Noname manuscript No.
(will be inserted by the editor)

Human Behavior Sensing: Challenges and Approaches

Xiwen Liu[1] · **Haiming Chen**[1] · **Antonio Montieri**[2] · **Antonio Pescapè**[2]

Received: date / Accepted: date

Abstract In recent years, Activities of Daily Living Scale (ADLs) is widely used to evaluate living abilities of the patients and the elderly. So, the study of behavior sensing has attracted more and more attention of researchers. Behavior sensing technology is of strong theoretical and practical value in the fields of smart home and virtual reality. Most of the currently proposed approaches for tracking indicators of ADLs are human-centric, which classify activities using physical information of the observed persons. Considering the privacy concerns of the human-centric approaches(e.g. images of home environment, private behavior), researchers have also proposed some thing-centric approaches, which use environmental information on things(e.g. the vibration of things) to infer human activity. In this paper, by considering the unified steps in both the human-centric approaches and the thing-centric approaches, we make a comprehensive survey on the challenges and proposed methods to do behavior sensing, which are signal collection, preprocessing, feature extraction, and activity recognition. Moreover, based on the latest research progress, we post a perspective from our standpoint, discussing future outlook and challenges of human behavior sensing.

Keywords Activities of Daily Living Scale · Behavior Sensing · Human-centric Approaches · Thing-centric Approaches

Xiwen Liu
137303561@qq.com
✉Haiming Chen
chenhaiming@nbu.edu.cn
Antonio Montieri
antonio.montieri@unina.it
Antonio Pescapè
pescape@unina.it

[1]Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China
[2]Department of Electrical Engineering and Information Technology, University of Napoli Federico II, Napoli, Italy

1 Introduction

Aging population has become one of the main concerns in both developed countries and developing ones. According to a report from the World Health Organization (WHO) (2011), most countries are facing a situation in which the percentage of elderly persons is becoming larger than ever before. Persons may suffer from many kinds of diseases with a higher probability when getting older. Knowledge about a person's ability to undertake normal Activities of Daily Living (ADL) is an essential part of the overall assessment of the patients and the elderly and is important in determining the diagnosis and in evaluating change. In the Gerontological Society's recent contract study on functional assessment, a large assortment of rating scales, checklists, and other techniques in use in applied settings was easily assembled.

Activities of Daily Living Scale (ADLs) (Lawton et al. 1970)are general techniques for assessing function of older adults or those with physical disabilities, which are of paramount importance to evaluate the living abilities of the patients and the elderly, especially for those who need to be under medical control. There are many indicators of ADLs (Debes et al. 2016), such as leaving the house, using the toilet, taking a shower, going to bed, preparing dinner, using the fridge, making a phone call, getting a drink and so on. Traditionally, these indicators were usually evaluated by professional institutions by asking the involved persons to fill a questionnaire periodically, or requiring them to record their own activities manually and then collecting the recorded data into electronic forms. This method is not only inaccurate, but also obtrusive to the elderly or the patient's living. With the rapid development of sensing technology, many approaches have been exploited to make the ADLs assessment more objective and mitigate obtrusiveness for the assessed persons.

Currently one of the approaches widely adopted to track indicators of ADLs unobtrusively is recognizing human activities by behavior sensing technology, which extracts valuable signals that change with activity and classifies activities by exploiting some machine learning algorithms. The valuable signals triggered from human activities, such as accelerometer signal, WiFi signal, and visible light signal, for behavior sensing are usually very subtle. Moreover, the signals are affected by many factors, such as the surrounding environment, equipment placement, signal attenuation, etc. Thus, it is still extremely challenging for behavior sensing to accurately recognize activities.

At present, some reviews (Debes et al. 2016; Lawton et al. 1970; Spagnolo et al. 2014; Srivastava et al. 2012) have been done on behavior sensing technologies, which divided the existing approaches into four categories: vision-based, light-based, sensor-based, and WiFi signal-based approaches, being all human-centric. Differently, this paper divides behavior sensing technology into two categories, which are referred to as human-centric and thing-centric approaches respectively, for the first time. Briefly speaking, human-centric approaches sense human activities by analyzing the signals generated by these activities, whereas thing-centric approaches take into account the disturbance characteristics of signals caused by things. Aiming to provide useful guidance for future work on this area, by considering the common functional parts of both the human-centric and thing-centric approaches, we analyze the challenges and existing methods to realize behavior sensing from four aspects, which are signal sampling, signal preprocessing, feature extraction, and activity recognition.

The remainder of the manuscript is organized as follows. Section 2 introduces two categories of-behavior sensing in detail. Moreover, we define a general framework of behavior sensing comprising four aspects of both human-centric and thing-centric sensing. We further deepen these aspects and how they has been implemented by the recently proposed approaches in the successive sections, namely signal sampling in Section 3, signal preprocessing in Section 4, feature extraction in Section 5, and activity recognition in Section 6. Finally, based on the analyzed results, we present a perspective from our standpoint in Section 7.

2 Behavior Sensing

In this section, we firstly provide a taxonomy of behavior sensing techniques, discussing both human-centric and thing-centric approaches. Then, we show a framework of behavior sensing that is independent of the specific approach taken into account.

2.1 Taxonomy of behavior sensing

(1) **Human-centric approach:** Most of the existing approaches are human-centric (Srivastava et al. 2012), which are featured by classifying activities using physical information of the observed persons. The physical information can be collected with or without sensors attached to the observed persons. We refer to the approaches with attached sensors as invasive/positive recognition, while calling the approaches without attached sensors as non-invasive/passive recognition.

Some representative existing works on these two types of human-centric sensing are listed in Table 1. As proposed by Fujinami et al. (2011), Keally et al. (2011), Lee et al. (2015), Galluzzi et al. (2015), and Chikhaoui et al. (2018), sensors (e.g. RFID sensors, body sensor networks, accelerometer in microphone and smart watch, etc.) were carried by the persons, and used to collect information about human activity. On the contrary, the approaches proposed by Erickson et al. (2011), Lao et al. (2009), Wilson et al. (2011), Wu et al. (2015), Yang et al. (2017), Nguyen et al. (2018), Liu et al. (2019) did not require the observed persons carrying sensors, but performed the detection of their activity through fixed infrastructures, such as PIR sensors (Erickson et al. 2011), cameras (Lao et al. 2009), Kinect sensor (Chikhaoui et al. 2017), WiFi signal sensing (Liu et al. 2019; Xin et al. 2018; Wu et al. 2015; Wilson et al. 2011), visible or reflective light sensing (Yang et al. 2017; Nguyen et al. 2018), ambient sensors (Alemdar et al. 2017). For example, camera-based approaches and optical equipment-based approaches collect human motion images to extract and identify human activity, whereas infrared-based approaches perform imaging of the human body in bad lighting conditions. In Wang et al. (2014), the authors proposed a method for deducing human activity from the indoor location of the observed person. Finally, hybrid approaches based on the combined used of sensors and cameras (Mitchell et al. 2014) have been proposed to combine the advantages of positive recognition in accuracy and passive recognition in convenience.

Although these approaches make the ADLs assessment more objective and mitigate obtrusiveness for the assessed persons, they are based on rich information about persons's lives and biometrics (i.e. human-centric), which raise some severe privacy concerns. For example, the camera-based approaches has the potential to allow older adults to remain in their homes longer than may otherwise be possible. Some images of home environment and private behavior unrelated to research are also captured by these human-centric approaches (Caine et al. 2016).

Table 1 Activity Recognition based on Human-centric Sensing.

Positive	Passive
RFID sensor (Fujinami et al. 2011)	Passive infrared (PIR) sensors (Erickson et al. 2011)
Body sensor networks (Keally et al. 2011)	Cameras (Lao et al. 2009)/ Kinect sensor (Chikhaoui et al. 2017)
Smart watch (Lee et al. 2015)	Wireless sensing (Liu et al. 2019; Xin et al. 2018; Wu et al. 2015; Wilson et al. 2011)
Wrist-worn sensors (Galluzzi et al. 2015)	Visible/reflective light sensing (Yang et al. 2017; Nguyen et al. 2018)
Accelerometer (Chikhaoui et al. 2018)	Ambient sensors (Alemdar et al. 2017)
	Indoor localization-based (Wang et al. 2014)
Hybrid (Sensors and Cameras) (Mitchell et al. 2014)	

Table 2 Activity Recognition based on Thing-centric Sensing.

Sensor	Type of Measurement	Tracked ADL Indicator
Contact switches (Dickerson et al. 2011)	Opening/Closing	Object usage
Binary sensors (Morales et al. 2013)	Opening/Closing	Object usage
Vibration sensors (Chen et al. 2019)	Used/Not Used	Object usage
RFID sensors (Yang et al. 2011)	Object information	Object usage
Wattmeter (Franco et al. 2008)	Consumption information	Electrical object usage

(2) **Thing-centric approach:** Thing-centric activity recognition refers to using environmental information on things to infer human activity. Some representative examples of thing-centric sensing are listed in Table 2. Some researchers proposed to take contact switches (Dickerson et al. 2011), binary sensors (Morales et al. 2013), vibration sensors (Chen et al. 2019), RFID (Yang et al. 2011) to detect object usage, and infer human activities using such kind of environmental information on things. In Dickerson et al. (2011), the authors placed some contact switches on the microwave, the oven, the spice cabinet, the refrigerator, and the freezer to detect activities of daily living. In Morales et al. (2013), binary sensors were adopted to measure the opening or closing of doors and cupboards and the use of electric appliances; additionally, motion sensors were used to recognize ADLs of elderly persons living on their own. In Yang et al. (2011), the authors leveraged information of the objects which humans touch while taking actions for daily living to conduct activity recognition. In Franco et al. (2008), the authors recorded the electricity consumed by room lights and various appliances and then translated it into the probability of a particular ADL.

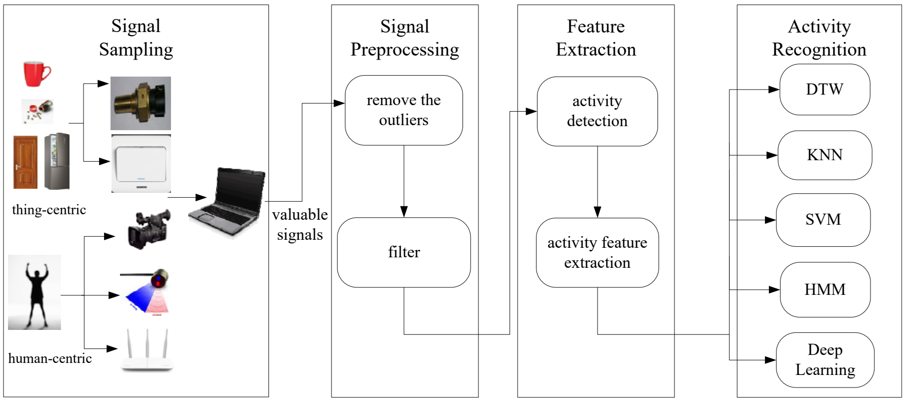


Fig. 1 The framework of behavior sensing.

Compared with the human-centric approach, the thing-centric approaches eliminate the privacy concerns of the observed persons, but still have a number of limitations. For example, one of the most prevalent is the multiple-persons interference problem, which consists in recognizing the acting person who generates the vibration of action transceiver, when there are two or more persons simultaneously inside an area.

2.2 Framework of behavior sensing

As pointed out in the introduction section, no matter which kind of approaches are implemented for behavior sensing, they are based on processing valuable signals that change with activity. As shown in Fig. 1, they both include the following four steps. Firstly, *signal sampling* collects valuable raw signals from various devices (e.g. cameras, light sensors, contact switches, etc.). Secondly, the raw signals are *preprocessed* to reduce the noise by means of two phases: (i) removing the outliers and (ii) applying filters (e.g. digital filter, principal component analysis, etc.). Then, *feature extraction* is performed, also in two different phases; firstly, the activities are detected by segmenting the denoised signal, then effective features are extracted from each segment. The last operation is *activity recognition* that leverages Machine Learning or Deep Learning-based classifiers fed with the features extracted in the previous step.

According to the working process of behavior sensing, as summarized in Fig. 1, in the following sections, we review the approaches employed to accomplish each step and the related challenges faced for signal sampling, signal preprocessing, feature extraction, and activity recognition.

3 Signal Sampling

Signal sampling is the first step in behavior sensing. The collection of environmental signals is crucial to successfully recognize human activities. For either human-centric or thing-centric behavior sensing, pioneering works on this area (Fujinami et al. 2011; Keally et al. 2011; Lee et al. 2015; Galluzzi et al. 2015; Dickerson et al. 2011; Morales et al. 2013; Chen et al. 2019; Yang et al. 2011; Franco et al. 2008) were proposed collecting different types of signals via wearable or embedded sensors. Conversely, later works (Erickson et al. 2011; Lao et al. 2009; Wilson et al. 2011; Liu et al. 2019; Wu et al. 2015; Yang et al. 2017; Nguyen et al. 2018) explored fixed sensing infrastructures to acquire signals.

The challenge in signal sampling lies in finding solutions that can collect signal with a minimal requirement on the sensing infrastructure, or even removing the need to carry a dedicated device. Another challenge encountered by the researchers is to keep the sensing environment stable. Indeed, a small variation of the sensing environment, such as fluctuation of lighting or change of the line of sight, could have a severe impact on signal sampling. In the following, we report an analysis of the existing methods for signal sampling. The latter three types (i.e. Camera Based, Infrared Based, and Wireless Signal Based) of methods are representatives of the sensing infrastructure-based signal sampling for behavior sensing.

Sensor Based: Collecting valuable activity-related signals via wearable sensors is crucial to conduct the subsequent steps of activity recognition. In Casale et al. (2011), the authors presented a custommade wearable system for human action recognition, which was based on processing signals of accelerometer and inertial sensors. In Hao et al. (2013), a practical system, called *iSleep*, was developed to monitor an individual's sleep quality using an off-the-shelf smartphone. It used the built-in microphone of the smartphone to detect the events that were closely related to sleep quality. In Yatani et al. (2012), acoustic sensors were used to record the sound produced by the user's throat area. Thus, researchers could distinguish user activities such as eating, drinking, talking, laughing, and coughing. In Dickerson et al. (2011), a cohesive set of integrated wireless sensors were attached to the things and connected to a mobile device to create a real-time depression monitoring system for the home. The data collected were multi-modal, which represented several different behaviors including sleep quality, weight, and activities of daily living. The data aggregated across multiple behavioral domains were used to help in caregivers' diagnostic assessment and therapeutic treatment planning, as well for patients in the management and tracking of their symptoms.

Camera Based: Cameras are popularly used in collecting images of human activities, which are then processed to extract features and identify human activities through computer vision methods (Herath et al. 2017; Kerola et al. 2014; Yang et al. 2012; Chikhaoui et al. 2017). In Herath et al. (2017), the authors investigated several aspects of the existing solutions for action recognition using cameras and focus on solutions that benefit from deep architectures. In Kerola et al. (2014), the authors presented spectral graph skeletons (SGS), a novel graph-based method for action recognition from depth cameras, and leveraged the SGWT framework to create an overcomplete representation of an action signal lying on a 3D skeleton graph. In Yang et al. (2012), a new data set of 3.6 Million accurate 3D human poses, named Human3.6M, was introduced, which recorded the performance of persons under 4 different viewpoints by cameras. Authors leveraged Human3.6M for training realistic human sensing systems and for evaluating the next generation of human pose estimation models and algorithms.

The main advantage of camera-based approaches is the relatively high recognition accuracy and for that reason, it is the most widely used. However, it is limited by the range of lighting and the line of sight of the camera.

Infrared Based: The infrared (IR) is electromagnetic radiation (EMR) with longer wavelengths than those of visible light and it is therefore generally invisible to the human eyes. Infrared-based sampling utilizes the physical properties of infrared rays and the principle of infrared emission. When a certain part of the human body or thing is in the infrared region, IR receiver diode will receive the infrared rays emitted by IR led due to the human body block. Then the received signal will be transmitted to the controller through the integrated circuit, and the controller will perform the corresponding instructions. In Yun et al. (2014), the authors took advantage of the infrared ray to collect the signals of human behaviors, placing pyroelectric IR sensors in a hallway for monitoring persons.

As opposed to camera-based sensing, this method is not affected by the light, but it has high requirements on the sensing platform and devices.

Wireless Signal Based: WiFi-based behavior sensing was firstly proposed by researchers, using special hardware (Pu et al. 2013; Kellogg et al. 2014; Lyonnet et al. 2010; Adib et al. 2014), such as Software Defined Radio (SDR), to collect signals. Conversely, in later works (Aly et al. 2013; Kosba et al. 2012; Sabek et al. 2012), the researchers tried collecting Received Signal Strength (RSS) to realize human activity recognition. More recently, the researchers have proposed methods to recognize activities by acquiring channel state information (CSI) with phase and amplitude information extracted from wireless signals. In Wang et al. (2016), WiFall detected a falling activity in an indoor environment employing CSI. In Zhou et al. (2013), the authors leveraged CSI to estimate occupancy in multipath-rich indoor scenarios. Several works leveraged CSI for recognizing various fine-grained behaviors, such as daily activity awareness (Xi et al. 2014; Xin et al. 2018), gesture recognition (Li et al. 2016a; Tan et al. 2016), person identification (Zeng et al. 2016), and breathing detection (Liu et al. 2015).

Table 3 compares the four signal sampling approaches described above in terms of the required equipment, sensing distance, cost and accuracy. The main drawback of sensor based approaches is that they are invasive, which means that sensors need to be attached to or carried by the observed persons. Additionally, some special-purpose sensors are too expensive to be widely used. The main advantage of the camera-based approaches is the relatively high recognition accuracy. However, the line of sight of the camera is limited by the range of lighting. As opposed to the camera, the infrared-based approaches is not affected by the light, but it has high requirements on the sensing platform and devices. The wireless signal based approaches are implemented with either commercial WiFi module, SDR (Software Defined Radio), or other special hardward. It can achieve large coverage at low cost, but with moderate accuracy, especially in a space with many persons coexisting.

According to the different preprocessing methods, the signals collected through the devices described above and approaches will be saved as a file or directly as streaming data. Taking the "WiFi-based signal" as an example, in the following sections, we will introduce the challenges and approaches of the remaining steps (i.e. signal preprocessing, feature extraction, and activity recognition) for behavior sensing.

Table 3 Comparison among works on signal sampling.

Signal Sampling	Equipment	Distance	Cost	Accuracy
Sensor Based	Dedicated sensors	close range	high	high
Camera Based	Camera	visible distance	high	high
Infrared Based	sensing platform and devices	visible distance/no-visible distance	high	moderate
Wireless Signal Based	WiFi/SDR/ special hardware	no-visible distance	low	moderate

4 Signal Preprocessing

The signal collected by wearable sensors or fixed sensing infrastructure in different forms contains not only the thing-centric or human-centric information related to the interested behavior but also a large amount of environmental noise. Thus, to improve the accuracy of sensing, the signal must be preprocessed.

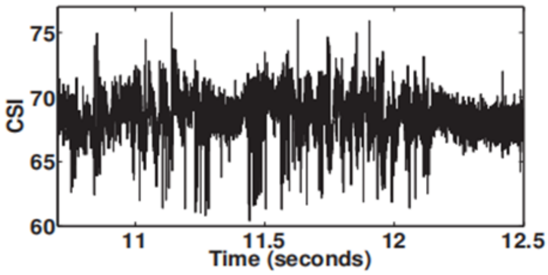
There are two key challenges in signal preprocessing. The first technical challenge is removing the outliers. The reason for the outliers is that a sudden change of the state inside the equipment will result in a mutation in the collected signal. Thus, if we do not remove the outliers, they may affect the results of the subsequent activity recognition.

The second technical challenge is that signals are often too noisy to be directly used for human activity recognition. Noise is generated randomly in the surrounding environment and several high-frequency sensing components are very likely to be affected. Hence, effectively removing the noise coexisting with the subtle valuable signal of monitored objects is challenging.

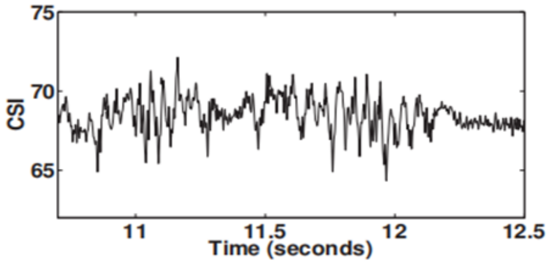
Next, we will have an overview of the methods for both removing outliers and denoising.

Removing Outliers: An outlier is an individual value, which deviates significantly from other observations of the samples it belongs to. Outlier detection can be statistic-based, density-based, and migration-based. The principle of outlier detection is to give a confidence probability and confidence limit, and any error that exceeds this limit is treated as an outlier. A statistic-based method of removing outliers is to use a Hampel filter based on the median absolute deviation, which calculates the mean μ and standard deviation σ in the sliding window, and sets a range $([\mu-\gamma\times\sigma, \quad \mu+\gamma\times\sigma])$ according to μ and σ. When the value in the window is not within this range, it is considered an outlier and is replaced by the mean of the window. γ is an adjustable parameter. It can perform a small correction on the signal to obtain a set of data with fewer outliers.

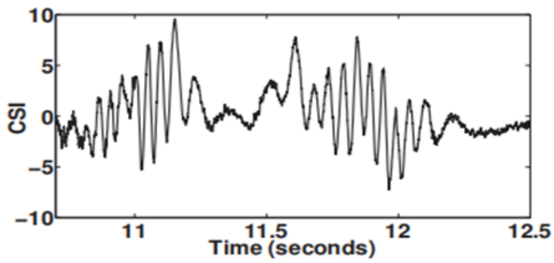
Denoising Filter: The frequency of human activities is common in the low-frequency range. At present, there are many filtering methods for denoising, of which digital filter and principal component analysis (PCA) are two main methods. Digital filter is to select a low-pass filter or a band-pass filter to filter the noise of the signal according to the frequency band. PCA reduces the dimensionality of obtained signals, choosing the principal components that represent the most common variations among all time-series. Meanwhile, PCA helps in removing uncorrelated noisy components from the signals by taking advantage of the correlated variations in CSI time-series. Fig. 2(a) depicts a row signal (in terms of CSI) as acquired by the sensing device; the signals obtained by applying the digital low-pass (e.g. Butterworth) filter and the PCA are reported in Fig. 2(b)



(a)



(b)



(c)

Fig. 2 Original signal, filtered signal and signal after PCA. (a) Original signal. (b) Butterworth low-pass filter. (c) PCA-based denoising (Wang et al. 2015).

and 2(c), respectively. Comparing the original signals with the filtered signals, we can notice that the noise is effectively removed.

Low-Rank Matrix Decomposition Based Denoising: The Low-Rank Matrix Decomposition is mainly used for image processing, such as denoising and deblurring. It can also be applied to denoise in the non-image domain. The principle of the matrix low-rank decomposition algorithm is to treat the degraded image as a set of low-dimensional data plus noise.

Suppose D is a blurred image, and according to the low-rank decomposition, it can be formulated as the following combinatorial optimization problem:

$$\begin{array}{l} \min rank(X) + \|E\|_0 \\ \text{s.t. } D = X + E \end{array}, \tag{1}$$

where E and X are unknown, X denotes a clear image, and E denotes a noise with sparsity; $\|\cdot\|_0$ is the zero (L0) norm. Since$\|\cdot\|_0$ is non-convex, the problem formulated by the equation (1) is NP-hard. Robust principal component analysis (RPCA) is used to solve E and X. First, a weighted factor $\lambda(>0)$ is introduced, then to turn the problem (1) into a convex optimization problem, the L0-norm is relaxed to L1-norm problem, as shown in Equation(2).

$$\begin{array}{c} \min \|X\|*+\lambda\|E\|_1 \\ \text{s.t. } D = X + E \end{array}, \tag{2}$$

where $\|X\|*$ is the nuclear norm of matrix X, which is the sum of its singular values. $\|\cdot\|_1$ is the L1-norm, which represents the sum of the absolute values of each element of the matrix.

There are many ways to solve the problem (2), such as the iterative reweighted least squares algorithm (Fornasier et al. 2011), the augmented Lagrange multiplier method (Lin et al. 2010), and the singular value threshold (Cai et al. 2010) among the others.

In Wu et al. (2018), *TW-See* partitioned the raw signals into two components, being the indoor physical environment noise signal and the CSI values that changed with activities. In details, Xuangou Wu *et al.* proposed an Opposite robust PCA (Or-PCA) approach to remove pulses and burst noise and obtained the correlation between human activity and its resulting changes in CSI values. Fig. 3(a), 3(b), and 3(c) show the raw signal, the environmental noise obtained by the low-rank decomposition algorithm, and the changed CSI values, respectively.

5 Feature Extraction

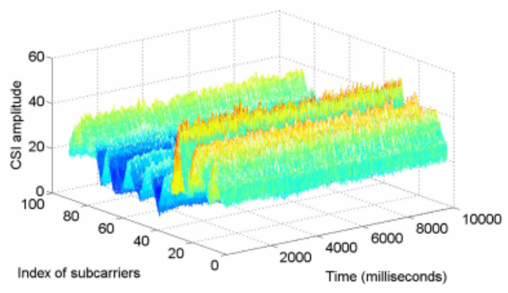
A large amount of environmental noise has already been removed from the signal obtained after preprocessing. Such signal is then processed to extract features for recognizing activities. For feature extraction, there are two key technical challenges. The first technical challenge is to segment the time series to identify the start time and end time of each activity (i.e. activity detection). It is difficult to determine the time period when activities occur because different persons have different habits, which can lead to different durations of activity.

The second technical challenge is to extract distinguishing activity features for generating the classification models. The main effect of human activity on the received signal is that the waveform presents a rising edge, a falling edge or a pause. It is very important to find the distinguishing features from the various changing patterns, which will be used to classify the activities.

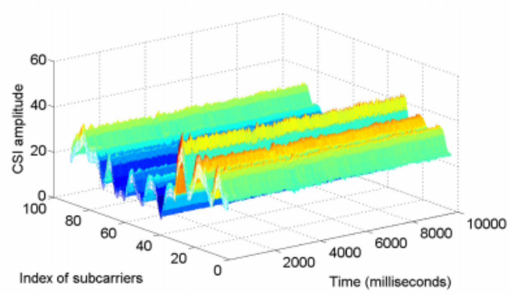
This section will describe several methods of segmentation (i.e. activity detection), followed by a brief introduction on the methods for feature extraction.

5.1 Segmentation

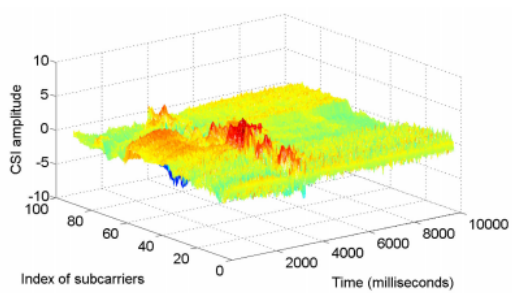
In the following, we discuss three segmentation methods, namely the segmentation based on the Local Outlier Factor, the (human) activity indicators, and that performed by adopting a sliding window.



(a)



(b)



(c)

Fig. 3 Raw CSI streams and denoising CSI streams based on the low-rank decomposition algorithm. (a) 90 Raw CSI streams. (b) Background environment CSI values. (c) The changed CSI values (Wu et al. 2018).

LOF Based: The Local Outlier Factor (LOF) is an anomaly detection algorithm based on the concept of local density, where the locality is defined by the distances of k nearest neighbors. By comparing the local density of an object (e.g. a point p) to the local densities of its neighbors, one can identify the regions with similar density and the points which have a substantially lower density than their neighbors. These are considered to be outliers. If the distance between the points is farther, the lower the density of the point p, the more likely it is to be an outlier. On the contrary, if the distance is closer, the density will be higher, and it is more likely to be not an outlier.

For example, *WiFall* (Wang et al. 2016) used the LOF-based anomaly detection algorithm to separate the corresponding anomaly patterns and recognize human activities. LOF indicates the probability that a point would be an outlier or not. At first, WiFall learned a model of the stable situation (i.e. the person is resting). Then, it computed the LOF of the data by comparison with the stable model to detect anomaly patterns. Researchers obtained nine LOF lists as a result of the analysis conducted on a CSI dataset. A LOF value of around 1 indicated that the point was located in a region of uniform density and it was not considered as an outlier (i.e. no activity occurred). On the contrary, if an outlier occurred in a stream, this stream was considered to be abnormal. For all nine streams, if the majority of them showed anomalies, WiFall considered the activity corresponding to this dataset as an anomaly, which meant that the human activity occurs.

Activity Indicator Based: Activity indicators are those which can represent the occurring activities. Some researchers have proposed to realize activity detection based on activity indicators, such as the amplitude, the phase, and other related information of the signal. The activity indicator is usually used along with a threshold. In this way, the activities are detected by judging whether the activity indicator is within the threshold range.

For instance, *WiStep* (Xu et al. 2018) and *CARM* (Wang et al. 2015) leveraged the correlation of signal subcarriers to calculate the activity indicator. In particular, they calculated the correlation matrix and performed eigendecomposition of the correlation matrix to calculate the eigenvectors and principal components. Authors found that, in the case of no activity, the variance of the second principal component h_2 was small, and the second eigenvector q_2 varied randomly over neighboring subcarriers. On the contrary, in the presence of an action the CSI series were correlated, q_2 varied smoothly over neighboring subcarriers, and the principal component h_2 had higher variance. Therefore, the authors defined the activity indicator as the ratio of the variance of the second principal component to the differential of the second eigenvector q_2, as shown in Equation (5):

$$E^2\left\{h_2\right\} = \frac{1}{L}\sum_{l=1}^{S}\left(h_2(l) - \overline{h_2}\right)^2 \tag{3}$$

$$\delta\left\{q_2\right\} = \frac{1}{S-1}\sum_{l=2}^{S}\left|q_2(l) - q_2(l-1)\right| \tag{4}$$

$$R = \frac{E^2\left\{h_2\right\}}{\delta\left\{q_2\right\}} \tag{5}$$

where L denotes the length of the CSI time-sequence, S denotes the number of subcarriers, h_2 is the mean of the second principal component, R denotes the activity indicator, and $|q_2(l) - q_2(l-1)|$ is the difference in coefficients for neighboring subcarriers. R can be used to determine whether the activities occurred. Comparison of the three activity indicators is shown in Fig. 4, where we can see that the change of R (orange squares) is more obvious than the other two indicators (note that the Amplitude is reported in logarithmic scale).

Sliding Window Based: Time sequences of activities differ depending upon the direction of activities or their duration (i.e. activity characteristics), so researchers need to distinguish effective fragments that contain action information

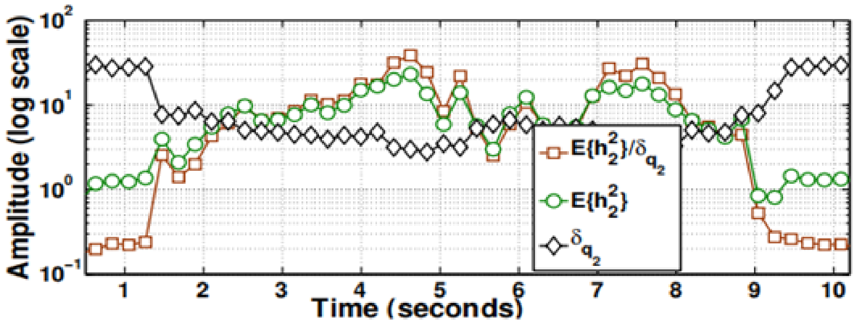


Fig. 4 Comparison of the three activities indicators (Wang et al. 2015).

based on action characteristics. For some fine-grained activities, the change of the signal is not as obvious as for the coarse-grained activities, so it is difficult to determine the start- and end-points of the activity. Therefore, some researchers have adopted a sliding window to segment the signal and have focused on the local By sliding the window, it is determined whether there is any activity occurs in the window.

For example, Ali et al. (2015) detected the start- and end-points of the activity (i.e. keystroke recognition) through a sliding window. Firstly, the algorithm calculated the mean absolute deviation (MAD) for each CSI time series and for each window of size W (Equation (6)). Secondly, the algorithm added mean absolute deviations (Δm_j) in each waveform to calculate a combined measure ΔM_j of MAD in all p waveforms (Equation (7)). Thirdly, the algorithm compared ΔM_j to a heuristically set threshold Thresh. Let $\delta_j = \Delta M_j -$ *Thresh*, then $\delta_j > 0$ indicates that the current window j contains significant variations in CSI amplitudes. Fourthly, the algorithm compared δ_j to its value in the last window δ_{j-1} and incremented the value of i_u by 1 when $\delta_j - \delta_{j-1} > 0$ and du by 1 when $\delta_j - \delta_{j-1} < 0$. Repeating the above steps until the values of i_u and d_u exceed empirically predefined thresholds I_u and D_u, respectively, the algorithm detected the start of the keystroke.

$$\Delta m_j[k] = \frac{\sum_{i=j}^{j+W} \left| Z_{t,r}^{(k)}(i) - \bar{Z}_{t,r}^{\{k\}}(j : j + W) \right|}{W} \tag{6}$$

$$\Delta M_j = \sum_{k=2}^{p} \Delta m_j[k] \tag{7}$$

where $\bar{Z}_{t,r}^{\{k\}}(j : j + W)$ represents the vector of means of the k^{th} projected CSI stream in the j^{th} window.

5.2 Activity Feature Extraction

After completing the activity detection, the next step is activity recognition. However, if the original waveform of the signal was directly classified, it would bring large computation cost and low accuracy. Hence, the exclusive features associated with each activity are generally extracted based on the waveform to improve activity recognition. Table 4 reports the activity features used in the literature grouping them into three categories: time-domain, frequency-domain, and time-frequency domain features. We discuss each category in the next subsections.

Time-domain Features: The original time-domain signal can directly be used to extract the feature vector through probability statistical methods. If this is the case, we can also refer to the extracted time-domain features as statistical features. Currently, commonly used time-domain features are Mean, Variance, Minimum, Maximum, Inter-Quartile Range (IQR), Root Mean Square (RMS), slope, and so on.

For example, in Sigg et al. (2013), the authors extracted time-domain features (i.e. Mean and Variance of the signal amplitude) for activity recognition. As shown in Fig. 5, the activities are easier to be distinguished based on these two features (center and bottom figures for mean and variance, respectively) as compared with the raw signal amplitude (top figure). *TW-See* (Wu et al. 2018) and *WiFall* (Wang et al. 2016) also extracted time-domain features for activity recognition (cf. Table 4).

Frequency-domain Features: Frequency-domain features, like time-domain features, are also statistical features, which are extracted from the signal frequency. When we perform FFT on the original waveform, valuable information can be found and extracted as frequency-domain features. Currently, commonly used frequency-domain features are Energy, Spectrum Entropy, Power Spectral Density, and so on. For instance, Xu et al. (2018) used frequency-domain features for feeding the walking step counting system proposed in their work. In Zeng et al. (2016), the authors conducted experiments to verify that time-domain features and frequency-domain features of different activities are observed in the WiFi CSI signals. Fig. 6 shows the different (normalized) coefficients of the FFT profile for the four activities (without the DC component) detected in the CSI data. The four activities taken into account are: (a) there is no person in the room, (b) a person is sitting and performing routine activities (e.g. typing, moving objects on a desk, etc.), (c) a person is standing while performing routine activities (e.g. using her phone, writing on whiteboard, etc.), and (d) a person is walking. It can be seen that the frequency-domain features of each activity are different, thus representing valuable information for activity recognition.

Time-frequency Domain Features: The time-frequency domain features preserve the characteristics in both the time and frequency domain. Discrete wavelet transform (DWT) is a typical method of extracting time-frequency domain features. It provides every detail of a signal for analysis in time and frequency domain at multi-scale. Directly extracted features will lose a lot of detail information because of the difference in activity speed. DWT can obtain the wavelet coefficients of each frequency band, which solves this problem well, and also acts as a filter since it can remove high-frequency noise.

CARM (Wang et al. 2015) applied DWT to decompose the PCA components into 12 levels that correspond to the speed of different parts of the body and span the frequency range from 0.15 Hz to 300 Hz. *Wi-Finger* (Li et al. 2016a) extracted finger waveform as features by DWT. Fig. 7(a) to Fig. 7(c) show the finger waveform by DWT for finger gesture No.1, No.4 and No.7, respectively. The component in every colored bar of each figure is the averaged result of every 6 subcarriers.

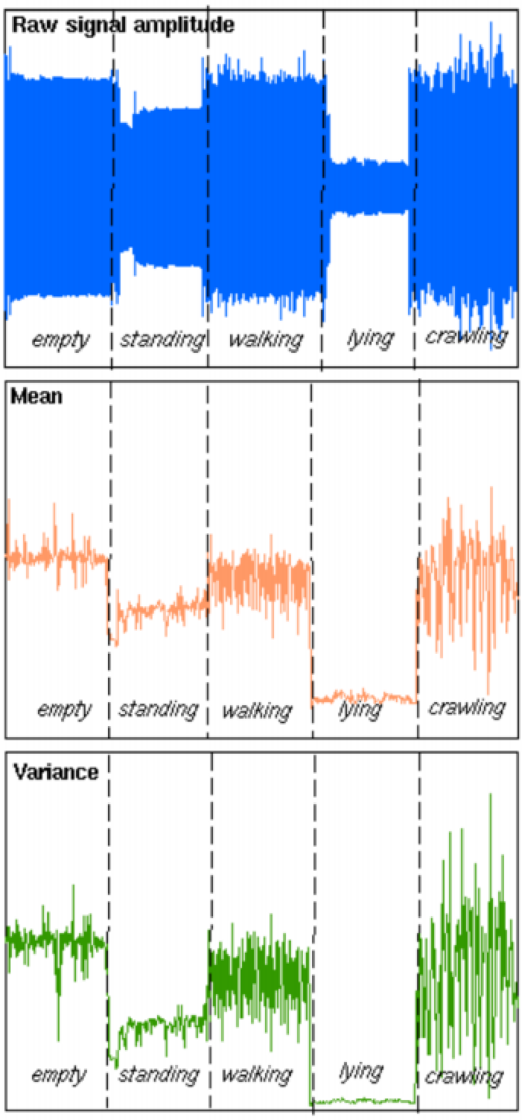


Fig. 5 Time-domain features (Sigg et al. 2013).

The three categories of features described above and their integration in the recent works on activity classification are listed in Table 4. Considered works are ordered starting from the most recent. From Table 4, we can see that the researchers can choose different features in the time domain according to the different perceptual actions, whereas FFT and DWT are the unique features used in the frequency- and time-frequency domain, respectively. Because the time-domain features can be more easily extracted than the frequency-domain ones, and are usually sufficiently

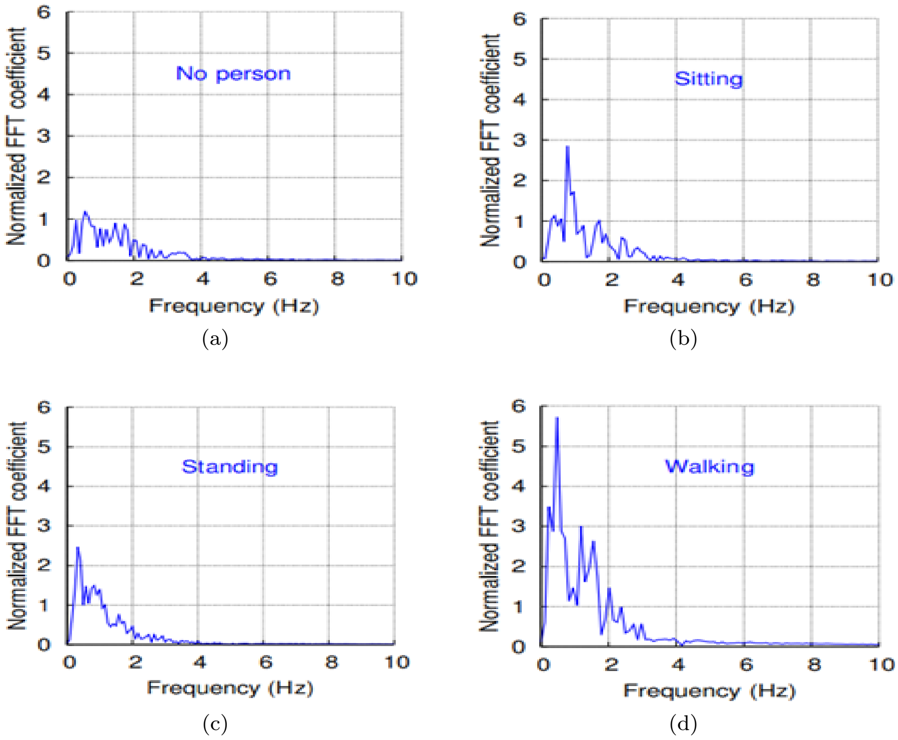


Fig. 6 Comparison of FFT coefficients for different activities (Xu et al. 2018).

diverse for classifying activities, they are most widely employed by the considered works.

Table 4 Commonly used features in activity recognition.

Year	Paper	Time Domain							Frequency Domain	Time-Frequency Domain
		STD	MAD	IQR	Max/Min	Entropy	Mean	Var	FFT	DWT
2018	TW-See(Wu et al. 2018)	✓	✓	✓	✓					
2018	WiStep(Xu et al. 2018)								✓	
2016	WiFall(Wang et al. 2016)	✓	✓	✓						
2016	Wi-Finger(Li et al. 2016a)									✓
2016	WiWho(Zeng et al. 2016)				✓				✓	
2015	CARM(Wang et al. 2015)									✓
2013	Sigg et al. (2013)						✓	✓		

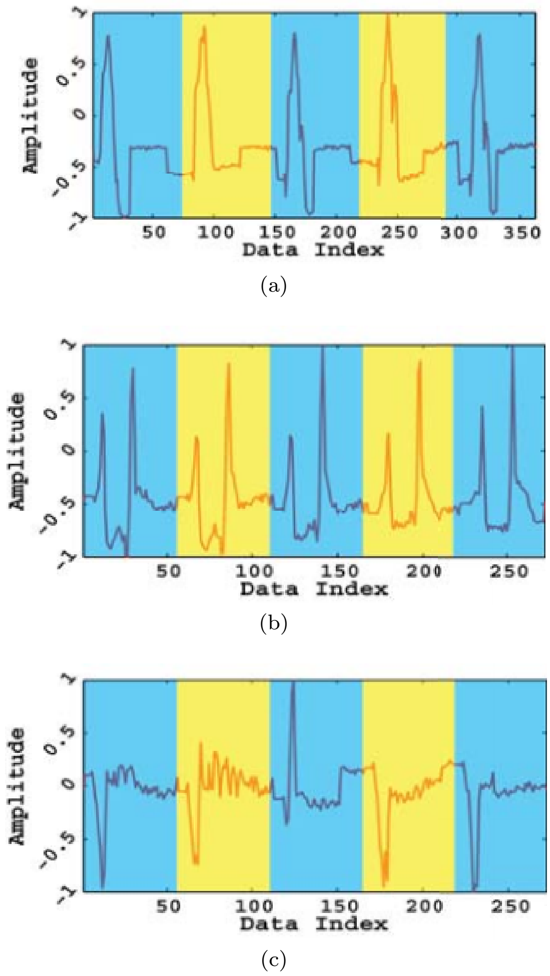


Fig. 7 DWT of time-series for finger gestures. (a) Finger features for gesture No.1. (b) Finger features for gesture No.4. (c) Finger features for gesture No.7 (Li et al. 2016a).

6 Activity Recognition

The activity recognition step employs various classification algorithms fed with the features extracted in the previous step and leveraged to recognized different human activities. The challenge in activity recognition is building the classification model so that it is robust for different persons in different environments. Indeed, for the same activity to a certain degree-different persons perform it differently and even the same person could perform it differently at different times.

The direct way to determine the type of recognized activity is matching the extracted activity features with the known activity templates. Dynamic Time Warping (DTW) is a typical algorithm for carrying out this method.

Another alternative is to utilize supervised (i.e. the action data are labeled with their actual types) Machine Learning-based classifiers to discriminate between the different types of activity (viz classes). In this case, the recognition model is built (viz. trained) with the activity features extracted from the data set collected in the application environment that represent the training set. Then some new activity data (i.e. activity features extracted from samples not included in the training set) are taken as the test set and used as the input of the trained classifier. Finally, the action tags are output after completing activity recognition. The performance of the classifiers are then computed comparing the tags returned in the test phase with the actual tags (viz. labels) of the test samples. In next subsections, we discuss two examples of Machine Learning-based classifiers commonly used in human activity recognition: K-Nearest-Neighbor (KNN) and Support Vector Machine (SVM).

Other recognition methods, such as Hidden Markov Model (HMM), which are based on probability statistical models have also been proposed. They treat the sequence of activities as a sequence of states, and law of state conversion is represented by the state transition function.

In addition, the researchers have begun to use Deep Learning techniques to improve the accuracy of activity recognition.

In the following, we briefly describe the various classification methods together with the state-of-the-art works in which they are employed.

DTW: Dynamic Time Warping (DTW) is a well-known algorithm to find an optimal alignment between two given (time-dependent) sequences (Daubechies et al. 1992). Compared with the Euclidean distance (a common basic choice), DTW is more suitable for calculating the distance between two waveforms. *E-eyes* (Wang et al. 2014) calculated the MD-DTW distance between the test CSI measurements and all the known activities. If the MD-DTW distance was less than a threshold, then E-eyes regarded the corresponding CSI measurements-labeled with the minimum distance as the activity identified as the test measurements. Similarly, in *Wi-Finger* (Li et al. 2016a), the authors exploited the MD-DTW to match the extracted finger waveform with the known (viz. baseline) template. The principle of the DTW algorithm is dynamic programming. Indeed, DTW has high robustness, but it has large computation costs and a strong dependence on the baseline template. This means that the accuracy of the template must be ensured during the identification process; otherwise, the results would be seriously affected.

KNN: The principle of the KNN classifier is that if the majority of the K most similar samples (viz. neighbors) in the feature space belong to a certain category, the analyzing sample also belongs to this category. If the distribution of the pretrained sample number is non-uniform between the classes, this will also affect the recognition result, therefore undersampling or oversampling techniques are leveraged to mitigate the impact of imbalanced data (Liu et al. 2012). Yan Wang's *E-eyes* (Wang et al. 2014) and Li's *Wi-Finger* (Li et al. 2016a) both used KNN as the classifier for recognizing in-place and walking activities, and digits finger-grained gestures, respectively.

SVM: The SVM classifier was first introduced in 1964 and it was quickly developed in the 1990s when researchers derived a series of extended algorithms. Support Vector Machine (SVM) introduces kernel functions (e.g. Polynomial, Gaussian, Sigmoid kernel, etc.) to map linear inseparable input data into a high dimensional linear separable feature space, with the aim of finding a set of opti-

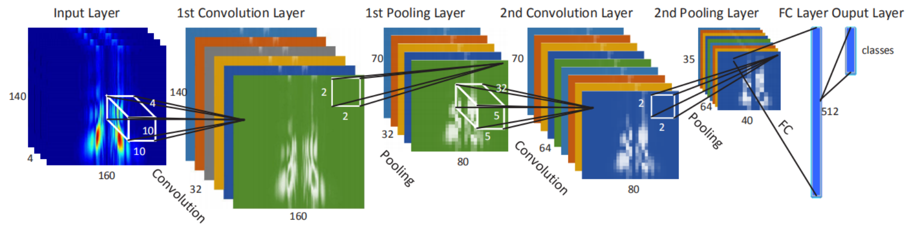


Fig. 8 Convolution Neural Network architecture proposed in *UltraGesture* (Ling et al. 2018).

mal hyperplanes in the high dimensional feature space. SVM has the great disadvantage of being highly computationally demanding since computing and storage requirements increase rapidly with the number of training vectors (its core is a quadratic programming problem). *WiFall* (Wang et al. 2016) employed SVM for activity recognition. It extracted seven features (i.e. Normalized STD, Offset of Signal Strength, Period of Motion, MAD, IQR, Signal Entropy, and Velocity of Signal Change) as input to achieve fall detection for a single person with high accuracy (i.e. up to 96% precision and 18% false alarm).

HMM: Hidden Markov Model (HMM) is a type of Markov chain, which can be described by two state sets, namely the hidden state S (which cannot be obtained by direct observation) and the observable state O (which can be obtained by direct observation), and three probability matrices, namely the original state probability matrix π (i.e. the probability matrix with hidden state at the original time t=1), the hidden state transition probability matrix A (i.e. the transition probability between hidden states in HMM model), and the observation state transition probability matrix B (i.e. the transition probability between observable states in HMM model). The hidden state is not directly visible, whereas the observable state, which is dependent on the hidden state, is visible. Specifically, each observable state is generated by a hidden state with a corresponding probability density distribution. Generally, a HMM can be represented by a triplet (A, B, π), which has a good ability to solve problems with dynamic time series. *CARM* (Wang et al. 2015) adopted the HMM and constructed a model for each activity using the training samples of that activity. Moreover, it constructed an activity model for the situation without any activity. To estimate average vector and covariance matrix of each activity state and the transition probabilities of the HMM, the authors leveraged the Baum-Welch algorithm, and finally identified the type of activity with up to 96% accuracy.

Deep Learning: Deep Learning allows to train classifiers directly from input data (or coarse-grained features) by automatically learning structured and complex feature representations, thus limiting the need of designing handcrafted domain-expert driven features (e.g. statistical features). The most common structure of Deep Learning classifiers is the artificial neural network, characterized by a complex hierarchical structure comprising multiple hidden (viz. intermediate) layers and therefore commonly referred to as "deep". Deep Learning architectures include deep neural networks, deep belief networks, recurrent neural networks, convolutional neural networks, and so on, which have produced results comparable to and in some cases superior to human experts (Aceto et al. 2019b). Ling et al. (2018) presented *UltraGesture*, which was a Channel Impulse Response

(CIR)-based ultrasonic finger motion perception and recognition system. To take into account important feature properties and guarantee generality, authors designed a Convolutional Neural Network (CNN) classifier and took a period of continues CIR measurements as its input. Fig. 8 shows the CNN architecture of UltraGesture. Since the authors considered the CIR measurement as an image, they designed UltraGesture employing two 2D-convolutional layers, each followed by a Max-pooling layer used to reduce complexity and mitigate overfitting, and a final Fully Connected layer before the soft-max output. Experimental results showed that the proposed CNN architecture was able to classify different gestures with high robustness, achieving an average accuracy up to 97% for 12 gestures including finger click and rotation. *TW-See* extracted 8 features from CSI and used a three-layer Back Propagation (BP) neural network to recognize different human activities under the scenarios where the WiFi signals pass through the wall. Authors pro-posed a three-layer BP neural network, including an input layer with 24 neurons, an output layer with 7 neurons, and a hidden layer with 14 neurons. Using the sigmoid activation function and cross-entropy cost function, they trained their model for 100 epochs, achieving an average 94% accuracy. The works presented (Li et al. 2016b; O'Shea et al. 2017) also devised Deep Learning classifiers, leveraging different compositions of convolutional and fully connected layers for behavior sensing behavior sensing.

Table 5 summarizes the above-described classification methods employed in the most recent works on activity recognition ordered by year. KNN has relatively lower computing cost compared with SVM and HMM, but may have higher accuracy than SVM and HMM. The computing cost of deep learning is higher than SVM and HMM, but its performance fluctuates with different recognition objects.

From Table 5, we can also find that DTW, KNN, SVM, and HMM are classification methods which are most widely adopted for activity recognition. DTW regards each feature vector as a time series and can be used to measure the distance (similarity) between two time series, even when they have different lengths or displacement on the time axis. KNN classifier is relatively low computational cost and is suitable to solve multi-classification problems. However, a parameter K (i.e. the number of nearest neighbors to be taken into account) needs to be manually adjusted so that the optimal parameters cannot be found adaptively, and different K values will result in different recognition accuracy. SVM can solve nonlinear classification problems. It has been successfully used in binary classification, multi-classification, and more generally hierarchical classification (Aceto et al. 2018). HMM has a good ability to solve problems with dynamic time series. However, HMM is used under the strong assumption that the observed feature vectors of the activity has Markov property (i.e. the conditional probability distribution of future states depends only on the present state), and has higher computing complexity. The latest works have all exploited Deep Learning for activity recognition. It has been revealed to be more accurate than the other methods through automatically extracting the most significant features. However, because it needs learning many parameters, which takes a lot of time to train, high-end hardware facilities and big data parallelism are exploited in accelerating the training phase of deep learning (Aceto et al. 2019a).

Table 5 Commonly used classification methods for activity recognition.

Year	Paper	Method	Cost	Accuracy
2018	UltraGesture(Ling et al. 2018)	Deep Learning	high	97%
2018	TW-See(Wu et al. 2018)	Deep Learning	high	94.46%
2017	O'Shea et al. (2017)	Deep Learning	high	90%
2016	Li et al. (2016b)	Deep Learning	high	85%
2016	WiFall(Wang et al. 2016)	SVM	moderate	90%
2016	Wi-Finger(Li et al. 2016a)	DTW&KNN	moderate	90%
2015	CARM(Wang et al. 2015)	HMM	moderate	96%
2015	Liu et al. (2015)	KNN	low	87%
2014	E-eyes(Wang et al. 2014)	DTW&KNN	moderate	96%

7 Summary, Outlook, and Challenges

Researchers have conducted a large number of human-centric or thing-centric behavior sensing studies to identify both coarse-grained and fine-grained activities, over recent years. However, there are still many challenges to be addressed in the future, concerning signal sampling, signal preprocessing, feature extraction, and activity recognition. In this paper, by considering the unified steps in both the human-centric approaches and the thing-centric approaches, we have made a review of the recently proposed approaches for behavior sensing.

Table 6 summarizes the different methods of signal sampling, signal preprocessing, feature extraction, and activity recognition integrated in the most recent works on human behavior sensing based on WiFi. We can see that most of these works employ low-pass or band-pass filter in the preprocessing step. Time-domain, frequency-domain and time-frequency domain features are often extracted in the feature extraction step. In the activity recognition step, deep learning is the state-of-the-art technique which can achieve good performance in recognition accuracy, but leads to high computational complexity.

Behavior sensing will provide important technical support for smart home and virtual reality in the future. For example, with China entering the aging society, it is an important task to realize the real-time monitoring of the state of the elderly. Falling down is a vital factor threatening the elderly. We can timely discover this emergent activity through technology of human behavior sensing, and take rescue response promptly to make home smarter. Virtual reality is commonly understood as a computer simulation technology that uses 3D graphics and devices to provide an interactive experience. So human-machine interface is one of the most important parts to realize virtual reality. Human behavior sensing is the fundamental to develop human-machine interface with good quality of experience (e.g. motion-sensing based controller) for advanced virtual reality.

Therefore, we want now emphasize the challenges and problems that need to be solved to achieve higher accuracy in human activity recognition, while keeping relatively low computational and storage complexity.

Extract Strong-correlation Features: At present, the commonly used time-domain, frequency-domain, and time-frequency domain features have been adopted to achieve satisfactory performance in behavior sensing. However, those features extracted from the raw measurements could not represent the real correlation between human activity and its resulting changes in feature values. Indeed, even for the same activity, if its amplitude, or its duration, or also the person-

Table 6 Comparison among works on human behavior sensing based on WiFi.

Year	Paper	Signal Preprocessing	Feature Extraction	Activity Recognition	Accuracy
2018	UltraGesture(Ling et al. 2018)	Down Conversion & Low-pass Filter	1st-order difference upon CIR (dCIR)	Deep Learning	Greater than 97% for 12 gestures
2018	WiStep(Xu et al. 2018)	Band-pass Filter	Frequency-domain features	Threshold algorithm	90.2% and 87.59% counting accuracies in two scenarios respectively
2018	TW-See(Wu et al. 2018)	Low-pass Filter & Opposite robust PCA	Time-domain features	Deep Learning	An average accuracy of 94.46%
2016	Li et al. (2016b)	Regularization	no	Deep Learning	76% recognition micro-accuracy and 85% average micro-accuracy
2016	WiFall(Wang et al. 2016)	Weighted Moving Average	Time-domain features	SVM	90% detection precision and 15% false alarm rate
2016	Wi-Finger(Li et al. 2016a)	Hampel & Low-pass Filter & Weighted Moving Average	Time-frequency domain features	MD-DTW	90.4%
2016	Wi-Who(Zeng et al. 2016)	Band-pass Filter	Time-domain & Frequency-domain features	Decision tree	2-6 people achieve 92%-80%
2015	CARM(Wang et al. 2015)	Low-pass filter & PCA	Time-frequency domain features	HMM	96%
2014	E-eyes(Wang et al. 2014)	Low-pass Filter	Histograms of the CSI values	DTW & KNN	96% TP 1% FP
2013	Sigg et al. (2013)	Low-pass Filter	Time-domain features	KNN	about 80%

al habit conditions are different, the collected signal will be most likely different. Furthermore, when recognizing multiple activities, signals of similar activities may be added with the increase of activity types to be recognized, which consequently increases the difficulty of recognition and could substantially reduce the accuracy. Finally, even for a specific single activity, the moving speeds of different parts of the body are changing. For example, the speed of the limbs is faster than the trunk when running. In that case, extracting strong-correlation features, such as using algorithms to directly calculate the speed, period, and other features that are not directly related to the signal waveform, can further improve the robustness of behavior sensing.

Improve the Robustness of Environments: Currently, the majority of behavior sensing studies are carried out in controlled environments (e.g. indoor WiFi-based recognition), which requires that the transmitter and the receiver are fixed at a certain location and the volunteers are asked to perform preset actions in the area between the transmitter and the receiver. The collected signals are then used to train the recognition model, which is utilized to classify human actions with high accuracy after a series of processing operation on the raw signals. The experimental environment, the placement scheme of detection equipment, the relative distance between the detected users and sensing devices, and the multi-person complex environment are all factors that seriously impact the accuracy of behavior sensing. The robustness of all the proposed approaches with respect to the environment needs to be improved, in such a way to apply the devised approaches in different environments without changing the recognition model while keeping the same (high) recognition accuracy.

Multi-person Sensing: Compared with the single-person behavior sensing, multi-person sensing is a more challenging problem that impacts the different steps of activity recognition. Indeed, in the multi-person scenario it is necessary to take into account the noise generated by several persons in the preprocessing stage. In the feature extraction step, the superposition signal of the activity and various features fusion needs to be considered. Finally, in the recognition phase, facing more complex features, the optimization requirements of the model are higher. Hence, it is extremely challenging to establish a multi-person detection mechanism and method for further research on behavior sensing.

Acknowledgements Partial work of this paper is supported by the Zhejiang Provincial Natural Science Foundation of China (LY18F020011), Ningbo Natural Science Foundation(2018A610154) and the K. C. Wong Magna Fund in Ningbo University.

References

Aceto G, Ciuonzo D, Montieri A, Pescapè A (2018) Multi-classification Approaches for Classifying Mobile APP Traffic. Journal of Network and Computer Applications 103(C):131–145

Aceto G, Ciuonzo D, Montieri A, Pescapè A (2019a) Know your Big Data Trade-offs when Classifying Encrypted Mobile Traffic with Deep Learning. In: IEEE/IFIP Network Traffic Measurement and Analysis Conference (TMA) 2019, pp 121–128

Aceto G, Ciuonzo D, Montieri A, Pescapè A (2019b) Mobile Encrypted Traffic Classification using Deep Learning: Experimental Evaluation, Lessons Learned, and Challenges. IEEE Transactions on Network and Service Management 16(2):445–458

Adib F, Kabelac Z, Katabi D, et al (2014) 3D Tracking via Body Radio Reflections. In: USENIX NSDI 2014, pp 317–329

Alemdar H, Ersoy C (2017) Multi-Resident Activity Tracking and Recognition in Smart Environments. Journal of Ambient Intelligence and Humanized Computing 8:513–529

Ali K, Liu AX, Wang W, et al (2015) Keystroke Recognition Using WiFi Signals. In: ACM MobiCom 2015, pp 90–102

Aly H, MYoussef (2013) New Insights into WiFi-based Device-Free Localization. In: ACM UbiComp 2013, pp 541–548

Cai JF, Cands EJ, Shen Z (2010) A Singular Value Thresholding Algorithm for Matrix Completion. Siam Journal on Optimization 20(4):1956–1982

Caine K, Fisk A, Rogers W (2016) Benefits and Privacy Concerns of a Home Equipped with a Visual Sensing System: A Perspective from Older Adults. In: The Human Factors and Ergonomics Society Annual Meeting 2016

Casale P, Pujol O, Radeva P (2011) Human Activity Recognition from Accelerometer Data Using a Wearable Device. Pattern Recognition and Image Analysis pp 289–296

Chen H, Liu X, Zhao Z, et al (2019) TaRad: A Thing-centric Sensing System for Detecting Activities of Daily Living. In: the 12th International Conference on Internet and Distributed Computing Systems(IDCS), Napoli, Italy 2019

Chikhaoui B, Ye B, Mihailidis A (2017) Feature-Level Combination of Skeleton Joints and Body Parts for Accurate Aggressive and Agitated Behavior Recognition. Journal of Ambient Intelligence and Humanized Computing 8:957–976

Chikhaoui B, Ye B, Mihailidis A (2018) Aggressive and Agitated Behavior Recognition From Accelerometer Data Using Non-Negative Matrix Factorization. Journal of Ambient Intelligence and Humanized Computing 9(5):1375–1389

Daubechies I, Heil C (1992) Ten Lectures on Wavelets. Computers in Physics 61

Debes C, Merentitis A, Sukhanov S, et al (2016) Monitoring Activities of Daily Living in Smart Homes: Understanding Human Behavior. IEEE Signal Processing Magazine 33(2):81–94

Dickerson R, Gorlin E, Stankovic J (2011) Empath: A Continuous Remote Emotional Health Moni-toring System for Depressive Illness. In: ACM Conference on Wireless Health 2011, pp 5–14

Erickson VL, Carreira-Perpinan MA, Cerpa AE (2011) OBSERVE: Occupancy-based System for Efficient Reduction of HVAC Energy. In: ACM IPSN 2011, pp 258–269

Fornasier M, H Rauhut RW (2011) Low-rank Matrix Recovery via Iteratively Reweighted Least Squares Minimization. SIAM Journal on Optimization 21(4):1614–1640

Franco GC, Gallay F, Berenguer M, Mourrain C, Couturier P (2008) Non-invasive Monitoring of the Activities of Daily Living of Elderly People at Home - A Pilot Study of the Usage of Domestic Ap-pliances. Journal of Telemedicine and Telecare 14(5):231–235

Fujinami T, Miura M, Takatsuka R, et al (2011) Toward Useful Services for Elderly and People with Disabilities, Springer, chap A Study of Long Term Tendencies in Residents Activities of Daily Living at a Group Home for People with Dementia using RFID Slippers, pp 303–307

Galluzzi V, Herman T, Polgreen P (2015) Hand Hygiene Duration and Technique Recognition using Wrist-worn Sensors. In: ACM IPSN 2015, pp 106–117

Hao T, Xing G, Zhou G (2013) iSleep: Unobtrusive Sleep Quality Monitoring using Smartphones. In: ACM SenSys 2013, p 4

Herath S, Harandi M, Porikli F (2017) Going Deeper into Action recognition: A survey. Image and Vision Computing pp 4–21

Keally M, Zhou G, Xing G, et al (2011) Pbn: Towards Practical Activity Recognition using Smartphone-based Body Sensor Networks. In: ACM SenSys 2011, pp 246–259

Kellogg B, Talla V, Gollakota S (2014) Bringing Gesture Recognition to All Devices. In: USENIX NSDI 2014, pp 303–316

Kerola T, Inoue N, Shinoda K (2014) Spectral Graph Skeletons for 3D Action recognition. In: ACCV 2014, pp 417–432

Kosba AE, Saeed A, Youssef M (2012) Robust WLAN Device-free Passive Motion Detection. In: IEEE WCNC 2012, pp 3284–3289

Lao W, Han J, de With P (2009) Automative Video-based Human Motion Analysis for Consumer Surveillance System. IEEE Trans Consumer Electronics 55(2):591–598

Lawton M, Brody E (1970) Assessment of Older People: Self-maintaining and Instrumental Activities of Daily Living. Nursing Research 19(3):278

Lee S, Kim Y, Ahn D, et al (2015) Non-obstructive Room-level Locating System in Home Environments using Activity Fingerprints from Smartwatch. In: ACM UbiComp 2015, pp 939–950

Li H, Yang W, Wang J, et al (2016a) Wi-Finger: Talk to Your Smart Devices with Finger-grained Gesture. In: ACM UbiComp 2016, pp 250–261

Li X, Zhang Y, Li M, I Marsic JY, et al (2016b) Deep Neural Network for RFID-based Activity Recognition. In: ACM MobiCom 2016, pp 24–26

Lin Z, Chen M, Ma Y (2010) The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-rank Matrices. arXiv:10095055

Ling K, Dai H, Liu Y, et al (2018) UltraGesture: Fine-Grained Gesture Sensing and Recognition. In: IEEE SECON 2018, pp 1–9

Liu J, Wang Y, Chen Y, et al (2015) Tracking Vital Signs during Sleep Leveraging Off-The-Shelf WiFi. In: ACM MobiHoc 2015, pp 267–276

Liu Q, Liu Z (2012) A Comparison of Improving Multi-class Imbalance for Internet Traffic Classification. Information Systems Frontiers 16(3):509–521

Liu X, Chen H, Zhang X, et al (2019) Human Action Counting and Recognition with Wi-Fi Signals. In: the 4th International Conference on Computing, Communications and Security (ICCCS), Rome, Italy 2019

Lyonnet B, Ioana C, Amin MG (2010) Human Gait Classification using Microdoppler Time-Frequency Signal Representations. In: IEEE Radar Conference-Proceedings 2010, pp 915–919

Mitchell D, Morrow P, Nugent C (2014) A Sensor and Video based Ontology for Activity Recognition in Smart Environments. IEEE EMBS 2014

Morales F, Toledo P, Sanchis A (2013) Activity Recognition using Hybrid Generative/discriminative models on Home Environments using Binary Sensors. Sensors 13(5):5460–5477

Nguyen V, Ibrahim M, Rupavatharam S, et al (2018) EyeLight: Light-based Occupancy Estimation and Activity Recognition from Shadows on the Floor. In: IEEE INFOCOM 2018

O'Shea TJ, West N, Vondal M, et al (2017) Semi-supervised Radio Signal Identification. In: IEEE ICACT 2017, pp 33–38

Pu Q, Gupta S, Gollakota S, et al (2013) Whole-home Gesture Recognition using Wireless Signals. In: ACM SIGCOMM 2013, pp 27–38

Sabek I, Youssef M (2012) Multi-entity Device-free WLAN Localization. In: IEEE GLOBECOM 2012, pp 2018–2023

Sigg S, Shi S, Ji Y (2013) RF-based Device-free Recognition of Simultaneously Conducted Activities. In: ACM UbiComp 2013, pp 531–540

Spagnolo P, Mazzeo PL, Distante C (eds) (2014) Human Behavior Understanding in Networked Sensing. Springer

Srivastava M, Abdelzaher T, Szymanski B (2012) Human-Centric Sensing. Philosophical Transactions pp 176–197

Tan S, Yang J (2016) WiFinger: Leveraging Commodity WiFi for Fine-grained Finger Gesture Recognition. In: ACM MobiHoc 2016, pp 201–210

Wang W, Liu AX, Shahzad M, et al (2015) Understanding and Modelling of WiFi Signal based Human Activity Recognition. In: ACM MobiCom 2015, pp 65–76

Wang Y, Liu J, Chen Y, et al (2014) E-eyes: Device-free Location-oriented Activity Identification using Fine-grained WiFi Signatures. In: ACM MobiCom 2014, pp

617–628
Wang Y, Wu K, Ni LM (2016) Wifall: Device-free Fall Detection by Wireless Networks. In: IEEE INFOCOM 2016, vol 16, pp 581–594
Wilson J, Patwari N (2011) See-through Walls: Motion Tracking using Variance based Radio Tomography Networks. IEEE Transactions on Mobile Computing 10(5):612–621
World Health Organization (2011) Global Health and Ageing. Tech. rep., US National Institute of Aging
Wu CS, Yang Z, Zhou Z, et al (2015) Non-invasive Detection of Moving and Stationary Human with WiFi. IEEE JSAC 33(11):2329–2342
Wu X, Chu Z, Yang P, et al (2018) TW-See: Human Activity Recognition through the Wall with Commodity Wi-Fi Devices. IEEE Transactions on Vehicular Technology 68(1):306–319
Xi W, Zhao J, Li XY, et al (2014) Electronic Frog Eye: Counting Crowd using WiFi. In: IEEE INFOCOM 2014, pp 361–369
Xin T, et al (2018) FreeSense: Human-Behavior Understanding using Wi-Fi Signals. Journal of Ambient Intelligence and Humanized Computing 9(5):1611–1622
Xu Y, Yang W, Wang J, et al (2018) WiStep: Device-free Step Counting with WiFi Signals. In: ACM IMWUT 2018, vol 1, p 172
Yang J, Lee J, Choi J (2011) Activity Recognition based on RFID Object Usage for Smart Mobile Devices. Journal of Computer Science and Technology 26(2):239–246
Yang X, Tian Y (2012) Eigen Joints-based Action Recognition using Naive-Bayes-Nearest-Neighbor. In: IEEE CVPRW 2012, pp 14–19
Yang Y, Hao J, Luo J, Pan SJ (2017) Ceilingsee: Device-free Occupancy Inference through Lighting Infrastructure based LED Sensing. In: IEEE Percom 2017
Yatani K, Truong KN (2012) Bodyscope: a Wearable Acoustic Sensor for Activity Recognition. In: ACM UbiComp 2012, pp 341–350
Yun J, Lee SS (2014) Human Movement Detection and Identification using Pyroelectric Infra-red Sensors. Sensors 14(5):8057–8081
Zeng Y, Pathak PH, Mohapatra P (2016) WiWho: Wifi-based Person Identification in Smart Spaces. In: ACM IPSN 2016, p 4
Zhou Z, Yang Z, Wu C, et al (2013) Towards Omni-directional Passive Human Detection. In: IEEE INFOCOM 2013, pp 3057–3065