A Survey of Human Action Recognition using Accelerometer Data

Amira Mimouna and Anouar Ben Khalifa

Abstract Recognizing human actions and analyzing human behaviors from accelerometer data has become a challenging task. Hence, Human Action Recognition (HAR) using inertial sensors have been addressed in a plethora of various review papers. This chapter provides a detailed state of art survey of HAR exploiting acceleration data. Considering different modalities, we prove that the accelerometer is one of the most promising sensors in this field by presenting an overview of its applications. In addition, we propose a comprehensive review of recent studies in this domain along different views: from data modalities to feature extraction and classification steps. Moreover, we list the most publicly available databases that include accelerometer data. Afterwards, we used a multi-level fusion framework that includes signal-level, feature-level, score level and the decision level fusion in order to improve the recognition performance. For the classification, we took advantage of the support vector machine with features from the time-frequency domain. The proposed framework was evaluated using three public datasets: WARD, MHAD and Realdisp. The results obtained from the fusion techniques indicate that the score level provides a satisfactory performance compared to the other levels and with the use of each accelerometer separately.

Keywords Human action recognition · Accelerometer · Accelerometer applications · Multi-level fusion

A. Mimouna (✉) · A. Ben Khalifa
LATIS- Laboratory of Advanced Technology and Intelligent Systems, Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, Sousse, Tunisia
e-mail: amira.mimouna@yahoo.com

A. Ben Khalifa
e-mail: anouar.benkhalifa@eniso.rnu.tn

1 Introduction

Do the human actions developed in a day present a good image of the overall home-ostasis of the person? The execution of an action such as sitting down or standing for a long period, rather than jumping or lying down, and the speed of the accomplishment of these tasks present valuable information about a person's daily activity. It reflects his vitality, therefore his state of health and even his psychological state. Hence, monitoring and supervising the activities of everyday living has become a crucial task to enhance the quality of our lives.

Human actions can be classified into four classes, relying on their complexity: gestures, actions, interactions and a group of activities (Aggarwal and Ryoo 2011; Jegham et al. 2019). Gestures consist of elementary movements of a body part, for example: 'raising an arm'. Actions include gestures that are temporally ordered, for instance: 'walking' or 'waving'. In addition, interactions involve two persons or more as 'two persons shaking hands', also it exists a human-object interaction between humans and objects, such as 'a person is giving a cup to another'; finally, group of activities that include several persons and/or objects such as 'a group having a meeting'.

The principal aim of human action recognition is automatically detecting and analyzing human activities, then interpreting continuously and successfully the situation (Chen and Shen 2017). Thus, this filed of research has been unavoidable in several areas including: health care (Ameur et al. 2016; Jain and Kanhangad 2018), surveillance (Lejmi et al. 2017), human-computer interaction (Nuno et al. 2017), virtual reality (Kwon et al. 2017), gaming (Namal et al. 2006), etc.

To guarantee the recognition and the analysis of human behavior, several researchers have exploited different types of technologies in their work, including cameras, Kinect, accelerometers, gyroscopes, microphones, MoCap (motion capture), RFID (radio frequency identification), etc.

In fact, the employment of microphones in the field of human behavior analysis is becoming more and more important in various fields, such as robotic assistance, action recognition, etc. However; the presence of noise and the distance of the person with regard to the microphone are still a challenge (Rodomagoulakis et al. 2016).

Although several works have used RGB cameras because it provides rich information of the scene, the recognition based on video sequence has its own limitations, such as sensitivity to lighting, background disorder and occlusion (Jegham and Ben Khalifa 2017; Chebli and Ben Khalifa 2018). In addition, this approach is limited to a fixed area of view outlined by the camera position and for many people, who feel uncomfortable when they are monitored continuously, cameras are intrusive (Cornacchia et al. 2017; Lejmi et al. 2019).

Based on depth sensors that provide 3D action data, the human action recognition has gained more improvement. For the Kinect, it is insensitive to changes in lighting and ensures recognition of actions in the dark. Nevertheless, the subject must always be present in the field of view of the Kinect and the images present different noise.

Table 1 A summary of some limitations of different sensors for human action recognition

Sensors	Limitations
Microphone	– Presence of noise
	– The distance of the person with regard to the microphone
Camera	– Conditions of acquisition
	– Occlusions
	– Intrusive
Kinect	– The person must always be in the Kinect's point of view
	– The images present different noise
	– Requires different material sources
MoCap	– Occlusions
	– Constrained place
	– Complex calibration and requires expensive equipment
RFID	– Objects must be equipped with RFID tags
	– The person must wear an RFID

Motion capture is a sector of research in full evolution. However, the use of such a technology requires a procedure of boring calibration and additional expensive equipment. Furthermore, MoCap has many challenges, for example: occlusion and a constrained space.

In the case of recognizing human actions from the radio-identification RFID, which informs us about the place of the person, RFID labels must equip the objects, which interact with the person, and the port of this sensor by the user is necessary.

A summary of some limitations of different sensors associated with human action recognition is presented in Table 1.

With the progress of microelectronics, human action recognition using wearable inertial sensors, such as the accelerometer or the gyroscope, has been acquiring more and more attentiveness from many researchers. Moreover, the integration of these sensors into different devices, which become a part of people's daily living (such as: smartphones, smart watches, sport medical bracelet, etc.) has opened the way to the advancement of the human action recognition. Among the technologies that recognize human activities, inertial wearable sensors seem to be the most promising. Indeed, their lightweight, small size, and low cost have attracted many researchers (Mimouna et al. 2018). Moreover, the low energy consumption and the reduced computational power provide a long-period recordings and continual interaction compared with based-image processing systems.

Undoubtedly, wearing these sensors is easy and using such a technology can ensure the recognition in darkness. Thanks to all these advantages, the accelerometer, which provides 3-axis accelerations, has been exploited in a diversity of applications in order to detect and analyze human activities.

Furthermore, to enhance the recognition performance, some researchers proposed to combine two different modalities to deal with several realistic events that may appear in the real world, for instance, fusing data from a depth image and data from a wearable inertial sensor as shown in Chen et al. (2015), Malawski and Gałka (2018).

To the best of our knowledge, this is the first research attempt to exploit the potential of the triaxial accelerometer and its employment in various fields especially in HAR. The aim of this chapter is: (i) to present an overview of the state of the art of accelerometers' applications and practicality we focus on the field of HAR exploiting accelerometer data, and (ii) to expose a fusion framework which consists of coupling several information acquired from numerous levels.

As we discussed, several modalities introduced to recognize human activities and as the accelerometer seems to be the most effective; we will present an accelerometer's review and its applications in Sect. 2. The third section is reserved for introducing the field of human action recognition exploiting accelerometers data, in this section we present challenges, various applications related to this field and several approaches employed to guarantee action recognition. Datasets based on inertial sensors are introduced in Sect. 4. We give a detailed description of the fusion framework in Sect. 5. The experimental results are reported in Sect. 6. The seventh section provides the conclusion.

## 2 Accelerometer's Review and Applications

Accelerometers are used to determine the measurement of changes in velocity. There exist two main modes of acceleration measured by this sensor: the first is the linear acceleration, which is the acceleration measured when the change in velocity is in the signal direction, and the second is the centrifugal acceleration, which is the measurement of the displacement of an object in a circle.

The triaxial accelerometer measures the acceleration following three directions X, Y and Z, as shown in Fig. 1 which represents accelerometer data acquired when moving the phone. It is a kinematic sensor existing in several devices. In addition to game consoles, mobile terminals and automobiles, accelerometers are now present in a large number of connected objects; we mention intelligent textiles, connected watches, cameras, prostheses, shoes, drones, robots, sports and medical bracelets, etc.

Thanks to its many benefits, nowadays, the accelerometer is present in a variety of applications which they will be detailed below.

Recently, monitoring road conditions become necessary to insure safety to vulnerable road users, and also to evaluate the state of the roads. Allouch et al. (2017), developed an android application named RoadSense to predict road conditions using the accelerometer and the gyroscope integrated into the smartphone. According to the results, it guarantees high performance with an accuracy of 98.6%.

In augmented reality, Unuma and Komuro (2015) proposed a natural 3D interaction system, the user can interact with virtual objects superimposed on the real image
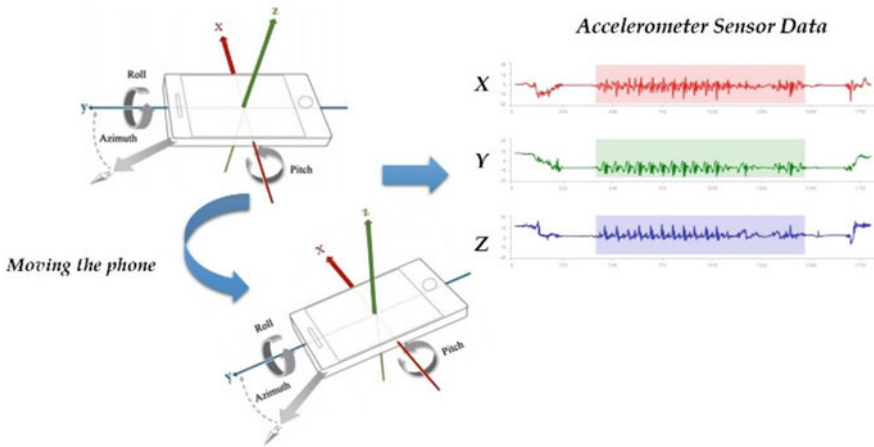
Fig. 1 Accelerometer sensor data acquired when moving the phone

using his hand. With the aim to insure natural interaction, a triaxial accelerometer is fixed on the depth camera. Thus, when the user pushes a virtual ball, it rolls immediately, and he can just find it when he displaces the mobile display even if the ball quits the screen.

Over the last decade, prosthetics have been evolving owing to the advancement of microelectronics sensors and their facility of incorporation to these prosthetics. In (Beyrouthy et al. 2016), an EEG mind-controlled prosthetic arm is developed. This smart prosthetic arm is controlled through brain commands and it is outfitted with a network of sensors. This smart network provides the prosthetic arm with normal hand movements and intelligent reflexes. Furthermore, the proposed prosthesis has been developed in order to ameliorate the quality of life of patients with a low cost.

In work environments, accelerometers embedded in mobile phones are used for detecting stress levels because it affects the health of workers. Data acquired from the accelerometer was utilised to differentiate humans' behaviours. For 8 weeks, 30 subjects with smartphones from two organizations participated in this study and they noted their stress levels three times while working. Besides, three levels are introduced: low, medium and high stress. An accuracy of 70% for user-specific modal was achieved (Garcia-Ceja et al. 2016).

Also based on a network of sensors embedded in a mobile phone, including the accelerometer and the GPS, Castignani et al. (2015) proposed a new application named SenseFleet, which is capable of detecting risky driving events by identifying several events, such as braking, steering, accelerating and over-speeding. Moreover, the obtained results show that the application is able to precisely identify risky events, it can also differentiate between the drivers' behaviours, for instance calm and aggressive drivers.

Air pollution caused by gaseous emission from vehicles has been increasing with the advancement of economy and vehicles. Traffic conditions are one of the most

Table 2 A summary of accelerometer's applications

Refs.	Fields	Applications
Chen and Shen (2017)	Traffic	– Monitoring roads conditions
		– Detection risky driving events
Jain and Kanhangad (2018)	Augmented reality	– The accelerometer provides the tilt of the depth camera to interact with virtual objects
Kowalczuk and Merta (2015)	Robotic	– Determination of the position of a mobile robot
Kyberd and Poulton (2017)	Medical	– Prosthesis
		– Diagnostic of the heart valve disease
Garcia-Ceja et al. (2016)	Healthcare	– Detecting and monitoring stress
Zhang et al. (2016)	Pollution	– Detecting driving events to estimate the quality of vehicle emission
Rastegari et al. (2017)	Industry	– Machinery maintenance
Höflinger et al. (2015)	Biology	– Monitoring movement patterns in animal models of disease

affecting elements of air pollution, thus, a method based on levels of service is proposed in Zhang et al. (2016), to estimate emissions under various traffic conditions. Accelerometer data was used to describe driving events, which are the characteristics of the vehicle movements that affect the quantity of emission.

In the field of industry, accelerometers are widely used to give an account of the vibration and its changes in the aim of permitting the user to monitor machines, to detect faults and to minimize its suspension. Rastegari et al. (2017) focus on condition based maintenance as regards to machine tools, particularly concentrating on vibration monitoring approaches. Hence, accelerometers are fixed to the spindle units, then, data are transferred to the computer as a dataset in order to be analysed.

A summary of accelerometer's applications is provided in Table 2 in the following.

In conclusion, the accelerometer is exploited in very fields, and is particularly employed to ensure human action recognition, this point will be detailed in the following section.

3 HAR Using Accelerometer Data

3.1 Challenges

Although the human action recognition using accelerometers data continues to progress, the recognition accuracy is affected by many challenges in this field. Firstly, people have different motion models: every subject has his unique style of execution as shown in Fig. 2.

Moreover, for the same person, the action may differ from one repetition to another: the action can be shorter or longer as provided in Fig. 3.

Furthermore, the placement of the on-body sensors presents an important defiance, for example: when a person is jogging, the data collected from an accelerometer attached to the wrist is different from data acquired from an accelerometer fixed to the thigh. Figure 4 presents signals recuperate from six different localizations.

In addition, the translation and the rotation of the sensor, when recording the action, may influence the measurement so it may affect the recognition performance. Thus, the number, the position and the type of the accelerometer are principally related to the application. Besides, the complexity of actions and the transition period between two successive actions lead to an additive challenge. Additionally, people performing multiple activities simultaneously might cause confusions.
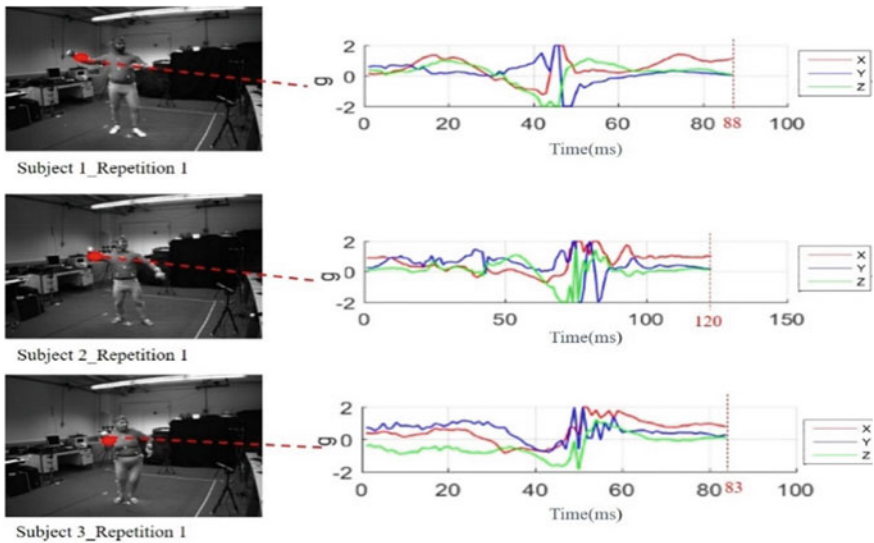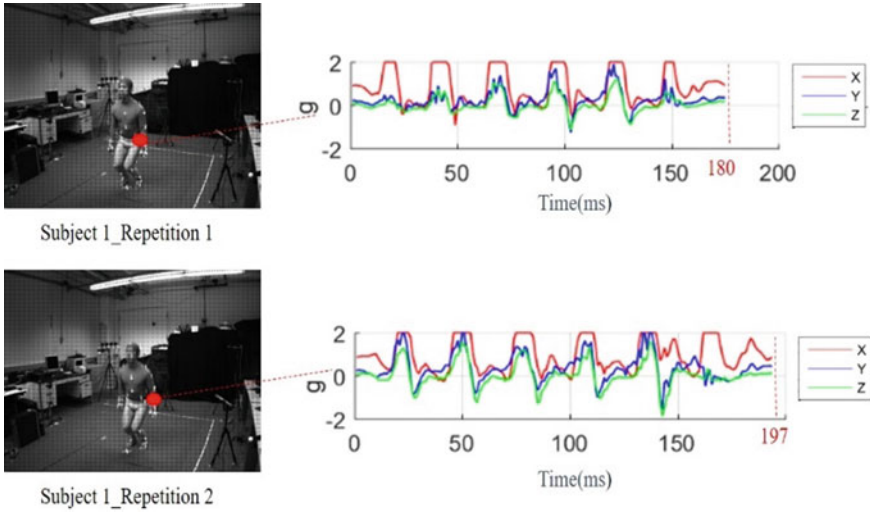

Fig. 2 Inter-class challenge

Subject 1_Repetition 1

Subject 1_Repetition 2

Fig. 3 Intra-class challenge

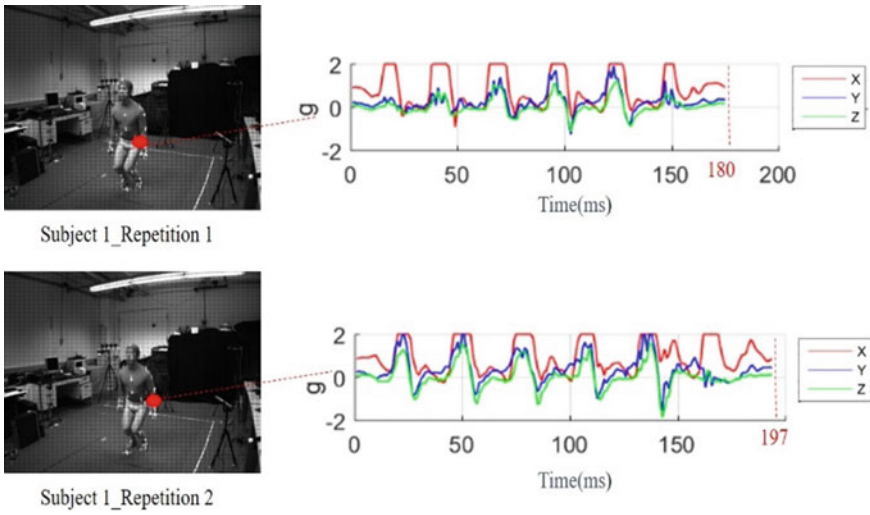Subject 1_Repetition 1

Subject 1_Repetition 2

Fig. 4 Signals acquired from six different positions

3.2 HAR Applications

Analysing human actions using wearable sensors, as the accelerometer, has become an increasingly unavoidable area of research in various fields including: medical, virtual reality, sport, security, surveillance, education, etc. In the following, we will expose several applications presented in Table 3 to outline the use of the accelerometer in HAR.

Surgeries are complex tasks accomplished in stressful areas (Zia et al. 2018). Therefore, the immersive virtual reality provides virtual environments to surgeons and trainees to be trained in realistic conditions to ensure the patient's safety and to attenuate errors. Various technologies are used in this field, including wearable sensors, which track the user's motions in order to gain surgical expertise (Dargar et al. 2015).

Laghari et al. (2016) focused on developing a biometric authentication application based on accelerometer data acquired from the smartphone. Indeed, the user performs his signature by handling the phone in his hand and moving it. Ten volunteers participated in this work; each subject had to perform his signature 6 times. The signal

Table 3 Applications of human action recognition using accelerometer data

Refs.	Application field	Applications
Dargar et al. (2015)	Education	– Training in virtual environments to gain expertise
Zhang et al. (2015)	Virtual reality	– Training
Laghari et al. (2016)	Security	– Biometric authentication application for smartphones
		– Medical adherence
Pepa et al. (2015), Kau and Chen (2015), Figueiredo et al. (2016)	Medical	– Diagnostic of various diseases: Parkinson's disease, epilepsy seizure, Alzheimer's disease, etc.
		– Fall accident detection
Sen et al. (2015)	Healthcare	– Monitoring sleep
		– Diet monitoring
		– Assessment of elderly people
Ferhat et al.	Surveillance	– Assessment of children's behaviours
Hung et al. (2015)	Automation	– Controlling the daily environment
Hidayat et al. (2016)	Gaming	– Detecting gestures and actions to play
Neto et al. (2009)	Industry	– Controlling an industrial robot
Koskimäki and Siirtola (2014)	Sport	– Recognizing gym exercises

matching was used as an identification approach. With regards to the traditional and the graphical techniques, this method is more secure with a false rejection rate of 6.87%.

Kalantarian et al. (2017) proposed an android application implemented on a smartwatch to detect various motions related to medical adherence. Furthermore, the system detects when the bottle is twisted to open it using the accelerometer data and then, the act of revolving the palm to retrieve the pill is identified using gyroscope data. Although the system is sensitive to how to remove the pill, it needed less human involvement for medication adherence with regard to nurses' calls or other forms.

Parkinson's disease is an advancing neurological disorder that affects the basal ganglia. Freezing of Gait (FOG) is one of the most frequent motor disorders for advanced Parkinson disease that can diminish the quality of life and it can be defined as a gait disturbance. Pepa et al. (2015) proposed a smartphone-based application that can detect FOG occurrences and is able to send an acoustic feedback to help patients restore walking. In addition, tested on 18 patients, this method provides an 82.34% of sensitivity.

Kau et al. (2015) used the triaxial accelerometer and the electronic compass integrated in the smartphone, which was located in the pocket of the subject to detect fall accidents. If the system detects a fall event, it will send the user's position identified by the GPS to the rescue center via Wi-Fi or the 3G network. Thus, the user can receive medical help straightaway. An accuracy of 92% is achieved using this algorithm with 450 test actions of 9 types that include a fall event.

Wearable inertial sensors are nowadays used to assist therapeutic movements. In (López et al. 2015), two sensors are worn on the forearm and the upper arm to identify the quality of the patient's movements and observe his/her recovery. Besides, the aim of the study is to define intra and inter-group dissimilarity between a given number of movements accomplished by young people, with regard to motions given by therapists.

Human action recognition is used to analyse children's behaviour and to follow their health and development. Indeed, children's actions can be limited to walking, playing, sitting, sleeping and hand motion. A kindergarten system was developed using acceleration information acquired from the accelerometer fixed on the child's hand, then, these information were anlysed to present a global state of the child's health to parents and child-minders (Kurashima and Suzuki 2015).

The assessment of the elderly people during their daily life became a crucial challenge in order to ensure their safety, autonomy and healthcare. Ferhat et al. focused on recognizing and monitoring elderly people using three inertial units that were mounted on the chest, the right thigh and the left ankle. Additionally, based on real-time technique and data transmission, the subject's motions were continually monitored by healthcare suppliers all along daily activities and abnormal events are detected to intervene.

Over the last decade, home automation has become an important field of research to control the daily environment. In (Hung et al. 2015), a hand gesture recognition belt was developed using an accelerometer and a gyroscope to control a LED array

lamp. Indeed, when the user shakes his hand up, the LED turns on and inversely. Consequently, as the user's palm is shaking, the luminosity of the LED can dims.

In the gaming word, the advancement occurs expeditiously. Hidayat et al. (2016) used a Wii remote as a controller of a fighting game. The Wii remote transfers data obtained from the accelerometer that detects gestures or motions of the hand. Then, when the movement is identified, it will be visualized in the built based-Unity 3D game as a player's action.

Neto et al. (2009), developed a system based on two triaxial accelerometers, for the purpose of controlling an industrial robot rather than programing it with typical techniques. Furthermore, the sensors were fixed on the human arms in order to capture its gestures and postures, so the robot can start the movements approximately while the user begins to perform a motion. Besides, a higher performance was achieved using this approach with a recognition rate of 92%.

3.3   Related Work

Human action recognition using acceleration information has been employed in several application areas mentioned previously; in fact, various approaches described in this section have been proposed to address this challenge.

Pre-processing is considered as a one of the most critical steps that includes replacing missing data or filtering it. Before the feature extraction step, raw data acquired from sensors are generally divided into small segments using windowing technique. In fact, various windowing approaches are used in this level: (i) sliding window that is the most commonly used owing to its facility of implementation and its guaranteed high accuracy, it consists of dividing signals into fixed length windows with or without overlap; (ii) the defined activity windows that resides with the division of the data based on the detection of activity changes; (iii) the defined event windows, where pre-processing is needed to find particular events; (iv) the dynamic sliding window that was developed to overpass the fixed-length of the sliding window technique, the main idea of this novel activity signal segmentation approach is that the window size could be dynamically adapted by using the signal information to determine the most effective segmentation.

Afterwards, feature extraction is considered as a crucial step; which consists of extracting quantities to characterize each performed action. Many researchers tended to extract features commonly from: time domain, frequency domain and time-frequency domain. Time domain characteristics include mean, maximum, median, skewness, variance, etc. Frequency-domain features incorporate peak frequency, signal energy, also the calculation of the power spectral density (PSD) and the utilization of the Fast Fourier Transform (FFT), etc. Furthermore, wavelet transform is the most common technique used to extract features from the time-frequency domain. Adding to this, it exists other techniques employed to extract features from accelerometer signals to differentiate actions such as the Dynamic Time Warping (DTW).

In many works, researchers employed a feature selection process, which consists of selecting a subset of appropriate features from the original features, because the use of inappropriate or redundant characteristics may decrease the performance of the classifier. This process reduces the number of features and the computation time. Generally, it exists three classes in feature selecting: (i) filter methods, (ii) wrapper methods, (iii) hybrid methods. The filter-based method evaluates features without any classifier, so it classes a set of selected features according to the estimated weights of each feature. Different from the filter methods, wrapper, which ensures often the best results, uses classifier accuracies to evaluate the selected subset. Eventually, the hybrid methods consist of selecting the most appropriate features due to some internal parameters of the machine-learning algorithm.

Feature vectors, obtained after extraction/selecting features from raw data, are used in order to train the classification algorithm. Indeed, to ensure this step, many machine learning techniques are employed, which are divided into two principal approaches: supervised and unsupervised methods. In addition, the supervised techniques are based on labeled activity data such as K-nearest neighbours (K-NN), Artificial Neural Networks (ANN-s), Support Vector Machines (SVMs), Decision tree (DT) and Random Forest (RF). Concerning the unsupervised approaches, which are linked with unlabeled data, we can site the Hidden Markov Model (HHM), the K-means, and the Gaussian Mixture Models (GMMs).

Some of the common works introduced to recognize and analyse human actions are presented in the following.

In (López et al. 2015), Lopez et al. proposed a novel method to detect and characterize walking and jogging using a triaxial accelerometer. Actually, the kurtosis of wavelet coefficients or the autocorrelation of the acceleration data was used for the detection. This methodology was tested on three different datasets of walking and jogging.

Lubina et al. (2015) evaluated the application of artificial neural networks (ANNs) to recognize human activities using accelerometer signals. Five accelerometers were fixed on the back, two on the waist laterally and two on the ankles, and 25 subjects were called to perform a set of predefined actions such as sitting down and walking. The obtained signals were firstly filtered using a median filter, then they were partitioned into non-overlapping windows with a length of 0.5s. Afterwards, statistical features were extracted, such as the mean, the sum of squares and the root mean square to train the ANNs. Despite the fact that the implementation of the Fisher Linear Discriminant shows that some features help to discriminate similar actions, none of the axes or the features or the sensors can be neglected.

For monitoring daily life activities, Wang et al. (2016) used a single wearable accelerometer that was attached to the waist and the left ankle respectively with a view to diminish the effect of sensor placement. An ensemble empirical mode decomposition (EEMD), which is a time-analysis technique is introduced in this study. Then, feature selection is insured using a game theory to select relevant features. K-NN and SVM are employed to classify human activities captured from the waist and the ankle. Compared with other works, the results obtained using the proposed method, which selects fewer features, show a better classification.

Monitoring sleep has gained the attention of numerous researchers as it affects our psychological and emotional health. Therefore, Yunyoung et al. (2016) focused on identifying sleep quality based on the triaxial accelerometer and the pressure sensor, and they used various physiological parameters. Additionally, data obtained from the accelerometer determined the sleeping posture and activity. Besides, the proposed algorithm based on a sensor fusion framework effectively detected sleeping and waking situations.

Luštrek et al. (2015) suggested an approach to recognize indispensable lifestyle activities of diabetic patients, using sensors embedded on the smartphone, in order to monitor their lifestyle since it affects the disease. A set of activities was introduced in this study such as eating, sleeping, working, and transport. Five volunteers carried a smartphone and an EEG monitor during two weeks. Furthermore, several features were derived from sensors, such as the user's location, the ambient sound and the acceleration features to train various classifiers to recognize the user's action, such as SVM, RF, and Naïve Bayes. Based on different experiments, the results obtained show that the vote provides a higher accuracy, which combine several machine learning algorithms. To improve the classification rate, they proposed to introduce a final machine learning approach, thus, the accuracy went from 0.77 to 0.88. Nevertheless, it exists some misclassification between the activities such as eating and out.

Noor et al. (2015) proposed a new approach of activity signal segmentation using triaxial accelerometer that consists of a dynamic sliding window. The main aim of this method is to recognize static and dynamic activities as well as transitional activities. Initially, a small window size is adjusted to segment static and dynamic activity signals, then the window length is extended in order to encompass the signal that it can be sometimes longer than the initiated window. Moreover, the dynamic sliding window is used to automatically determinate the optimum window size while the signal is being evaluated. A triaxial accelerometer was fixed on the right waist and three subjects performed several actions such as walking, sitting to lying, standing to sit, etc. and each subject repeated each action five times. For pre-processing, a moving average filter is employed, then a 3s sliding window is used to segment the signal, after that the window length is limited to 1.5s with 50% overlapping rate with the previous window. 117 features are extracted from raw data including standard deviation, spectral entropy, maximum, etc. Afterward, relevant features are selected using Relief-F method. Decision Tree was chosen to classify activities which provided an accuracy of 96% and the transitional activities were effectively recognized.

In (Tran and Phan 2016), sensors integrated on the smartphone were used to develop an android real-time system that is able to recognize human actions. Six actions were introduced such as walking, lying down and sitting. Furthermore, SVM was employed to classify the actions and 248 features were extracted from raw data including mean, minimum, energy, etc. The android system compares the performed activity with its model. Thus an accuracy of 89.59% is achieved using this method. A summary of several approaches introduced for human action recognition using accelerometer data is provided in Table 4.

Table 4 A summary of various approaches introduced for human action recognition using accelerometer data

Refs.	Database	Number and placement of the accelerometer	Feature extraction	Feature selection	Classification	Accuracy	
Ferhat et al.	6 subjects 12 actions	3 accelerometers fixed on: the chest, the right thigh and the left ankle	– 11 time domain features: mean, median, peak to peak, variance, etc. – 6 frequency domain features: energy spectrum, energy of the wavelet coefficients,etc. – Fast Fourier Transform – Wavelet analysis	Random Forests – k-NN	k-NN – RF – SVM – HMM – K-means – GMM	: 98% RF SVM HMM K-means GMM	: 98% : 92% : 67% : 51% : 64%
Noor et al. (2015)	3 subjects 11 activities	1 accelerometer mounted on the right waist	– Sliding window technique: initial length is 3s then the size 1.5 s with 50% overlapping – Dynamic sliding window – Time-domain features	Relief-F	– Decision Tree	DT	: 96%
Tran and Phan (2016)	10 subjects 6 actions	The phone is placed in the pocket	– Sliding window technique – Features: mean, entropy, etc.	–	– SVM	SVM :	89.59%
Safi et al. (2016)	6 subjects 12 actions	3 accelerometers placed on: the chest, the right thigh and the left ankle	– The raw data are used as an input for classification algorithm	–	– K-means – GMM – HMM – HMMR	K-means: GMM: HMM: HMMR:	77% 70% 86% 97%

(continued)

Table 4 (continued)

Reference	Database	Number and placement of the accelerometer	Feature extraction	Feature selection	Classification	Accuracy
Chen et al. (2015)	MHAD	6 accelerometers worn on: the wrists, the ankles and the hips	– Sliding window technique. – Statistical features: rms, variance, mean, std		A_1:86% –SVM A_5:56%	A_2: 85% A_3:71% A_6 : 57%	A_4 :72%
Erdas et al. (2016)	15 subjects 7 activities	The mobile phone is fixed on the chest	– 17 time-domain features: mean, rms, minimum, etc. – 6 frequency-domain features: Fast Fourier Transform – 9 time-frequency domain features: Wavelet Transform	– Chi-square selection – Correlation-based feature selection (CFS) – ReliefF – Gain-ratio-based selection	– RF – k-NN – SVM	RF: 88% K-NN: 80% SVM: 67%
Zhao et al. (2014)	23 subjects 6 actions	The smart watch is worn on the dominant hand	– The Short-time Fast Fourier Transform using a hamming window is employed	–	– HMM	HMM: 97%
Lee et al. (2017)	5 subjects 3 activities	The smartphone is carried in various locations (held in hand, carried in a bag, placed in a pocket)	– Sliding window technique transformed into a magnitude vector	– Acceleration data – 1D- CNN – RF	1D-CNN:	92.71%

4   Datasets

A large number of public human action recognition datasets have been introduced based on inertial sensors. We distinguish uni-modal and multimodal databases. This section consists of a review of various databases that have been included to recognize human actions captured from accelerometer data.

4.1   Uni-Modal Databases

4.1.1   MIT PlaceLab Dataset

This Dataset is one of the first public databases in this field of research. To record this dataset, five accelerometers and a wireless heart rate monitor were utilised, each accelerometer is mounted on the left and right arm, the left and right leg and one on the hip. During a four-hour period, one person is asked to perform a set of activities wearing these sensors, including house-holding activities, such as preparing a recipe, cleaning the kitchen, doing the laundry and other types of everyday tasks, for instance talking to the phone or answering emails. However, data existing in this database are collected from one person, which could present a real problem because each person has its own way to perform activities, so the characteristics of the action are poorly represented.

4.1.2   UC Berkeley WARD Dataset

WARD (Wearable action recognition database) is a public human action recognition dataset developed by the University of California. It consists of continuous sequences of human actions measured by a network of wearable motion sensors. The sensors are attached at five body locations: the two wrists, the waist, and the two ankles. Each wireless sensor includes a triaxial accelerometer and a biaxial gyroscope. The database contains 20 subjects: 13 male and 7 female and includes a rich set of activities that involve some of the most frequent actions in the daily life, such as standing, sitting, walking and jumping. It is true that WARD covers the most typical human actions and includes a sufficient number of persons, but some of the data is missed due to battery failure.

4.1.3   USC-HAD

A single inertial sensor was used to evaluate 12 different actions performed by 14 subjects (7 males and 7 females): each action is repeated four times. This database includes a considerable number of subjects of different sexes and the activities con-

sidered are among the most basic and common human activities in people's daily lives. However, data is acquired form a single accelerometer

4.1.4   REALDISP (REAListic Sensor DISPlacement)

Realistic sensor displacement is a benchmark dataset dedicated for human action recognition. This set was collected to evaluate the effects of sensor displacement in activities recognition, "which can be caused by a loose fitting of sensors, or a displacement by the users themselves". Indeed, three scenarios were introduced: ideal-placement, self-placement, and induced-displacement. The first scenario is "Ideal placement" or default scenario, where sensors are arranged by the instructor on pre-defined locations of the body. The second scenario is the "Self-placement", where the user is asked to position 3 sensors himself on the body part specified by the instructor. This scenario tries to simulate some of the variability that may occur in the day-to-day usage of an activity recognition system, involving wearable or self-attached sensors. And for the last scenario, the instructor introduces a de-positioning of the sensors using rotations and translations with respect to the ideal placement. This database consists of 33 different physical activities that can be classified as warming up, cooling down and fitness activities and it includes 17 subjects. Data was measured from nine different sensors that contain a 3D accelerometer; a 3D gyroscope, a 3D magnetic field orientation and a 4D quaternion that are attached overall body parts.

Table 5 lists a summary of some uni-modal publicly available databases using accelerometers for human action recognition.

4.2   Multimodal Databases

4.2.1   CMU Multimodal Activity Database

This Database was developed in the Carnegie Mellon University that contains different multimodal measures of the human activity of subjects performing the tasks involved in cooking and food preparation. It contains video, audio, RFID tags and motion capture system based on-body markers and physiological sensors such as galvanic skin response (GSR) and skin temperature. In addition, 43 subjects were asked to perform food preparation and cook five recipes while the sensors were placed all over the body: both forearms and upper arms, left and right calves and thighs, abdomen, and both wrists. This set involves a very large population but it is specific to just cooking activities.

Table 5 Summary of uni-modal publicly available databases using accelerometer data for human action recognition. N_s: Number of Subjects. N_A: Number of Accelerometers

Database	N_s	Actions	N_A	Sensor locations	Comments
MIT Place Lab	1	10 activities: Prepare a recipe Do a load of dishes Search for items, etc.	5	Right arm, Left arm Right leg, Left leg Hip	– Just one person performing the activities – Activities considered are imprecise
USC Berkeley WARD	20	13 activities: Stand Sit Walk Forward Walk left-cercle, etc.	5	Left wrist, Right wrist, Front centre of the waist Left ankle, Right ankle	– Part of the data is missing
USC-HAD	14	12 activities: Walk forward Walk left Walk right Walk up stairs, etc.	1	Front right hip	– Data collected form a single accelerometer
Realistic sensor displacement benchmark dataset (Realdisp)	17	33 activities: Walking Jogging Running Jump up Trunk twist, etc.	9	Left calf, Left thigh Right calf, Right thigh Back, Left lower arm Left upper arm, Right lower arm, Right upper arm	– The dataset covers a wide range of physical activities and a number of wearable sensors

4.2.2 OPPORTUNITY Dataset

The opportunity dataset is collected from a European research project called OPPOR-TUNITY, which concentrated on daily home activities especially on preparing breakfast. This dataset includes different modalities such as accelerometers, gyroscopes, magnetometers, microphones, and cameras. 12 subjects were asked to perform a sequence of daily morning activities including grooming a room, preparing and drinking coffee. Different modalities were used to collect data, such as a camera, a microphone, an accelerometer, and a gyroscope.

4.2.3   Berkeley MHAD: Multimodal Human Action Database

MHAD contains temporally synchronized and geometrically calibrated data acquired from an optical motion capture system, multi stereo cameras from multiple views, depth sensors, accelerometers and microphones. 11 subjects (7male and 7 female) participated in the data collection and were asked to perform 11 actions with five repetitions for each action, including jumping in place, jumping jacks, bending, waving two hands. Prior to each recording, the subjects were given instructions on what action to perform; however, no specific details were given on how the action should be executed (i.e., performance style or speed). In addition, six accelerometers were fixed on the wrists, ankles and hips, and the two Kinect were placed in opposite directions. This database contains 660 action sequences.

4.2.4   UTD-MHAD: University of Texas at Dallas Multimodal Human Action Dataset

UTD-MHAD is a publicly available multimodal human action recognition data set collected from a Kinect and a wearable inertial sensor measuring a 3-axis : accelerometer, velocity signals and magnetic strength. The dataset contains 8 subjects (4 female and 4 male) and 27 different actions: right arm swiping to the left, right arm swiping to the right, right hand waving, two-hand clapping, right arm throwing, crossing the arms, etc. Each person repeats each action four times with the wearable inertial sensor fixed on the subject's right wrist or right thigh depending on whether the action was mostly an arm or a leg type of action.

4.2.5   Huawei/3DLife Dataset

The Huawei/3DLife is a multimodal dataset developed for a 3D human reconstruction and action recognition Grand Challenge in 2013. For this challenge, two datasets were provided: Dataset 1 contains a synchronized RGB-plus-Depth video captured by five Kinects, as well as multiple-Kinects audio and eight inertial sensors covering the whole body. The inertial sensors were placed on: the left wrist, the right wrist, the chest, the hips, the right ankle, the left ankle, the right foot and the left foot. This dataset includes two sessions with different spatial arrangements of the sensors. 17 subjects performed a set of 22 repetitive actions, and each action was performed 5 times. It consists approximately 3740 captured gestures. The performed actions can be classified into i) Simple actions that involve mainly the upper human body, ii) Training exercises, iii) Sports related activities and iv) Static gestures.

   With regard to Dataset 2, it was captured in Berlin and includes synchronized multi-view HD video streams of multiple humans doing multiple actions. It consists of 7 individuals performing a set of 26 different body movements.

4.2.6 Multimodal Kinect-IMU dataset

This dataset has been originally collected to investigate transfer learning among ambient sensing and wearable sensing systems. Nevertheless, the dataset may be also used for gesture spotting and continuous activity recognition. It includes data for three activity recognition scenarios, namely HCI (gesture recognition), fitness (continuous recognition) and background (unrelated events). It comprised synchronized 3D coordinates of 15 body joints, measured by a vision-based skeleton tracking system (Microsoft Kinect), and the readings of 5 body-worn inertial measurement units (IMUs. A single subject performs five kinds of geometric gestures with the right hand in alternation 48 times. The locations of the IMUs devices are: the left lower arm, the right lower arm, the back, the left upper arm and the right upper arm.

Table 6 lists a summary of some multi-modal publicly available databases involving accelerometer sensor for human action recognition..

5 Fusion Framework

Although human action recognition promises to be highly effective, the exploitation of multi-level fusion approaches can guarantee an excellent rate thanks to the wealth of the information available in all stages of the human action recognition process: acquisition, feature extraction, classification and decision. Thus, we introduce a fusion framework that utilises accelerometers data.

Fusing data is the process of coupling data acquired from numerous sources (in our case several accelerometers) allowing to assess the accuracy of the system. Indeed, we distinguish two categories of merging: before correspondence and after correspondence. The first category concerns the signal-level fusion and the feature level fusion, and the second category involves fusion at the score level and fusion at the decision level. The four levels of fusion shown in Fig. 5, are presented in the following.

5.1 Signal-Level Fusion

The signal presents the modality acquired on-line or off-line (ex. Speech, Accelerometer signal, Image, Video, etc.). At this level, the fusion is only possible when the data are compatible: the sources produce signals of the same type. In our study, the signal fusion technique includes the combination of 3-axes signals from the accelerometer (X-axis, Y-axis and Z-axis).

Table 6 Summary of multimodal publicly available databases using accelerometer data for human action recognition. N_s: Number of Subjects

Database	N_s	Actions	Sensors	Accelerometers locations	Comments
CMU Database	43	Food preparation Cooking five recipes	Camera/ Microphone Motion capture Accelerometer, etc.	Both forearms, Upper arms, Left/ Right calves, Thighs, Abdomen, Both wrists	– This set is concentrated on a very specific category of actions (cooking)
Opportunity	12	Groom room, Prepare coffee, Clean up, etc.	Accelerometer/ Gyroscope Microphone/ Camera, etc.	Wrist, Chest Limb, Shoulder Foot	– The database covers only morning activities
Berkeley MHAD	11	11 actions : Jumping in place, Jumping jacks, Bending, Punching, etc.	12 cameras/ 2 Kinect 6 accelerome-ters/ motion capture/4 microphones	Wrists Ankles Hips	– Data are temporally synchronized and controlled
UTD-MHAD	8	27 actions : jogging in place, walking in place, sit to stand, etc.	– Kinect Wearable inertial sensor	The right wrist or the right thigh	– The dataset covers a wide range of actions
Huawei/3DLife: Dataset 1	17	22 actions : Throwing, Jumping Jacks Squats, etc.	−5 Kinect -8 inertial sensors	Right/Left wrist Chest, hips, Right/Left ankle, Right/Left foot	– This set focuses on various types of actions
Multimodal Kinect-IMU dataset	1	Five kinds of geometric gestures :Drawing a triangle, etc.	– Kinect -5 inertial sensors	Left /Right lower arm Back Left/ Right upper arm	– Just one person performed the gestures

5.2 *Feature Level Fusion*

Features or attributes are characteristics extracted from the raw data. The feature fusion level is the combination of the different feature vectors, obtained either from the same modality or from different modalities. Therefore, the merging at this level can consider homogeneous feature vectors and heterogeneous feature vectors.

Table 7 Score level fusion rules. T is the number of matchers and s_j presents the normalized scores of the jth matcher. w_j corresponds to the Equal error rate (EER) of the jth and F represents the fusion score

Rule	Equations
Sum	$F = \dfrac{1}{T} \sum\limits_{j=1}^{T} s_j$
Max	$F = \max(s_1, s_2 ..., s_T)$
Min	$F = \min(s_1, s_2 ..., s_T)$
Product	$F = \prod\limits_{j=1}^{T} s_j$
Weighting	$F = \sum\limits_{j=1}^{T} w_j s_j$

5.3 Score Level Fusion

A score is a measure of similarity that corresponds to the distance between the test sample and the reference sample. In fact, the fusion at this level presents a compromise between the richness of the information and the facility of the implementation. Actually, each classifier produces a matching score or several scores and the merging process combines these measures to obtain the final score which will be then used to produce the final decision. There exist two main approaches to combine scores: the classification of scores and the combination of scores. Several rules exploited to ensure the fusion of scores are presented in Table 7.

5.4 Decision Level Fusion

It processes the outputs of the different classifiers. The decision level fusion consists in assembling the decisions obtained from each classifier in order to obtain the final decision. There are several methods for merging decisions, such as the AND and the OR logic operator type rules, the majority vote, and the Dempster–Shafer theory.

Thus, we focused our attentiveness on characterizing the human actions in order to gain a better classification accuracy by employing a fusion framework exploiting every information obtained along the human action recognition process.

For the feature extraction, we opted for the discrete wavelet transform. The Wavelet transform of a function $f(x)$ is calculated using (1) as follows:

$$W_f(i, \tau) = \int_{-\infty}^{+\infty} f(x) \psi_{i,\tau}^{\star}(x) dx \tag{1}$$

$$\psi_{i,\tau}(x) = \frac{1}{2^i} \psi \left(\frac{x - \tau}{i} \right) \tag{2}$$

ψ is the wavelet mother.

The wavelet transform decomposes initially raw data into approximation coefficients by employing low pass filter and detail coefficients by a high-pass filter. Various levels are constructed as follows: the approximation signal required from the previous level is decomposed into approximation and detail coefficients. The desired decomposition level is determined after several repetitions of this process.

With regard to the classification, we exploit support vector machine SVM.

The experimentations associated to the fusion framework are presented in the following section.

6 Experimental Results and Analysis

6.1 Results

To evaluate the effectiveness of our methodology, we chose three databases: WARD, MHAD and Realdisp which were detailed in Sect. 5. The position of each accelerometer related to each dataset is presented in Fig. 6.

In the interest of the fusion framework, we aimed firstly to select the sensors that guarantee better classification rates for each dataset. In fact, these sensors will be afterwards exploited in the fusion approach in order to provide a higher performance.

Therefore, we proposed to evaluate each accelerometer individually. The collected data from the accelerometer sensor were firstly divided into N temporal windows using the sliding window technique. The window length related to each database is 6 for WARD, 15 for MHAD and 9 for Realdisp. The number of segments N was
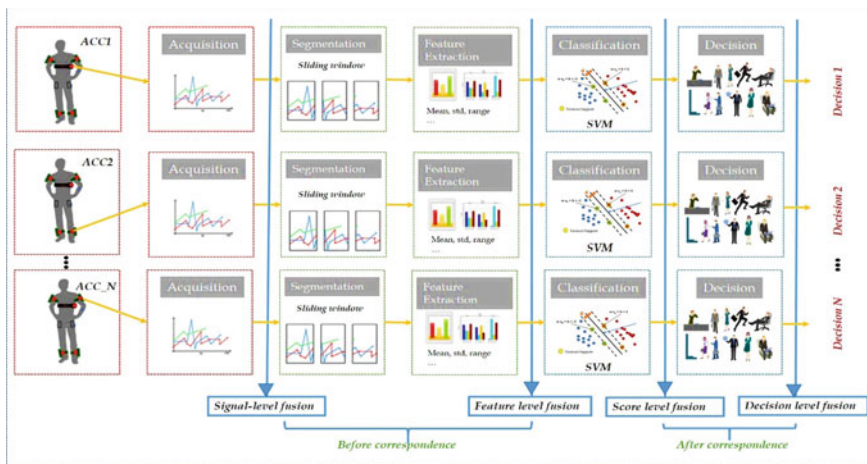
Fig. 5 An example of multi-level fusion using accelerometer data
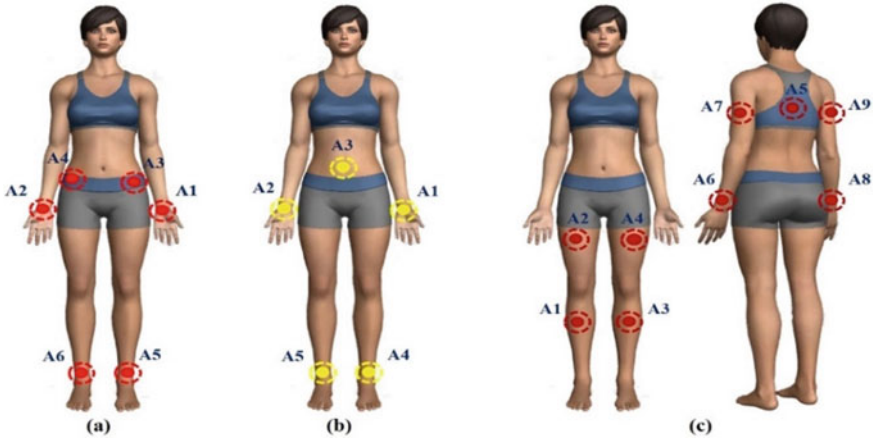
Fig. 6 Positions of the accelerometers for the three databases. **a** MHAD. **b** WARD. **c** Realdisp

Table 8 Recognition rates (%) for all accelerometers for the 3 databases

Accelerometer	Recognition rate (%)		
	WARD	MHAD	Realdisp
A_1	64	88	41
A_2	64	86	49
A_3	54	36	30
A_4	72	68	50
A_5	71	45	41
A_6	–	50	51
A_7	–	–	51
A_8	–	–	43
A_9	–	–	48

determined based on several experimentations. Then, the features were extracted from each window using the discrete wavelet transform with a Daubechies2 as a wavelet mother. From the approximation coefficients, we extracted the mean and the standard deviation and concerning the detail coefficients, we extracted the minimum and the root mean square. These measures are computed over the three directions (X, Y and Z) within each temporal window. Furthermore, we took advantage of the SVM with RBF kernel to classify the actions. We considered 12 subjects for training and 8 subjects for the test. Concerning MHAD, 7 subjects were reserved for training and 5 persons for testing. Finally, for Realdisp 10 subjects were provided for the learning base and 7 persons were preserved for testing.

Thus, the results relative to each accelerometer of the three datasets are presented in Table 8.

Table 9 Recognition rates (%) for all levels of fusion

Database	Fusion level						
	Signal	Feature	Score			Decision	
			Sum	Max	Product	AND	Vote
WARD	75	88	88	82	88	72	78
MHAD	89	94	94	91	97	86	89
Realdisp	53	76	77	67	78	51	63

According to the results shown in Table 8, we can observe that for the WARD database the accelerometers attached to the ankles (A_4 & A_5) provide a better performance in terms of accuracy. Indeed, the actions introduced in this dataset are essentially linked with the motions of the feet such as "walking", "going up and down the stairs", "Jumping", etc. Accordingly, the sensors fixed on the ankles ensure better recognition rates compared to the other sensors.

Form the results obtained from the six accelerometers considered on the MHAD dataset, we notice that the accelerometers mounted on the left and the right wrist ensure a better classification rates (A_1 & A_2). Based on the type of the actions considered in this dataset, which are related to the hand motions (e.g. "clapping", "waving", "punching"), the accelerometers worn on the hands can classify correctly the classes. With regard to the accelerometers attached to the ankles, they are not able to generate useful information because the actions are relatively static.

Regarding the Realdisp dataset, the accelerometers A_4, A_6 and A_7 attached respectively to the right thigh, the left lower arm and the left upper arm, seem to be the most effective to distinguish the human actions introduced in this dataset. In fact, the actions employed focus on the trunk, upper and lower extremities including actions of translation, jumping and physical activities. Therefore, a part or all of the body is moving during the performance of the actions; otherwise, the recognition rates relative to the 9 accelerometers distributed at different positions are convergent.

After evaluating each sensor separately and with a view to obtain a higher recognition rate and improve the classification, we proposed to employ the multi-level fusion techniques. We fused the signals acquired from the 3 axis acceleration data and we combined the features from the time-frequency domain from each chosen sensor. For the score level, the Sum, the Max and the Product rules were exploited. Finally, for the decision level, the rules AND and OR were employed.

In this step, we suggest involving the 3 accelerometers that guarantee the best performance for each dataset based on the positions of the sensors and the results obtained. For WARD database accelerometer number 1, accelerometer number 4 and accelerometer number 5 are chosen. In addition, A_1, A_2 and A_4 are part of the stage of fusion for MHAD. And for Realdisp, only A_4, A_6 and A_7 are involved. Thus, the results are presented in Table 9.

6.2 Discussion

We compared the recognition rates of the multi-level fusion framework with the accuracies obtained from each accelerometer individually, we noticed that combining signals, features, scores or decisions guarantees a higher performance. In fact, our approach exploits every information available in the recognition process from acquisition to decision and leads to good results for the employed datasets as we listed in Table 9.

From Table 9, we outline that the matching score level fusion outperformed the other levels of fusion and achieved favorable performance compared with the utilization of each sensor individually. Actually, compared with the other levels of coupling, this level provides richer information as it fuses the distances between the test samples and the reference samples.

Moreover, the classification accuracy using fusion scores is higher than the performance found in the literature for MHAD database which is 97% against 94% in Chen et al. (2015).

This improvement leads to a discrimination between most of the actions as we can see from Figs. 8 and 9 which represent the confusion matrices when fusing scores related respectively to MHAD database and WARD database

In the intention to evaluate the effectiveness of our method, we compare the confusion matrix related to the use of each accelerometer individually with the matrix obtained from the coupling of scores. Thus, we consider MHAD database as an example, Fig. 7a, b and c correspond respectively to the confusion matrix of A_1, A_2 and A_4.

As seen in Fig. 7a, the accelerometer A_1 worn on the left wrist provides a good discrimination between the actions as the accomplishment of most of the actions requires the contribution of the left hand differently (waving, punching, throwing, etc.). However, it can't differentiate action 4 "Boxing" from action 7 "Clapping" owing to the similarity of the behavior of arms. In addition, the misclassification that occurs between action 8 "Throwing" and action 11 "Standing" can be explained by the fact that the posture of the left hand is the same in these actions so the accelerometer generates similar raw data.

From Fig. 7b, we notice that the distinction between action 5 and action 6 is difficult using the accelerometer A_2 fixed on the right, besides, the action 6 "waving using the right hand" can be considered as a subset of the action 5 "waving using both hands".

Figure 7c shows the confusion matrix when using the accelerometer A_4 mounted on the right hip, we notice that the recognition of classes: 9, 10 and 11 is improved by this sensor because of its contribution to the accomplishment of these tasks: To stand up then sit, Sit and Stand up. However, the system thus is unable to distinguish between the other classes.

As seen in Fig. 8, combing the scores acquired from these sensors leads to a discrimination between most of the actions, nonetheless, there remain some slight
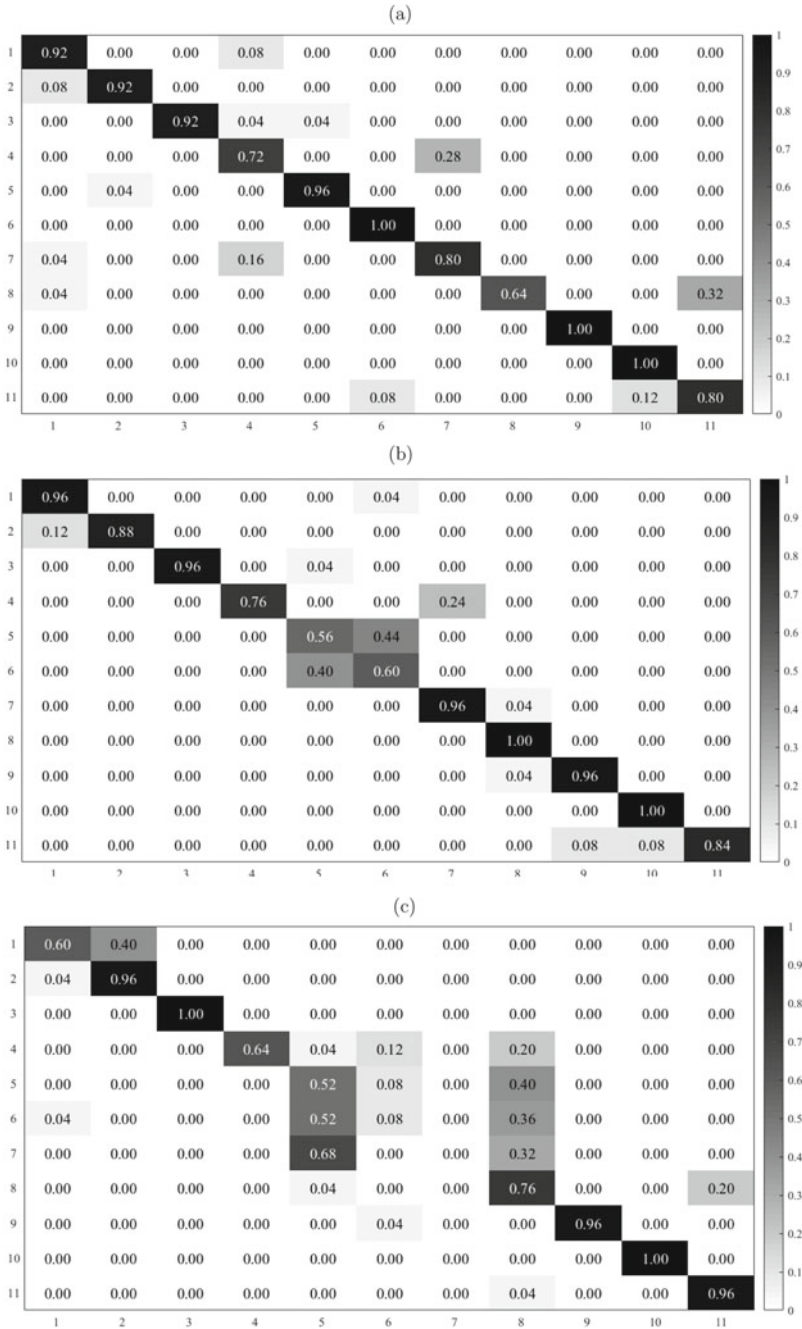
Fig. 7 Confusion Matrix related to MHAD **a** when using A_1 **b** when using A_2 **c** when using A_4 (Actions: 1. Jump, 2. Jack, 3. Bend, 4. Punch, 5. Wave 2 hands, 6. Wave using the right hand, 7. Clap 8. Throw 9. Sit+Stand, 10. Sit, 11. Stand)
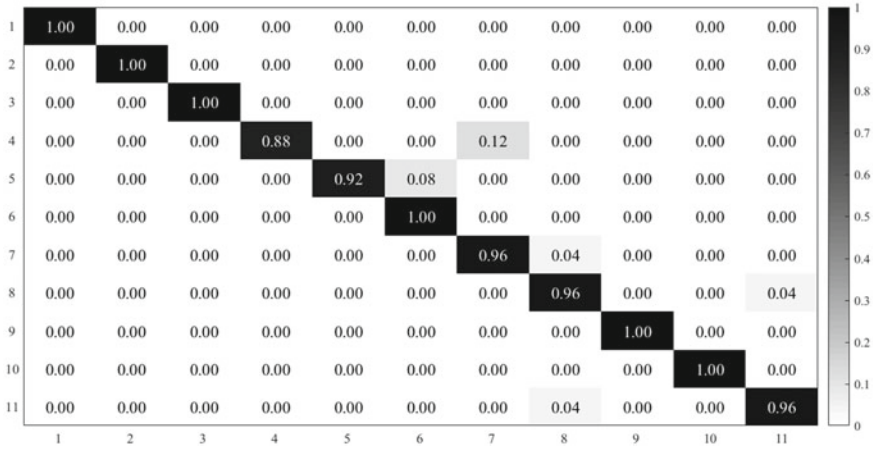
	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.88	0.00	0.00	0.12	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.92	0.08	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.04	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.04
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.96

Fig. 8 Confusion matrix of MHAD database related to fusion scores for A_1 & A_2 & A_4 (rule: Product)

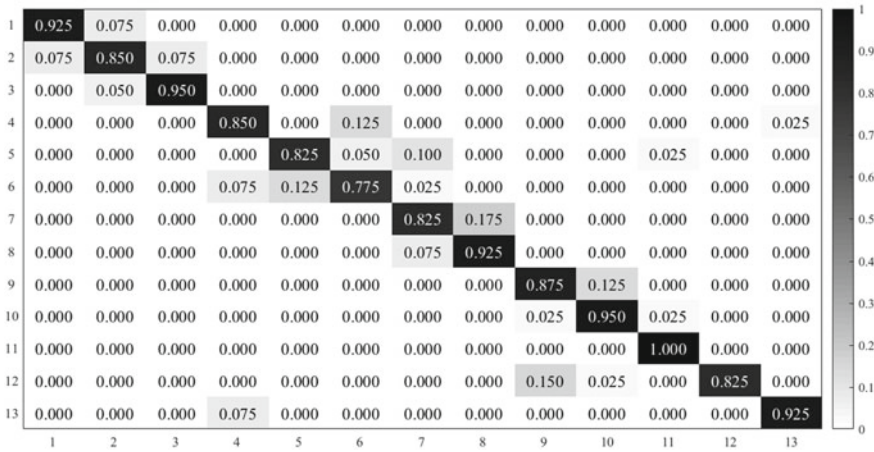|	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.925	0.075	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.075	0.850	0.075	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.050	0.950	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.850	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.025
5	0.000	0.000	0.000	0.000	0.825	0.050	0.100	0.000	0.000	0.000	0.025	0.000	0.000
6	0.000	0.000	0.000	0.075	0.125	0.775	0.025	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.825	0.175	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.075	0.925	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.875	0.125	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.950	0.025	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.150	0.025	0.000	0.825	0.000
13	0.000	0.000	0.000	0.075	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.925

Fig. 9 Confusion matrix of WARD database related to fusion scores for A_1 & A_4 & A_5 (rule: Product)

misclassifications between the action "punching" and "clapping" because of the similarity of hand movements.

As regards to WARD database, the misclassification occurs between the most similar actions as "walk forward", "walk right" or "walk left" as shown in Fig. 9. Indeed, the walking speed differs from one person to another so the differentiation of these actions is a challenging task.

Finally, the classification accuracies of our method are encouraging as it decreases the number of misclassifications and provides important recognition rates.

7 Conclusions

In this chapter, an overview of different methodologies for human action recognition using accelerometer data have been introduced. After we covered diverse sensors used to recognize human actions, we proved that the accelerometer seems to be the most efficient thanks to its benefits in this area of research. Furthermore, various applications related to human action recognition in many areas were outlined, and different approaches existing in the literature were reviewed. Moreover, we reported some publicly available datasets from human action recognition where the accelerometer data was provided. Afterward, a multi-level fusion framework was introduced using acceleration data from the most efficient accelerometers for each dataset used to evaluate this work. The multi-level fusion framework included a signal level, a feature level, a score level and a decision level. According to the results, the recognition rates were improved however; there remains some slight misclassification between the most similar classes.

References

Aggarwal, J., & Ryoo, M. (2011). Human activity analysis: A review. *ACM Computing Surveys*, *43*(3), 1–47.

Allouch, A., Koubaa, A., Abbes, T., & Ammar, A. (2017). RoadSense: Smartphone application to estimate road conditions using accelerometer and gyroscope. *IEEE Sensors Journal*, *17*(13), 4231–4238.

Ameur, S., Ben Khalifa, A., & Bouhlel, M.S. (2016). A comprehensive leap motion database for hand gesture recognition. In *7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications* (pp. 514–519).

Beyrouthy, T., Kork, S. K. A., Korbane, J. A., & Abdulmonem, A. (2016). EEG mind controlled smart prosthetic arm. In *IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies, EmergiTech* (pp. 404–409).

Castignani, G., Derrmann, T., Frank, R., & Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. *IEEE Intelligent Transportation Systems Magazine*, *7*(1), 91–102.

Chebli, K., & Ben Khalifa, A. (2018). Pedestrian detection based on background compensation with block-matching algorithm. In *15th International Multi-Conference on Systems, Signals & Devices (SSD'18)* (pp. 473–477).

Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). Improving human action recognition using fusion of depth camera and inertial sensors. *IEEE Transactions on Human-Machine Systems*, *45*(1), 51–61.

Chen, Y., & Shen, C. (2017). Performance analysis of smartphone-sensor behavior for human activity recognition. *IEEE Access*, *5*, 3095–3110.

Cornacchia, M., Ozcan, K., Zheng, Y., & Velipasalar, S. (2017). A survey on activity detection and classification using wearable sensors. *IEEE Sensors Journal*, *17*(2), 386–403.

Dargar, S. Kennedy, R., Lai, W., Arikatla, V., & De, S. (2015). Towards immersive virtual reality, (iVR): A route to surgical expertise. *Journal of Computational Surgery*, *2*(2), 1–26.

Erdaş, Ç., Atasoy, I., Açıcı, K., & Oğu, H. (2016). Integrating features for accelerometer-based activity recognition. *Procedia Computer Science*, *98*, 522–527.

Figueiredo, I. N., Leal, C., Bolito, L. P., & Lemos, A. (2016). Exploring smartphone sensors for fall detection. *MUX: The Journal of Mobile User Experience*, *5*(2), 1–17.

Garcia-Ceja, E., Osmani, V., & Mayora, O. (2016). Automatic stress detection in working environments from smartphones' accelerometer data: A first step. *IEEE Journal of Biomedical and Health Informatics*, *20*(4), 1053–1060.

Hidayat, A. A., Wasista, S., & Pratiwi, Y. P. (2016). Development of fighting genre game(boxing) using an accelerometer sensor. In*International Conference on Knowledge Creation and Intelligent Computing* (pp. 201–206).

Höflinger, F. Rui, Z., Tobias, V., Enrique, G., Adnan, Y., Christina, S., Kerstin, K., & Leonhard, M. (2015). Motion capture sensor to monitor movement patterns in animal models of disease. In *IEEE 6th Latin American Symposium on Circuits & Systems* (pp. 1–4).

Hung, C.-H., Bai, Y.-W., & Wu, H.-Y. (2015). Home appliance control by a hand gesture recognition belt in LED array lamp case. In *IEEE 4th Global Conference on Consumer Electronics, GCCE* (pp. 599–600).

Jain, A., & Kanhangad, V. (2018). Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sensors Journal*, *18*(3), 1169–1177.

Jegham, I., & Ben Khalifa, A. (2017). Pedestrian detection in poor weather conditions using moving camera. In *IEEE/ACS 14th International Conference on Computer Systems and Applications, AICCSA* (pp. 358–362).

Jegham, I., Ben Khalifa, A., Alouani, I., & Mahjoub, M. A. (2019). MDAD: A multimodal and multiview in-vehicle driver action dataset. In M. Vento & G. Percannella (Eds.), *Computer analysis of images and patterns, CAIP, 2019* (Vol. 11679, pp. 518–529). Lecture notes in computer science. Cham: Springer.

Kalantarian, H., Alshurafa, N., & Sarrafzadeh, M. (2016). Detection of gestures associated with medication adherence using smartwatch-based inertial sensors. *IEEE Sensors Journal*, *16*(4), 1054–1061.

Kau, L.-J., & Chen, C.-S. (2015). A smart phone-based pocket fall accident detection, positioning, and rescue system. *IEEE Journal of Biomedical and Health Informatics*, *19*(1), 44–56.

Koskimäki, H., & Siirtola, P. (2014). Recognizing gym exercises using acceleration data from wearable sensors. In *IEEE Symposium on Computational Intelligence and Data Mining* (pp. 321–328).

Kowalczuk, Z., & Merta, T. (2015). Evaluating the position of a mobile robot using accelerometer data. In *Advanced and intelligent computations in diagnosis and control* (pp. 131–143).

Kurashima, S., & Suzuki, S. (2015). Improvement of activity recognition for child growth monitoring system at kindergarten. In *41st Annual Conference of the IEEE Industrial Electronics Society* (pp. 2596–2601).

Kwon, B., Junghwan, K., Kyoungoh, L., Yang, K., Sangjoon, P., & Sanghoon, L. (2017). Implementation of a virtual training simulator based on 360° multi-view human action recognition. *IEEE Access*, *5*, 12496–12511.

Kyberd, P., & Poulton, A. (2017). Use of accelerometers in the control of practical prosthetic arms. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(10), 1884-1891.

Laghari, A., Waheed-ur-Rehman, & Memon, Z. A. (2016). Biometric authentication technique using smartphone sensor. In *13th International Bhurban Conference on Applied Sciences and Technology* (pp. 381–384).

Lee, S.-M., Yoon, S. M., & Cho, H. (2017). Human activity recognition from accelerometer data using convolutional neural network. In *IEEE International Conference on Big Data and Smart Computing* (pp. 131–134).

Lejmi, W., Ben Khalifa, A., & Mahjoub, M. A. (2017). Fusion strategies for recognition of violence actions. In *IEEE/ACS 14th International Conference on Computer Systems and Applications, Hammamet* (pp. 178–183).

Lejmi, W., Ben Khalifa, A., & Mahjoub, M. A. (2019). Challenges and methods of violence detection in surveillance video: A survey. In M. Vento, G. Percannella (Eds.), *Computer analysis of images*

and patterns. CAIP, 2019 (Vol. 11679, pp. 62–73). Lecture notes in computer science. Cham: Springer.

López, J. D., Sucerquia, A., Duque-Muñoz, L., & Vargas-Bonilla, F. (2015). Walk and jog characterization using a triaxial accelerometer. In *IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing* (pp. 1406–1410).

Lubina, P., & Rudzki, M. (2015). Artificial neural networks in accelerometer-based human activity recognition. In *22nd International Conference Mixed Design of Integrated Circuits & Systems* (pp. 63–68).

Luštrek, M., Mitja, L., Cvetkovi'c, B., Mirchevska, V., Stefan, J., Kafali, Ö., Romero, A. E., & Stathis, K. (2015). Recognising lifestyle activities of diabetic patients with a smartphone. In *9th International Conference on Pervasive Computing Technologies for Healthcare* (pp. 317–324).

Malawski, F., & Gałka, J. (2018). System for multimodal data acquisition for human action recognition. *Multimedia Tools and Applications*, *77*, 23825–23850.

Mimouna, A., Ben Khalifa, A., & BenAmara, N. E. (2018). Human action recognition using accelerometer data: Selective approach. In *International Multi-Conference on Systems, Signals & Devices* (pp. 467–472).

Namal, S., Senanayake, A., Vincent, C., James, C., & Rolland Sirisinghe, G. (2006). Analysis of soccer actions using wireless accelerometers. In *IEEE International Conference on Industrial Informatics, Singapore* (pp. 664–669).

Neto, P., Pires, J. N., & Moreira, A. P. (2009). Accelerometer-based control of an industrial robotic arm. In *The 18th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 1192–1197).

Noor, M. H. M., Salcic, Z., & Wang, K. I.-K. (2015). Dynamic sliding window method for physical activity recognition using a single tri-axial accelerometer. In *IEEE 10th Conference on Industrial Electronics and Applications, ICIEA* (pp. 102–107).

Nuno, M., João, F., João, V., Mohammad, S., & Pedro, N. (2017). Human behavior and hand gesture classification for smart human-robot interaction. *Procedia Manufacturing*, *11*, 91–98.

Pepa, L., Verdini, F., Capecci, M., & Ceravolo, M. G. (2015). Smartphone based freezing of gait detection for Parkinsonian patients. In *IEEE International Conference on Consumer Electronics* (pp. 212–215).

Rastegari, A., Archenti, A,. & Mobin, M. (2017). Condition based maintenance of machine tools: Vibration monitoring of spindle units. In *Annual Reliability and Maintainability Symposium* (pp. 1–6).

Rodomagoulakis, I., Kardaris, N., Pitsikalis, V., Mavroudi, E., Katsamanis, A., Tsiami, A., & Maragos, P. (2016). Multimodal human action recognition in assistive human-robot interaction. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 2702–2706).

Safi, K., Mohammed, S., Attal, F., Dedabrishvili, M., & Amirat, Y. (2016). Recognition of different daily living activities using hidden Markov model regression. In *3rd Middle East Conference on Biomedical Engineering* (pp. 16–19).

Sen, S., Vigneshwaran, S., Archan, M., Rajesh, K.B., & Youngki, L. (2015). The case for smartwatch-based diet monitoring. In *IEEE International Conference on Pervasive Computing and Communication Workshops* (pp. 585–590).

Tran, D. N., & Phan, D. D. (2016). Human activities recognition in android smartphone using support vector machine. In *7th International Conference on Intelligent Systems, Modelling and Simulation* (pp. 64–68).

Unuma, Y., & Komuro, T. (2015). Natural 3D interaction using a see-through mobile AR system. In *IEEE International Symposium on Mixed and Augmented Reality* (pp. 84–87).

Wang, Z., Donghui, W., Jianming, C., Ahmed, G., & Mohammad, A. H. (2016). A triaxial accelerometer-based human activity recognition via EEMD-based features and game-theory-based feature selection. *IEEE Sensors Journal*, *16*(9), 3198–3207.

Yunyoung Nam, Y. K. J. L. (2016). Sleep monitoring based on a tri-axial accelerometer and a pressure sensor. *Sensors*, *16*(5), 1–14.

Zhang, Y., Fei, Y., Xu, L., & Sun, G. (2015). Micro-IMU-based motion tracking system for virtual training. In *34th Chinese Control Conference* (pp. 7753–7758).

Zhang, L., Ou, M., Fu, X., & Yan, X. (2016). Using smartphones to estimate vehicle emission under urban traffic levels-of-service. In *12th World Congress on Intelligent Control and Automation* (pp. 1758–1763).

Zhao, X., Zhiming, G., Tao, F., Shishir, Shah, & Weidong, Shi. (2014). Continuous fine-grained arm action recognition using motion spectrum mixture models. *Electronics Letters*, *50*(22), 1633–1635.

Zia, A., Yachna, S., Vinay, B., Eric, L. S., & Irfan, E. (2018). Video and accelerometer-based motion analysis for automated surgical skills assessment. *International Journal of Computer Assisted Radiology and Surgery*, *13*(4), 443–455.