of the digital AGC used by the Wi-Fi network card is relatively high, the amplitude of the practical receiving signal cannot be made up in the CSI computing.

The phase noises in the CSI data mainly consist of the packet boundary detection (PBD) error, sampling frequency offset (SFO) error and central frequency offset (CFO) error. The mathematical model of the CIS data phase noises can be written as below:

$$\varphi_k = \theta_k + k \bullet (\lambda_b + \lambda_o) + \beta \tag{4}$$

where, k denotes the subcarrier number; φ_k denotes the actually-accepted CSI numerical phase; θ_k denotes the phase caused by the signal while propagating in the space; λ_b denotes the PBD error; λ_o denotes the SFO error of the sampling frequency; β denotes the CFO error of the central frequency.

Considering the desynchronisation of the receiving and sending devices, the time deviation τ_o which is constant within a short period of time is introduced to the module conversion process. Then, the phase deviation can be written as $\theta_{sfo} = 2\pi(f - f_c)\tau_o = 2\pi \bullet k \bullet \Delta f \bullet \tau_o = k\lambda_o$, where k is the subcarrier number. The PBD error in the CSI matrix data is reflected as the phase shift related to the subcarrier number that is almost constant over several minutes.

The PBD error is introduced during the PBD. The correlator sensitivity to be matched with the packet lead code leads to the introduction of a random time deviation i τ_b, which is reflected as the random phase deviation, $\theta_{pbd} = 2\pi(f - f_c)\tau_b = 2\pi \bullet k \bullet \Delta f \bullet \tau_b = k\lambda_b$. According to Ruppert et al. (2012), λ_b obeys the zero-mean Gaussian distribution. The PBD error in the CSI matrix data is reflected as the zero-mean Gaussian distribution phase deviation that is related to the subcarrier number, the same between antennas, and random between frames.

3 Interference filtering algorithm

Affected by the external ambient factors, the CSI raw datasets that are collected cannot avoid noises. Therefore, before using the CSI for gesture recognition, it is necessary to first establish an interference filtering mechanism to realise denoising processing of CIS raw data to improve the system robustness.

3.1 Shortages of current interference filtering methods

In response to noises in CSI raw data, the current interference filtering methods generally use a lowpass filter to remove the high-frequency noises or use an outlier to cope with or even remove the abnormal value. But after being processed by the current interference filtering methods, CSI datasets might still be left with many noises. If these datasets are directly applied to the follow-up Wi-Fi gesture recognition system, the system's recognition might be affected.

Concerning shortages of the current interference filtering methods, this paper proposes an interference filtering algorithm, according to which the Butterworth lowpass filter and the main component analysis method are combined to realise denoising of the CIS raw data. Since the background noises in CSI signals are concentrated in the high-frequency part, Butterworth lowpass filter is first used to remove the high-frequency noise components. In order to address the remaining noises in CSI data after lowpass filtering, this paper makes use of the correlation between carrier waves to remove residual noises.

3.2 Butterworth lowpass filter

Butterworth lowpass filter is an electronic filter which has found wide applications in the communications field. The frequency response curve of its transmission bands is the smoothest. In other words, Butterworth lowpass filter technology possesses the largest flatness frequency response. Therefore, in the primitive CSI data denoising process, this paper chooses Butterworth lowpass filter.

3.3 Filtering based on the principal component analysis (PCA)

Butterworth lowpass filter can be used to remove high-frequency noises in CSI raw data. Because of the minimum stopband attenuation existing in Butterworth lowpass filter, CSI data after going through the filter might still be found with noises. In this paper, the PCA is used to remove the residual noises in CSI data based on the correlation between carrier wave data of CSI. This process is realised mainly through the correlation among various carrier waves of CSI.

3.4 Denoising based on the PCA

The PCA is a statistical method. A group of variables which might be correlated can be converted into a group of linearly irrelevant variables through orthogonal transformation. This group of variables after transformation is known as principal components (Billah and Waheed, 2020). Many research fields require many data. Meanwhile, multiple dimensions which can reflect the attributes of things should be largely observed to find out the rules behind. However, the multidimensional large samples, while providing information for analysis, will, to some extent, increase the workload because of many high-dimensional samples. Additionally, the correlation among various variables can also cause inconvenience to the analysis process. If various indexes were separately analysed, the analysis results might lack completeness. But if the indexes were reduced blindly, some relevant information might be lost, making it hard to obtain the desired conclusions. Therefore, use of PCA can fully maintain the primitive data characteristics, and obtain principal components to replace the raw data via the transformation of the linear combinations. Various principal components thus obtained are irrelevant. Therefore, PCA usually find applications in data denoising or data dimension reduction. The principal components obtained through the PCA will be aligned according to the eigenvalue. The smaller the serial number is, the more the data characteristic information is included. Therefore, this paper adopts the PCA to reduce the sub-carrier-wave dimensions and remove the residual redundant information in CSI after Butterworth filter. So PCA can also help with denoising.

3.4.1 CSI data matrix, H.

First of all, this paper chooses an antenna pair of CSIs to obtain the CSI data matrix among various sub-carrier-waves:

$$H = \begin{bmatrix} H(1,1) & \cdots & H(1,56) \\ \vdots & \ddots & \vdots \\ H(Len,1) & \cdots & H(Len,56) \end{bmatrix} \tag{5}$$

where, CSI data matrix, *H*, contains a continuous number (*Len*) of samples collected by one antenna pair. Every sample composes of CSI data with 56 OFDM sub-carrier waves. Therefore, the $Len*56$ matrix, $H(i, j)$, denotes the CSI data value of the *j* sub-carrier-wave of the *i* sample. Therefore, every column in *H* denotes the CSI temporal sequence whose length is *Len*.

3.4.2 CSI standardised matrix, Z

Before obtaining the principal component of CSI, it is necessary to conduct demeaned processing of *H*. Following that, normalisation is carried out to obtain the correlation coefficient matrix, *Z*. The equation for the demeaned processing and normalisation can be written as equation (6):

$$Z_{i,j} = \frac{h_{i,j} - \overline{h_j}}{\overline{s_j}} \tag{6}$$

where, $Z_{i,j}$ and $h_{i,j}$ are $H(i, j)$ in the CSI data matrix and $Z(i, j)$ in the normalised matrix; $\overline{h_j}$ and $\overline{s_j}$ are the mean and covariance of the *j* column in the matrix, *H*. Therefore, the normalised form, *Z*, of the CSI matrix can ensure the mean of every CSI flow to be 0, and the standard deviation (STD) to be 1.

3.4.3 Covariance matrix of Z

The CSI normalised matrix, *Z*, can be used to compute the covariance matrix whose dimension is 56*56. The covariance matrix equation for a random variable set, $\{X_1,\ X_2,\ \cdots,\ X_n\}$, can be written as below:

$$R = \begin{bmatrix} \operatorname{cov}(X_1, X_1) & \cdots & \operatorname{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n, X_1) & \cdots & \operatorname{cov}(X_n, X_n) \end{bmatrix} \tag{7}$$

where, cov(*X*, *Y*) denotes the covariance between the random variables, *X* and *Y*, which can be used to measure the statistical value between two random variables. The statistical value can be given by equation (8) below:

$$\operatorname{cov}(X, Y) = \frac{\sum_{i-1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \tag{8}$$

3.4.4 Characteristic decomposition

After the covariance matrix, *Z*, is deduced via the CSI normalised matrix, the correlation coefficient matrix of CSI data can be obtained. The corresponding eigenvalue after the characteristic decomposition of the correlation coefficient matrix can be obtained via characteristic decomposition. Following that, ten maximal eigenvalues are selected, and the ten eigenvectors of the ten eigenvalues constitute the eigenvector matrix, *Q*, whose dimension is 56*10.

3.4.5 Reconstruction of the CSI matrix

The eigenvector matrix, *Q*, obtained above can be used to reconstruct the PCA matrix, $Z^{\{1:10\}}$, of the CSI matrix, *Z*:

$$Z^{\{1:10\}} = Z * Q \tag{9}$$

where, every column of the PCA, $Z^{\{1:10\}}$, includes the CSI principal components mapped by the CSI raw data matrix, H.

As mentioned above, the noises existing in the data value of various sub-carrier-waves are significantly correlated. The first principal component contains a majority of noises. Though the first principal component contains the characteristic information generated by the gesture, the characteristic information might still exist in the other nine principal components. Besides, various principal components obtained after PCA are independent from each other. Hence, PCA can effectively remove the noises in the first principal component under the condition of not destroying the characteristic information.

It can be seen that PCA can effectively remove the residual noises after lowpass filter under the prerequisite of not destroying the characteristics information. In this way, the interference caused by residual noises after the use of the current interference filtering methods can be addressed. Thus, the first principal component is deleted to obtain the matrix, $Z^{\{2:10\}}$. Regarding the characteristics of MIMO, the CSI data of nine antenna pairs can be obtained. Every antenna pair also include the data of 56 sub-carrier-waves. So, use of PCA can help process CSI data. On the one hand, the dimension-reducing function of PCA can be fully displayed to effectively bring down the CSI data processing amount. On the other hand, the correlation among different sub-carrier-waves in CSI data can be referred to for effective filtering of the residual noises after Butterworth lowpass filter. Therefore, PCA can help filter play a significant role in the interference filtering mechanism proposed by this paper.

4 Wi-Fi-based gesture recognition model

The CSI data obtained after interference filtering can be used to train the Wi-Fi-based gesture recognition model via machine learning. The CSI data obtained after the interference filtering algorithm proposed above are still a time series, which cannot be classified directly as characteristics. At the same time, the waveforms generated by different gestures are different. So, concerning CSI data waveform changes arising from gesture changes in this paper, the waveform is adopted as the gesture characteristic.

4.1 Characteristic extraction

Characteristic extraction is mainly realised through the relationship among attributes. For example, combination of different attributes can obtain new attributes. This can change the original characteristic space. The waveform generated by the gesture is different from each other. Therefore, this paper focuses on analysing the characteristics from the waveform perspective. Because the large number of CSI waveform data can increase the system operation cost, this paper first calculates the mean, mode, median, range, STD, mean absolute deviation (MAD), max and min of the CSI amplitude under different gestures, respectively, as the eigenvalue of each gesture on the time domain, and are listed as several standards to recognise the gesture from the statistical perspective. Among them, the mean, mode, median, median and range can reflect the integrated state of CSI wave amplitude; the STD reflects the degree of dispersion of CSI waveform amplitude; the MAD reflect the degree of deviation of CSI waveform amplitude; the max and min can reflect the extreme values of gesture waveforms.

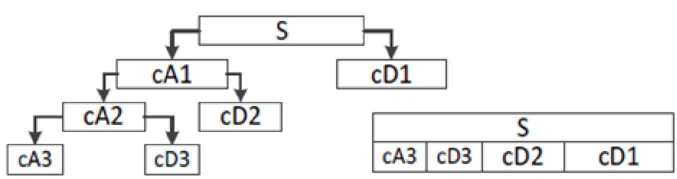
Besides, in order to reduce the computing cost caused by the large number of CSI waveform data, this paper adopts the discrete wavelet transform to process signal waveforms. To retain the approximation coefficient after wavelet transform as the eigenvalue of the gesture can retain not only the useful information of the characteristic waveform on the time domain and the frequency domain, but also reduce the computing amount of CSI data and improve the performance and efficiency of gesture recognition system.

The discrete wavelet transforms or DWT is to realise discretisation of the basic wavelet scale and translation according to the power series. It is a commonly-used time-scale (time-frequency) analysis approach. Therefore, DWT can well reflect its local characteristics interns of the time domain and frequency domain, and can realise multilevel resolution. Through DWT, the high-frequency part can obtain a high time resolution ratio and a lower frequency resolution ratio. The low-frequency part is to the opposite. Therefore, DWT is suitable for local signal feature extraction.

Use of DWT for characteristic extraction can, in addition to reducing the computing cost brought by the large number of CSI waveform data, avoid damaging the shape characteristics of gesture waveforms. Therefore, it is necessary to choose a suitable wavelet basis function to improve the system performance and accuracy. The commonly-seen wavelet base functions include Daubechies (dbN), Haar, Sym1ets (symN) and so on. Through experimental verification. This paper chooses db4 in Daubechies as the wavelet base function of DWT. This is a wavelet base function that is approximately symmetric, which can be written as dbN (N = 1, 2, 3, …, 10).

Figure 5 shows four principal component waveforms from the second one to the fifth one chosen after PCA filtering. The db4 wavelet base function is used for the three-layer DWT decomposition of the four waveforms. In this way, the waveform characteristic information of gestures can be maintained, and the CSI characteristic waveform data size can be compressed maximally. The detail coefficient, cD, obtained after DWT includes the high-frequency part of the wavelet, while the approximation coefficient, cA, includes the low-frequency component of the waveform.

Figure 5 Three-layer DWT decomposition of db4 wavelet transform



After three decompositions of the second principal component waveform, the low-frequency part of the waveform can well maintain the waveform characteristic, but the high-frequency components constitute a section of chaotic waveforms, which cannot be used for the classification of follow-up gesture recognition systems. Therefore, this paper removes the detail coefficient of the waveform, cD1 to cD3, and maintain the approximation coefficient, cA3, as the recognition characteristic. At the same time, the CSI waveform data are effectively compressed, and the extracted characteristics can be applied to the follow-up classifier.

After characteristic extraction, this paper extracts 17 eigenvalues from every principal component, including eight parameters (mean, mode, median, range, STD, average absolute deviation, maximum and minimum), and nine approximation coefficients obtained after DWT. Besides, the gesture waveform characteristics of four principal components after filtering constitute a characteristic space containing 68 eigenvalues. See Table 1:

Table 1 Characteristic extraction

Classification of eigenvalues	*Chosen eigenvalues*	*Number*
Eigenvalues in statistics	Mean, mode, median, range, standard deviation, average absolute deviation, maximum, and minimum	8 * 4 = 32
Eigenvalues after DWT	Approximation coefficient	9 * 4 = 36

4.2 Characteristic selection

After characteristic extraction, 68 eigenvalues are obtained. In order to select better characters from the original characteristic data se for the construction of the machine learning model, the chosen characteristics should be capable of providing more accurate description for the model. Therefore, the characteristic selection aims at improving the prediction accuracy and constructing a machine learning model at a lower operation cost which can obtain a better explanation of the model. The commonly-seen characteristic selection steps are show as below.

- First, understand the problem-based model, and assess all independent variables that affect dependent variables.
- Second, after the independent variables are chosen, the characteristics should be selected.
- This paper uses the minimum redundancy maximum relevance (mRMR) (Abd and Abd, 2017) to finish characteristic selection of 68 characteristics.

4.2.1 Maximum relevance

Assume that *S* is a collection of characteristics, $\{X_1, \quad X_2, \quad \cdots, \quad X_n\}$. According to the following equation, n most relevant characteristics are chosen to ensure S to meet the requirement of maximal relevance of characteristics and categories:

$$\max D(S, c) = \max \frac{1}{S} \sum_{x_i \in S} I(x_i; c) \tag{10}$$

where, *S* denotes the collection of chosen characteristics; x_i denotes the *i* characteristic; *c* denotes the label of the category. It can be seen that the objective is to choose the collection, *S*, with n average mutual information that is the maximal.

4.2.2 Minimum redundancy

The above collection, *S*, might contain characteristics with a high degree of relevance, which means there is redundancy among characteristics. The redundancy of the collection, *S*, can be written as equation (11) below:

$$\min R(S) = \min \frac{1}{|s|^2} \sum\nolimits_{x_i, x_j \in S} I(x_i; x_j) \tag{11}$$

The ultimate objective is to work out the collection, S, of the mRMR through equation (12):

$$\max \Phi(D, R) = \max(D - R) \tag{12}$$

when D increases, R decreases, and the objective function will increase accordingly. In practice, this paper uses incremental search to seek the approximately optimal characteristic. If there is/are already $(n - 1)$ characteristic(s) constituting the characteristic collection, S_{n-1}, then it is necessary choose n characteristic from the remaining characteristics, $X - S_{n-1}$, to ensure the $\Phi(D, R)$ to be maximal. As shown in equation (13):

$$\max_{x_j \in X - s_{n-1}} \left(I(x_j; c) - \frac{1}{n-1} \sum\nolimits_{x_i \in s_{n-1}} I(x_j; x_i) \right) \tag{13}$$

mRMR algorithm is used to choose 18 characteristics with the highest relevance to the classification category. In other words, the relevance between the characteristics and the classification category is maximised, while the relevance between characteristics is minimised, thus reducing the redundancy of characteristic variables. Through characteristic selection, the eigenvalues extracted from the former section can be simplified to obtain a simplified characteristic dataset with better classification effects, which can improve the classification effects of the classifier in the following part, and effectively reduce the dimensionality of characteristics for the improvement of the system operation performance.

4.2.3 Classifier selection

Classification is to build the model through machine learning and according to the classification knowledge that is already known. The input unknown model is analysed to identify the category of the input model. The role of the classifier is to make use of training data that belong to the category already known to learn classification rules and classifiers. After that, the unknown data are classified or predicted. This paper chooses the support vector machine (SVM) as the gesture recognition classifier. The SVM is a classifier designed on the basis of statistics learning theories, and it belongs to supervised learning.

4.3 Support vector machine (SVM)

The basic form of the SVM can be written as below (Wang et al., 2016),

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 \quad s.t. y_i \left(\omega^T x_i + b \right) \geq 1, i = 1, 2, \cdots m \tag{14}$$

where, $\omega = \{\omega_1; \omega_2; \cdots \omega_d\}$ denotes the normal vector, which represents the direction of the division super-plane; b denotes the distance between the division super-plane and the original point. ω and b can decide the division super-plane, which can be written as (ω, b). The equation above revolves around a convex quadratic programming question.

Therefore, use of Lagrangian multiplier method can solve this question, which is to add the Lagrangian multiplier, $\alpha_i \geq 0$, to every constraint. So, the Lagrangian multiplier of the above question can be written as below:

$$L(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2 + \sum_{i=1}^{m} \alpha_i \left(1 - y_i \left(\omega^T x_i + b\right)\right) \tag{15}$$

where, $\alpha = \{\alpha_1; \alpha_2; \cdots \alpha_m\}$. After α is solved, the parameters, ω and b, are solved. Then, the model can be obtained:

$$f(x) = \omega^T x_i + b = \sum_{i=1}^{m} \alpha_i y_i x_i^T x + b \tag{16}$$

The above solution process should meet the condition of Karush-Kuhn-Tucker (KKT):

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(x) - 1 \geq 0 \\ \alpha_i \left(y_i f(x) - 1\right) = 0 \end{cases} \tag{17}$$

The category of the case, R, can be judged to solve $f(R)$,

$$f(R) = \omega^T R + b = \sum_{i=1}^{m} \alpha_i y_i x_i^T R + b \tag{18}$$

when $f(R)$, the case, R, belongs to the positive category; otherwise, it belongs to the negative category.

4.4 Nuclear function and parameter settings of SVM

When the linearly inseparable dataset is handled, the samples can be converted to a space of a higher dimension via the original space to ensure the samples to be linearly separable. Assume that $\Phi(x)$ is the eigenvector of x after mapping. In the characteristic space, the corresponding model of the division super-plane can be written as below:

$$f(x) = \omega^T \Phi(x) + b \tag{19}$$

The dual problem can be written as below in accordance with equation (20):

$$\begin{aligned} &\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \Phi(x)_i^T \Phi(x_j) \\ &s.t. \sum_{i=1}^{m} \alpha_i y_i = 0 \\ &\alpha_i \geq 0, i = 1, 2, \ldots, m \end{aligned} \tag{20}$$

Based on $\kappa(x_i, x_j) = <\Phi(x_i), \Phi(x_j)> = \Phi(x)_i^T \Phi(x_j)$, the dual problem can be solved:

$$\begin{aligned} f(x) = \omega^T \Phi(x) + b &= \sum_{j=1}^{m} \alpha_i y_i \Phi(x_i) + b \\ &= \sum_{j=1}^{m} \alpha_i y_i \kappa(x_i, x_j) + b \end{aligned} \tag{21}$$

where in the equation (21), $\kappa(\bullet,\bullet)$ denotes the kernel function. Since there are a few parameters involved in the radial basis function (RBF), this paper adopts RBF.

5 Experimental process and result analysis

The experimental environment is a one-person indoor environment, which can avoid other interference. This paper makes use of Atheros-CSI-Tool on the terminal to implement and receive the recv-CSI command and to realise exchange with AP in terms of its data package. The interval between AP and DP is around 0.6 m. The experimenter can have gesture operation between AP and DP. Four gestures in total are collected, and CSI data package without gestures is collected. The five data packages collect data for 200 times. The CSI data package of the odd number is chosen as the training set, while the CSI data package of the even number is adopted as the test set.

5.1 Effectiveness verification of interference filtering algorithm

Denoising by wavelet transform. The following results can be obtained

Figure 6 Approximate component of the ninth layer of wavelet transform (see online version for colours)

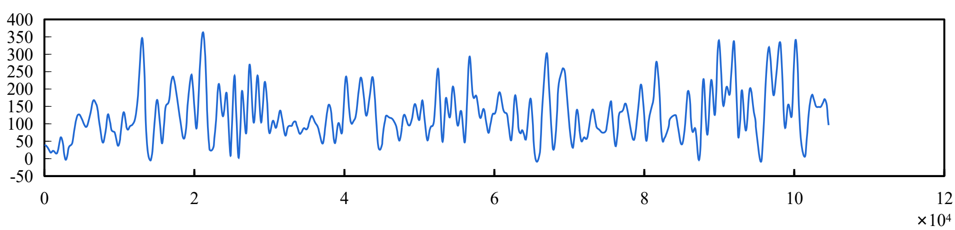


Figure 7 Approximate components of the seventh and eighth layers (see online version for colours)

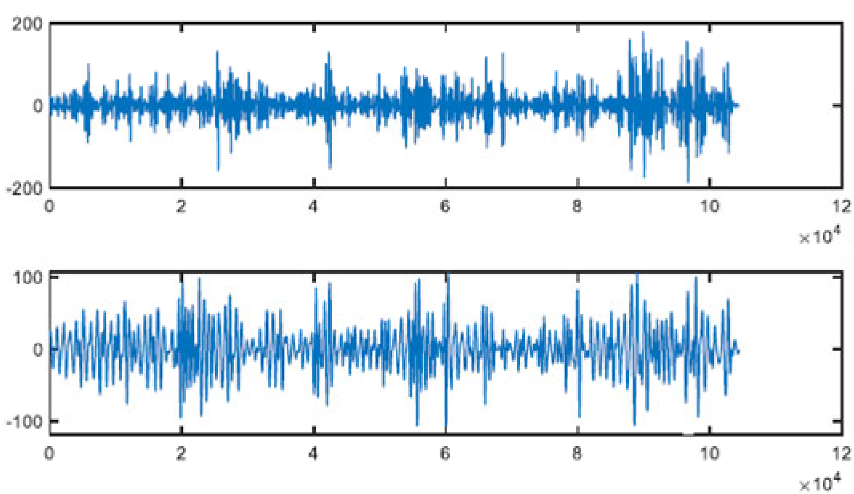


Figure 8 Detail component of the seventh and eighth layers (see online version for colours)

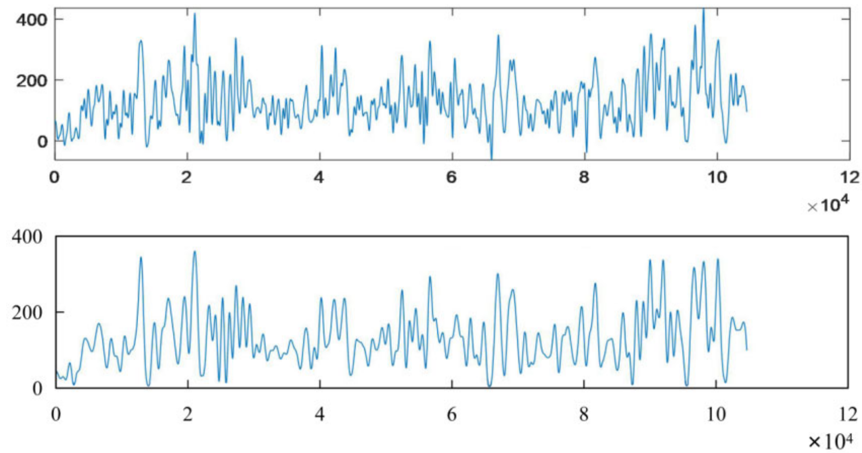
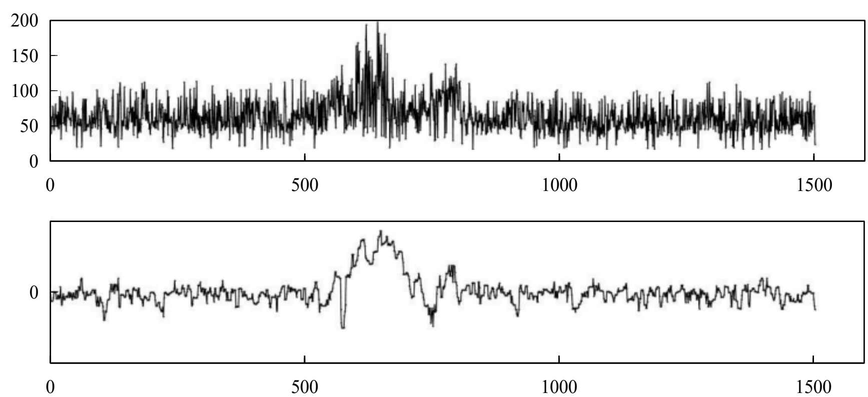


Figure 7 and Figure 8 show the signal denoised by the wavelet filter using DB wavelet for the principal component. It can be seen from the figure that both the low-frequency components of the eighth layer decomposition and the ninth layer decomposition can obviously show the introduction of Doppler frequency shift. Compared with the low-frequency components of the eighth layer decomposition, the low-frequency components of the ninth layer decomposition have significantly better noise reduction effect.

The validity of the interference filtering algorithm is verified. Figure 9 shows the waveform of the CSI raw data and the CSI waveform of the second principal component under the joint function of the Butterworth filter and principal component filter.

Figure 9 Front and back waveform oscillography before and after CSI filtering under the gesture of sliding leftwards

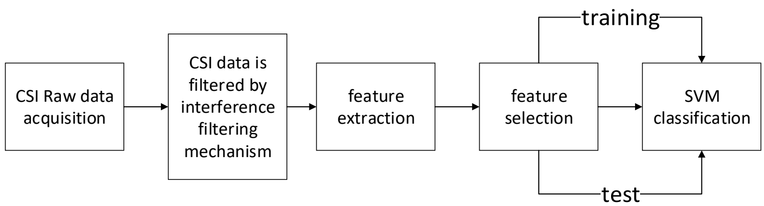


As shown in Figure 9, before use of the interference filtering algorithm, the CSI raw data waveform is found with a series of noises. If the CSI raw data are directly applied to the Wi-Fi gesture recognition system, the system accuracy might be significantly impaired. Therefore, this paper combines Butterworth filter and the principal component filter to realise denoising of the CSI raw data waveform. It can be seen that the second principal component waveform of PCA after denoising, compared with the original oscillography, has not only maintained the gesture waveform characteristics of the original oscillography, but also remove the noises existing in raw data.

5.2 Gesture recognition results

This paper collects CSI data, and uses the interference filter algorithm to realise filtering of raw data. After characteristic extraction and characteristic selection are completed, this paper divides the dataset into the training set and the test set, and uses SVM to train the gesture recognition model. After the model is obtained, the model obtained can be tested with the test set. The whole process is shown in Figure 10:

Figure 10 Gesture recognition system procedure



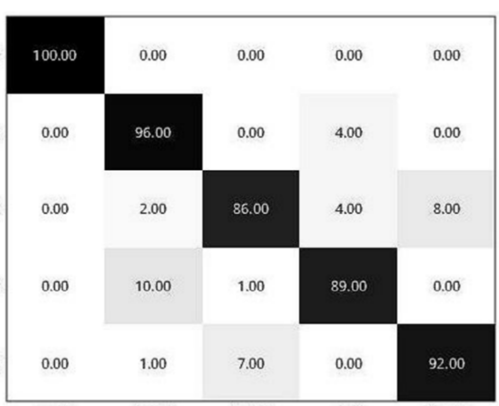
The test set of every category contains 100 pieces of gesture waveform data.

Table 2 Wi-Fi gesture recognition results

Slide leftwards	*Push forward*	*Shake fists*	*Cut downwards*	*Without gestures*	*Overall accuracy*
96%	86%	89%	92%	100%	92.6%

It can be seen that a high accuracy can be achieved under the condition of 'slide leftwards', 'shake fists', 'cut downwards' and 'without gestures', respectively. The identification rate of the gesture, 'push forward', is 86%, which is relatively low. The overall accuracy of the system is 92.6%, and the accuracy is generally acceptable. This verifies the robustness of the Wi-Fi-based gesture identification system.

Figure 11 Confusion matrix of gesture identification results



The confusion matrix is employed to present gesture identification results of the experiment. The x-coordinate shows the dataset type, while the y-coordinate shows the identification outcomes. In the confusion matrix, the deeper the colour is, the more accurate the identification outcomes are. From Figure 4, one can observe that the system has achieved a high identification rate on the whole, which provides solid evidence for the feasibility of the W-Fi-based gesture interference identification and filtering algorithm.

6 Conclusions

To sum up, this paper makes use of the OFDM multiple sub-carrier-wave amplitude information provided by the wireless CSI for fine-grit gesture recognition, and proposes an interference filtering algorithm for filtering of the noise interference in the filtering channel. Meanwhile, the Wi-Fi-based gesture recognition system is built. Through the experiment, the robustness and accuracy of the system are verified. First of all, concerning the shortages of the existing CSI data interference filtering methods, this paper combines Butterworth filter algorithm and PCA to not only remove the high-frequency noise components in the CSI raw data. On the other hand, based on the relevance of sub-carrier-waves, a large number of noises included in the first principal component are deleted to realise denoising of the CSI raw data and to improve the system accuracy. Besides, the CSI dataset after filtering is used for characteristic extraction and selection. Next, SVM is used as the classifier to train the gesture recognition model. The final test results verify the system robustness and accuracy. The overall accuracy of the gesture recognition system is as high as 92.6%. Nevertheless, the system still has room for improvement. First, the environment is full of various noises, thus resulting in the environmental instability, and bringing noises to data of the training set and test set. The interference filtering algorithm proposed by this paper can be further improved. Second, the environment is subject to the multipath interference, which can affect the CSI and further influence the system accuracy. In the future, different methods can be tried to inhibit the multipath interference so as to improve the system robustness. Third, more eigenvalues that are different can be extracted tentatively in an attempt to find out better characteristic space and improve the classifier design and performance. Fourth, while receiving the CSI data packets, this paper adopts the single receiving device. The future scholars can employ multiple receiving devices to improve the system accuracy.

Acknowledgements

This work was supported by the R&D Program of Beijing Municipal Education Commission (No. KM202110858001) and 2020Z012-KXZ.

References

Abd, A.M. and Abd, S.M. (2017) 'Modelling the strength of lightweight foamed concrete using support vector machine (SVM)', *Case Studies in Construction Materials*, Vol. 17, No. 6, pp.8–15.

Anmol, S., David, W., Wenjun, H. et al. (2011) 'Tool release: gathering 802.11n traces with channel state information', *Computer Communication Review: A Quarterly Publication of the Special Interest Group on Data Communication*, Vol. 41, No. 1, pp.53–53.

Behera, A., Wharton, Z., Liu, Y., Ghahremani, M. and Bessis, N. (2020) 'Regional attention network (RAN) for head pose and fine-grained gesture recognition', *IEEE Transactions on Affective Computing*, pp.1949–3045.

Bi, W., Li, X.Y., Yang, P., Li, G., Zhou, H., Li, Y. et al. (2018) 'Signal tracking using commodity WiFi', in *International Conference on Big Data Computing & Communications*, Chicago.

Billah, M. and Waheed, S. (2020) 'Minimum redundancy maximum relevance (MRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection', *Multimedia Tools and Applications*, Vol. 79, No. 1, pp.23633–23643.

Fazli, S. and Moeini, M. (2016) 'A robust image watermarking method based on DWT, DCT, and SVD using a new technique for correction of main geometric attacks', *Optik – International Journal for Light and Electron Optics*, Vol. 127, No. 2, pp.964–972.

Halperin, D., Hu, W., Sheth, A. et al. (2011) 'Tool release: Gathering 802.11 n traces with channel state information', *ACM SIGCOMM Computer Communication Review*, Vol. 41, No. 1, pp.53–53.

Kleisouris, K., Firner, B., Howard, R. et al. (2010) 'Detecting intra-room mobility with signal strength descriptors', in *Proceedings of the 11th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Chicago, IL.

Kosba, A.E., Saeed, A. and Youssef, M. (2012) 'RASID: a Robust WLAN device-free passive motion detection system', in *IEEE International Conference on Pervasive Computing and Communications*, Lugano, Switzerland.

Liu, J., Wang, Y., Chen, Y. et al. (2015) 'Tracking vital signs during sleep leveraging off-the-shelf WiFi', in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Hilton Head, SC, USA., pp.267–276.

Ming, L. and Bors, A.G. (2008) 'Principal component analysis of spectral coefficients for mesh watermarking', in *IEEE International Conference on Image Processing*, San Diego, CA. pp.441–444.

Nan, Y., Wei, W., Liu, A.X. and Kong, L. (2018) 'Qgesture: quantifying gesture distance and direction with WiFi signals', in *Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies*, Vol. 2, No. 1, pp.1–23.

Ruppert, G.C.S., Reis, L.O., Amorim, P.H.J. et al. (2012) 'Touchless gesture user interface for interactive image visualization in urological surgery', *World Journal of Urology*, Vol. 30, No. 5, pp.687–691.

Seifeldin, M. et al. (2013) 'Nuzzer: a large-scale device-free passive localization system for wireless environments', *IEEE Transactions on Mobile Computing*, Vol. 12, No. 7, pp.1321–1334.

Sheng, T. and Jie, Y. (2016) 'Fine-grained gesture recognition using WiFi', in *IEEE INFOCOM 2016 – IEEE Conference on Computer Communications Workshops* (*INFOCOM WKSHPS*), San Francisco, CA, USA, pp.257–258.

Siamak, Y., Hirokazu, N., Sankalp, D. et al. (2017) 'A survey on behavior recognition using WiFi channel state information', *IEEE Communications Magazine*, Vol. 55, No. 10, pp.98–104.

Wang, G., Zou, Y., Zhou, Z., Wu, K. and Ni, L.M. (2016) 'We can hear you with Wi-Fi!', *IEEE Transactions on Mobile Computing*, Vol. 15, No. 11, pp.2907–2920.

Wang, Y. et al. (2014) 'E-eyes: device-free location-oriented activity identification using fine-grained WiFi signatures, MobiCom 14', in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, New York, USA, pp.617–628.

Wang, Y., Wu, K. and Ni, L.M. (2017) 'WiFall: device-free fall detection by wireless networks', *IEEE Transactions on Mobile Computing*, Vol. 16, No. 2, pp.581–594.

Wei, W., Liu, A.X., Shahzad, M., Ling, K. and Lu, S. (2015) 'Understanding and modeling of WiFi signal based human activity recognition, MobiCom 15', in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, New York, USA, pp.65–76.

Wei, X. et al. (2014) 'Electronic frog eye: counting crowd using WiFi', in *IEEE INFOCOM 2014 – IEEE Conference on Computer Communications*, New York, USA, pp.361–369.

Wu, C., Yang, Z., Zhou, Z., Liu, X., Liu, Y. and Cao, J. (2015) 'Non-invasive detection of moving and stationary human with WiFi', *IEEE Journal on Selected Areas in Communications*, Vol. 33, No. 11, pp.2329–2342.

Wu, K. et al. (2013) 'CSI-based indoor localization', *IEEE Transactions on Parallel and Distributed Systems*, Vol. 24, No. 7, pp.1300–1309.

Xie, Y., Li, Z. and Mo, L. (2015) 'Precise power delay profiling with commodity Wi-Fi', *IEEE Transactions on Mobile Computing*, Vol. 18, No. 6, pp.1342–1355.

Yang, Z., Zhou, Z. and Liu, Y. (2014) 'From RSSI to CSI: indoor localization via channel response', *ACM Computing Surveys*, Vol. 46, No. 2, pp.25.1–25.32.

Youssef, M.A., Agrawala, A. and Shankar, A.U. (2003) 'WLAN location determination via clustering and probability distributions', in *IEEE International Conference on Pervasive Computing & Communications*, Fort Worth, TX, USA, pp.143–150.

Yue, Z., Zh, Y., K, Q., et al. (2019) 'Zero effort cross-domain gesture recognition with Wi-Fi', *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pp.313–325.

Zhou, Q. et al. (2018) 'From signal to image: enabling fine-grained gesture recognition with commercial Wi-Fi devices', *Sensors*, Vol. 18, No. 9, pp.3142–3163.