IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 5, 2017, accepted November 12, 2017, date of publication November 21, 2017, date of current version March 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2776098

Who Used My Smart Object? A Flexible Approach for the Recognition of Users

HAMDI AMROUN AND MEHDI AMMI
Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, Centre National de la Recherche Scientifique,
University of Paris-Sud, 91400 Orsay, France

Corresponding author: Hamdi Amroun (hamdi.amroun@limsi.fr)

ABSTRACT This paper deals with the authentication of the user of a connected object. We propose a flexible and nonintrusive method based on the use of two categories of everyday connected objects (i.e., smart watch and remote control). Data were collected during participants' interactions with a smart TV. The discrete cosine transform algorithm was used to extract the most informative features. Based on these features, four classification algorithms (deep neural network, support vector machine, Naïve Bayes classifier, and C45) were applied to the data in order to detect the user's identity. The classification was performed based on the recognition of four types of human activities (sitting, standing, walking, and lying down) through building four databases. Following this, a second classification was made for each data set activity type in order to identify the users. The results show that it is possible to discriminate between users according to their activities. The accuracy of recognition reached 91% for some participants within a certain activity configuration.

INDEX TERMS IoT, activity recognition, automatic classification, unconstrained environment.

I. INTRODUCTION

The use of connected objects has become more widespread in recent years, and many new uses have emerged, such as using a connected cup for hydration, for monitoring patients in everyday life, or for predicting the risk of falls in the elderly [1]–[4].

However, the benefits of all these services are subject to a major problem, that is, the ability to distinguish between the different users of the same connected object. This can be useful for instance in order to associate recorded data with the right user or, more essentially, to identify the real owner of the connected object. For example, a connected bottle can indicate whether or not a given user has drunk water. Similarly, a smart phone could recognize its real owner.

Today, there are several methods for detecting whether or not the user of a connected object is its owner, such as the use of passwords or finger print sensors to log on to the device. However, the use of passwords is unsuitable and inefficient, particularly for certain categories of users such as people with Alzheimer's who risk forgetting their passwords, or even involuntarily disclosing them to third parties [5]. In addition, the use of fingerprint sensors may not work, for example, due to hardware or software failures. This paper investigates an alternative to these methods. We propose a flexible (in term of used devices which does not present any constraint of port nor of use. Also in terms of method used and the environment of experimentation) and robust method for recognizing the users of connected objects based on an analysis of their physical activities.

In the current study, we use two commonly connected objects: an Apple TV remote and an Apple Watch. In practice, the user of the Apple TV or Apple Watch may encounter several problems with user recognition, for instance the risk of purchasing applications or paying subscriptions (music, movies, TV programs, video games etc.); there is therefore a need to detect whether or not the user is the real owner.

The approach adopted here is based on the recognition of four types of human activities: standing, sitting, walking and lying down. For each of these activities, each user is classified according to his/her behavior with respect to the use of the Apple TV remote and Apple watch, in an uncontrolled environment (i.e. in everyday life).

Two scenarios are studied here: the recognition of users of a single device and of two devices. The objective is to investigate the gain in the classification accuracy of users for the different configurations.

The core of our approach is to identify users of one or more connected objects using two levels of classifications. Users activities are classified through building databases for four activities: sitting, standing, walking and laying down. Then, for each database, a second classification is carried out to identify users in relation to their activities. Both classifications are made using a DNN algorithm. A comparative study with other algorithms (SVM, NB and C45) is carried out at the end of this paper.

The following section describes the present state-of-the-art methods used to detect the different users of a connected object and the various methods used to perform the classification of physical activities.

II. RELATED WORK

The methodology we propose for detecting users is subdivided into two phases of recognition: the recognition of human activity and the recognition of users themselves. There is there fore a need to deal with these two issues.

A. RECOGNITION OF HUMAN ACTIVITY

In order to be able to trace back an activity to a user, many studies of there cognition of human activity have subdivided the process of activity recognition into three stages, as follows:

1. Data collection through developing specific platforms.
2. Extraction, selection (and possibly merging) of the best features.
3. Classification using machine learning algorithms.

Research work on the recognition of activity focuses on Steps 2 and 3. Step 2 tries to find the best descriptors or attributes to provide as input to the automatic learning algorithms and step (3) to increase the classification accuracy.

Very few studies have focused on finding the best descriptors to extract for the learning process using IoT. For example, Da Silva and Galeazzo [25] developed a system for recognizing activity using a smart watch; the author extracted nineteen features using two techniques: the Fisher discriminant ratio (FDR) and principal component analysis (PCA). He used SVM to classify these features and obtained a performance of nearly 93% accuracy. He and Jin [26] how have focused on the other type of features such as autoregressive feature. He presented an autoregressive process to recognize human activity. These features were extracted for classification using SVM and obtained a classification performance of 92.25% accuracy.

Other kinds of descriptors have been extracted, such as the fast fourier transform (FFT) or discrete cosine transform (DCT), which was used by He and Jin [27] who developed an activity recognition model based on a single accelerometer. The authors of this study selected the DCT as a feature to extract from input signals, and used SVM for classification, reaching an accuracy of 97.51%.

A great deal of prior work exists on extracting basic features, such as that in [28], in which the authors extracted the mean, variance and standard deviation using different window sizes for signal slicing. They used two classification algorithms, the multi layer perceptron (MLP) and k-nearest neighbors (KNN), and obtained accuracies of 89.6% and 92.89%, respectively.

With the development of machine learning techniques, new approaches have emerged, such as deep learning models, and more specifically, those that have been used in other work such as the model developed by Zebin *et al.* [12] for the recognition of human activity, which was based on a somewhat complex architecture of a convolution neural network (CNN). This architecture learns features extracted automatically from the signal flow of the accelerometer and the gyroscope. The authors also tested the influence of the number of convolution layers and the size of the convolution kernel on classification performance, and compared the precision of their model with conventional models such as SVM and MLP. CNN achieved a performance of 97.1% in contrast to SVM (96.4%) and MLP (91.7%).

New research has also been proposed in [29] using a new CNN architecture called CNNs; this is composed of CNN-pf and CNN-pff. CNN-pf is a CNN with partial weight sharing in the first convolution layer and a total weight sharing in the second convolution layer; CNN-pff represents a CNN with both total and partial weight-sharing in the first and second convolution layers. This model was used to learn the multi modal characteristics of the input data stream. This model has been tested on public data sets for the classical models HMM, SVM, HCFR, 1DCNN, 2DCNN; the authors have shown that their model has an accuracy of between 91.24% and 99%.

Other studies have been proposed for completing the process of recognition of an activity by making changes at the CNN core level, as has been done by Chen and Xue [30] who developed an approach for recognizing human activities using a single accelerometer. They proposed an architecture based on a modification of the convolution kernel in order to adapt the characteristics of the accelerometer signals, and compared their results with existing models such as SVM with descriptor FFT, DCT and FT. These authors reported an accuracy of 9 3.8%.

DNNs have begun to emerge, and are being used to recognize human activity using smart objects; for example, Zeng *et al.* [31] have developed a new model that can automatically extract features from the sensor signals of a smart phone. This method, which is based on the CNN model, can capture local dependencies and in variances in the signals, as has been demonstrated in speech recognition or image processing. These authors obtained an accuracy of more than 96%.

B. USER RECOGNITION

Numerous studies have assessed the traceability of data of devices using applications for the collection of this data in order to determine whether a user is indeed the real owner of the device. For example, Datta and Manousakis [13] developed an application to record several types of information from the sensors of a smart phone and to determine whether the user is the owner of the device. They applied the SVM algorithm to classify users into two classes; this study used thirty-four participants and obtained a classification accuracy of about 85%, with false positives not binding to smart phone security.

Another study has addressed the problem of traceability from another aspect, in order to establish a profile of several users by studying user authentication; this research developed temporal and spatial time models to establish a probability model and create a threshold for detection [14].

Some researchers have tried to classify users using accelerometers, gyroscopes and magneto meters [15], and applying the SVM algorithm to classify their data. A classification accuracy of about 83% has been reported.

Various approaches have been developed in order to detect and classify users and to recognize the traceability of device data, based on continuous and passive authentication. These approaches classify users' identities according to their tactile movements.

Wu *et al.* [16] propose a method for profiling users based on tactile gestures and movements, otherwise known as behavioral biometric profiling. They showed that the way in which a user interacts with the touch screen reflects their unique physical and behavioral biometrics, i.e. up and down movements of a finger and the finger's pressure on the screen).

Authentication methods have also been developed in this context, such as continuous authentication methods for preventing data loss and the leakage of Android smart objects. Biometric authentication refers to the use of human features in order to label and identify users.

Recently, a significant number of research studies of the tactile dynamics of fingers and tactile striking on smart phones have begun to emerge; these have been applied to the identification of smart phone users. The phrase 'touch dynamics of keystrokes' refers to a collection of detailed information about touch screens, such as the touch or keystroke time, movement, scrolling, blinking and rotation of touch screens. The movement and dynamics of touch are used in the continuous authentication of smart phones [17].

These techniques (tactile dynamics and touch dynamics) have been widely applied to smart phones equipped with physical keys. As an example, Antal and Szabó [18] examined the performance of a touch screen based on keystroke dynamics. They interviewed forty-two participants in order to collect data on touch screen smart phones, and used SVM methods, naïve Bayes classifier and a random forest algorithm to classify users. They showed that these features significantly improved the accuracy of both processes.

From the point of view of software, many studies have developed applications to be used in the context of user detection. For example, an application was used in [19] to extract functionality from data in order to deduce the keywords that were typed on to a screen. A total of 70% of the strokes were correctly predicted.

Another method studied the biometrics of keystrokes on the touch screen of a smart phone [20]; these authors analyzed 20, 160 entries of the passwords of 28 participants. The authentication error rate was found to be between 26% and 36%.

Frank *et al.* [21] analyzed a set of 30 tactile features extracted from touch screen behaviors, and trained user profiles based on vertical and horizontal lines using the k-nearest neighbor and SVM algorithms. The result was very satisfactory, with an authentication error of between 0% and 4%.

An analysis of user behavior has also been developed in order to be able to classify users. Many studies have examined the micro-behaviors of smart phone users [22]; these micro-behaviors can identify users with an accuracy of up to 68%. In addition, the detection rate after 15 presses was 86% accuracy. Xu *et al.* [23] studied authentication using a new mechanism of continuous and passive authentication based on pinching, screen typing and handwriting, while Li *et al.* [24] have proposed a biometric model for smart phones in order to re-authenticate the current identity of users according to their finger movements without raising the finger from the screen. These authors used eight features from 75 participants; the results indicate that this model can reach an error rate of less than 4% accuracy.

All of the above studies were carried out in a controlled environment, mostly using smart phones and basic classification algorithms. In addition, none of these user identification methods was trained using pre-processed data. Indeed, no feature calculation has been attempted when the algorithm used for the detection of users was the DNN.

III. OBJECTIVES AND METHODOLOGY

The present paper proposes a more effective (in terms of method used, devices used and even in terms of the environment of the experiment) system for recognition of users of IoT devices.

Based on commonly connected objects with an embedded series of sensors, we develop a platform for the automatic collection and management of data.

We use an Apple framework which includes an Apple watch and an Apple TV remote. These each include an accelerometer, a gyroscope and a microphone. These two devices are generally used at home, and can provide an efficient approach for recognizing users, for instance, for providing an additional authentication tool for purchasing applications or paying subscriptions(music, movies, TV programs, video games, etc.).

Based on this platform, we address a series of issues in order to design an efficient processing pipeline for the recognition of users of an Apple TV remote control and an Apple watch, based on the recognition of their activities (walking, sitting, standing and lying down) using a DNN algorithm.

Initially, the study was carried out in a non-controlled environment. Participants were observed during daily activities without instructions; this configuration was relatively realistic in recreating daily life.

Following this, the DCT method was used to extract features from various data sources to classify them using the DNN algorithm.

Next, the data were processed using two levels of classification. The first level consisted of the automatic detection of participants' activities. This classification was performed using the DNN algorithm. The second level consisted of the classification of participants within each of the four activity categories, also using the DNN.

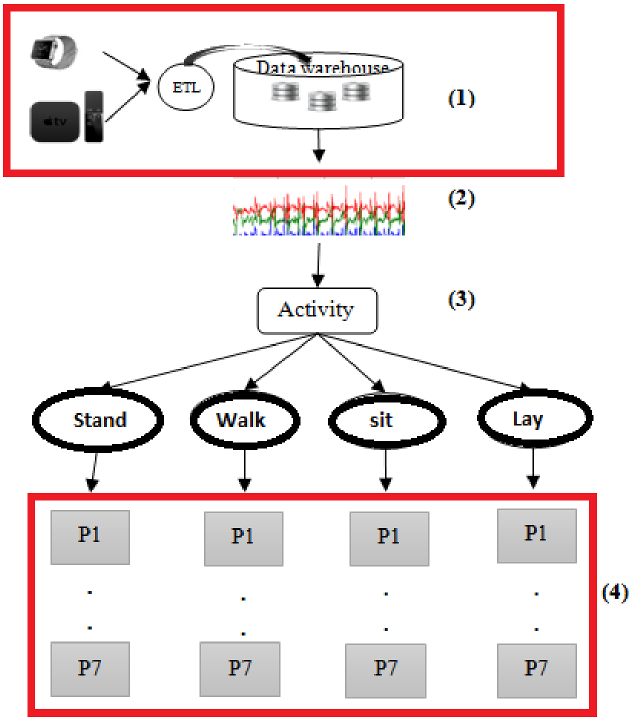

FIGURE 1. User detection steps based on activity recognition: (1) data collection; (2) data preparation; (3) activity recognition; (4) user classification.

Finally, a comparative study was conducted in order to assess the accuracy of the DNN used relative to three other basic algorithms: SVM, NB classifier and C45. Our approach is summarized in four steps, as shown in Fig.1, which presents the method of data collection, the environment of the experiment and the main adjustments made.

1. We extracted descriptors of the various signals, using DCT to concatenate and classify these based on the DNN model.
2. The activities were then classified into four classes: standing, sitting, lying down and walking.
3. For each type of activity, the behavior of the seven participants with respect to the Apple TV remote and Apple watch was studied during each activity. This made it possible to identify differences in behavior between the seven participants during the same activity. In addition, a behavioral profile for the users based on these four activities can be built up (see Fig.4).

The remainder of this paper is organized as follows. Section IV presents the user detection method developed in this paper, the DCT extraction and computation, the adjustments made and our approach towards classification of activity based on the DNN model. Section V describes the main results, and finally,the conclusions of the paper are presented.

IV. PARTICIPANT IDENTIFICATION

The identification of users was carried out in four stages, as shown in the diagram in Fig. 1:

The four steps shown in Fig. 1 are described in detail below.

A. DATA COLLECTION

The experiment took place in a room within a house over the course of one week. Seven participants (P1 to P7) aged between 25 and 48 years (four males and three females) were involved in the experiment for one week each.

Three IP cameras were fixed at different locations in the room to record the participants' activity. The videos were recorded using a local server. The participants were asked to carry an Apple watch and an Apple TV remote during the experiment. Both the Apple watch and the Apple TV remote contained an embedded accelerometer, a gyroscope and microphone.

All the data from the buttons of the Apple TV remote and the Apple watch were also extracted.

The sampling frequencies for the accelerometers, gyroscopes and audio were set to 128 Hz, 132 Hz and 8 KHz, respectively. The recording time was 3 hours and 50 minutes, twice daily over one week.

We developed an app to access, record and send the sensor data via Wi-Fi to a local server, which was stored in a SQL server database.

A data warehouse was created to integrate the data automatically from the database at the end of each record.

The videos and sensor recordings were synchronized and started simultaneously. The data were labeled using ELAN software.

Participants held the Apple TV remote in their hand, and the Apple watch was attached to their wrist during the experiment.

B. DATA PREPARATION

The data collected from the accelerometer, gyroscope, microphones and buttons of the Apple TV remote and the Apple watch were integrated into a data warehouse using Extract-Transform-Load (ETL) software.

In the data warehouse, data were organized using Data Marts.

The data from the various sensors were stored in different Data Marts (one Data Mart for the accelerometer data, another for the gyroscope data, another for the microphones and finally another for the buttons).

From these Data Marts, data files were extracted and analyzed in order to recognize activities (standing, sitting, lying down and walking).

The data from the signals and buttons were resized and concatenated or merged, and then used as classifier inputs

(Fig. 3). Recordings were triggered automatically via an IOS application when the participant returned home.

In order to carry out classifications (of activities and then users) using our DNN model, we chose to extract descriptors from the collected signals.

There were two types of data:

A. Button data: this corresponds to integers. When the user presses a button, it enters a value of 1 into the database; otherwise it enters 0.

B. Signal data (accelerometer, gyroscope, audio): this corresponds to a time series. The data was disseminated, and then a selection phase of descriptors was applied.

For signal data, we compute the DCT, or more precisely, the DCT-II [27], [32]. The DCT is a very good signal decorrelator. It also allows grouping of the energy using low-frequency coefficients due to its approximation of the Karhunen Loeve transform and its use of principal component analysis (PCA) [33].

The DCT applies a transformation to the starting signal, and thus most of its energy is projected within a restricted area (a reduced number of coefficients) of the transformed space. The transformation used is linear, in order to provide an analytical solution for the subsequent reconstruction of the signal.

The window width (Δt) is an important parameter for the adopted method. It is illustrated in Fig 2.

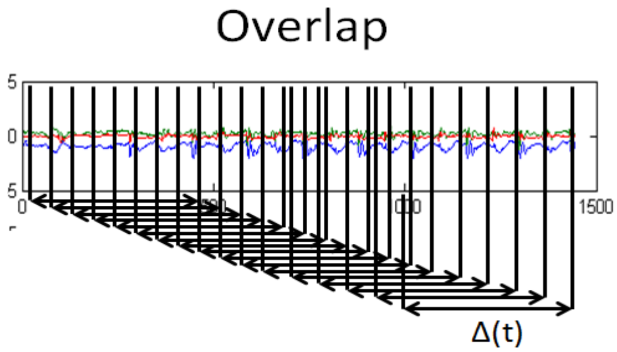

FIGURE 2. **Signal representing a walking pattern and the integration window (width of Δt).**

The extracted DCT showed the following characteristics:

A. **Window width:** The optimal window width is Δt = 512 with an overlap of 256.

B. **Feature size:**The size of the descriptor varied between 2 and 95, and the classification performance was calculated based on these dimensions. The classifier was the deep neural network model (DNN).
For the accelerometers and the gyroscope signals, we retain the first 48 DCT factors.

C. **DCT descriptor:** the DCT, as determined for each axis and for each sensor, was concatenated and used at the classification stage. The descriptors were stored in the database and used as input for the DNN.
For the audio signals, we used Open smile [34] to extract features. The DCT was extracted using the same window size as for the accelerometer and gyroscope signals.
All these features were extracted and concatenated from the sensor signals from the two devices and used as input classifiers. The DNN-based model was trained to classify the input activities' data.

C. ACTIVITY RECOGNITION

The descriptors and the button data for the two devices were used as input to the learning model to generate the classification of the activity into four classes.

In this paper, a DNN [35]–[39] was used as a classifier; this can extract features automatically and without requiring domain-specific knowledge of acceleration, gyroscope and microphone data.

The model could be made more responsive by skipping the process of extracting features.

The following notation is used to denote the components of the network:

- $I = h_0$: the input layer.
- h_i (i $= 1, 2, \ldots, \tau - 1$): the i^{th} hidden layer.
- O $= h_\tau$: the output layer.
- $w_i(\mathrm{i} = 1, \ldots, \tau)$: the connection weight matrix between h_i and h_{i+1}.
- $\rho_i(\mathrm{i} = 1, \ldots, \tau)$: the biases of neurons of layer h_i when they are activated by the h_{i+1} layer.
- ς_i (i $= 1, \ldots, \tau$) : the biases of neurons of layer h_i when they are activated by the h_{i-1} layer.
- Θ : all the network settings
- Ŧ: the training dataset
- $[f_{\theta(x)}]_i$: the score associated with the i^{th} label by our parameter network.

In addition, according to [40], for two adjacent layers h_{i-1} and h_i the activation functions can be defined as:

$$\mathrm{p}\ (h_{i-1,s} = 1|h_i) = \Gamma(\rho_{i,s} + \sum\nolimits_j w_{i,j}, h_{i,,j}) \tag{1}$$

$$\mathrm{p}\ (h_{i,t} = 1|h_{i-1}) = \Gamma(\varsigma_{i,t} + \sum\nolimits_j w_{i,j}, h_{i,j}) \tag{2}$$

$$\Gamma(x) = \frac{1}{(1 + \mathbf{e^{-x}})} \tag{3}$$

where Γ (x) is the logistic function [41].

The training process of the DNN is divided into two steps: *pre-training* and *fine-tuning*.

1) PRE-TRAINING

The pre-training is unsupervised, and an initial network is obtained using a greedy layer-wise training algorithm.

The goal of pre-training is to maximize the probability of generating training data. The probability of each set of training data assigned by the network was calculated using the energy function in Equation (4):

$$\mathrm{P}\ (\mathrm{I}) = \sum\nolimits_{h \in H} \boldsymbol{p}\ (\boldsymbol{v}, \boldsymbol{h}) = \frac{\sum_h \boldsymbol{exp}(-\boldsymbol{E}\ (\boldsymbol{I}, \boldsymbol{h}))}{\sum_{u,g} \boldsymbol{exp}(-\boldsymbol{E}\ (\boldsymbol{u}, \boldsymbol{g}))} \tag{4}$$

Hinton and Salakhutdinov [42] have proposed a method based on a layer-wise pre-training. This is used in order to obtain a suitable neural network, by treating the lower layer as a visible layer, v, and the upper layer as a hidden layer, h. Each couple of adjacent layers can be considered as a restricted Boltzmann machine (RBM) [43].

The entire network is constructed by training one RBM, which has the following energy function:

$$\mathrm{E}\,(\mathrm{v},\mathrm{h}) = -\sum\nolimits_{s,t} v_s w_{st} h_t - \sum\nolimits_{s} b_s b_v - \sum\nolimits_{t} c_i h_i \quad (5)$$

2) FINE TUNING

The fine-tuning is supervised, and the parameters of all the layers are updated using a back propagation algorithm.

The model was trained using the maximum likelihood of the training set by a gradient descent stochastic. The aim is to maximize the log-likelihood [44]:

$$\Theta \rightarrow \sum_{(x,y)\in \dot{T}} \log\left(y|x,\Theta\right) \quad (6)$$

where x is the input data and y corresponds to the labels. If x is a given example, the probability p is calculated from the outputs of the neural network by means of a soft max:

$$P(i|x,\theta) = e^{[f_{\theta(x)}]_i} \quad (7)$$

This allows us to easily express the log-likelihood:

$$\mathrm{Log\ p}\,(\mathrm{y}|\mathrm{x},\theta) = [f_\theta\,(x)]_y - \log(\sum_j e^{[f_\theta(x)]_j}) \quad (8)$$

Maximization of the log-likelihood using a stochastic gradient is carried out by randomly selecting a training example (x, y) and by performing a gradient descent:

$$\Theta \rightarrow \Theta + \varphi \frac{\delta \log py|x,\Theta)}{\delta\Theta} \quad (9)$$

where φ is the learning rate.

All the proposed architecture networks were performed using the Theano Library [45].

The effectiveness of the proposed method was evaluated using the database and tested using ten-fold cross-validation.

The number of hidden layers of our model was set to five and the numbers of neurons of the hidden layers were set as 850- 340-430-920-870. The other parameters of the network were fixed as the default parameter settings of Hinton's DBN package [42].

Once the activities have been classified, we detect the identity of the different users during each type of activity. This order of the methodology is important. That is to say the recognition of the activity then the detection of the users.

The following section explains the procedure followed for identifying users and the settings considered.

D. USER CLASSIFICATION

The classification carried out in Step (3) involves the four activities, with all participants considered together (that is, without taking account of the participants individually). For example, all of the "sitting" activities of participants were merged and classified against the other three remaining activities (standing, walking, and lying down).

In this part, the classification is made a little finer; a classification is made with respect to the participant's identity (see Fig.4).

The participants are characterized by a set of habits or behaviors (which are restricted here to the physical activities of sitting, standing, lying down and walking), with respect to the Apple TV remote and Apple watch (using the buttons of the different devices).

Based on this information, the identity of the user of the Apple TV remote is determined.

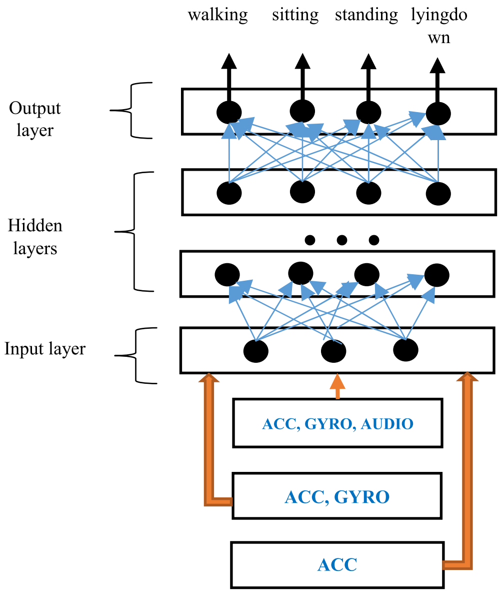

FIGURE 3. The architecture of the method used.

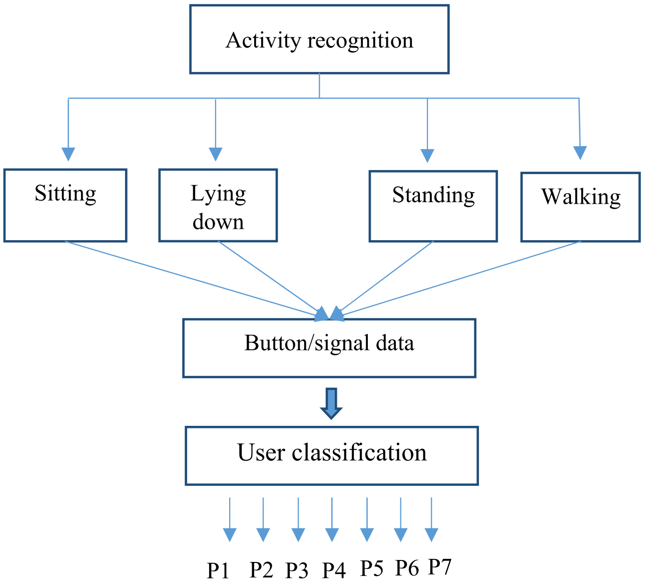

FIGURE 4. User identification process corresponding to the second classification level (cf. Fig.1, Step 4).

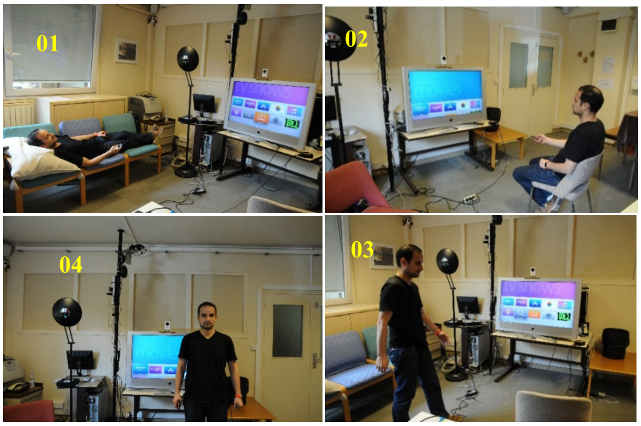

FIGURE 5. **The four activities studied: (01): the participant is lying down and holds the Apple TV remote in his hand; (02): the participant is sitting and holds the Apple TV remote in his hand; (03): the participant is walking and holds the Apple TV remote in his hand; (04): the participant is standing, holding the Apple TV remote in his hand.**

TABLE 1. **Classification results of activities using 1) the Apple watch alone; 2) the Apple watch and Apple TV together.**

Activity/device	Apple watch	Apple watch and Apple TV
Standing/other	93.11%	95.08%
Sitting/other	92.98%	94.48%
Lying down/walking	95.55%	97.39%

The participants' data were recorded independently and an attempt was made to classify them using the DNN. The seven participants performed the four activities (standing, sitting, walking, lying down) by manipulating the Apple TV remote and the Apple watch. The aim here was to classify the participants in relation to the four activities and the different actions performed using the Apple TV remote. Fig. 5 shows the four activities studied: lying down, sitting, standing and walking.

The recorded raw acceleration, gyroscope and audio signal streams were all cropped to the same size, with an overlap of 256 point samples. The length of the acceleration, gyroscope and audio data was about 4.16 seconds for each sensor.

V. RESULTS

Table 1 below presents the detailed results [46] of the classifications of the seven participants'activities involving the two devices. We used the Leave one out method to learn our model.

Note that performance recognition accuracies are high (between 92.98% and 95.55% for the Apple watch data) and more than 97% classification performance is achieved using the Apple watch and Apple TV remote data together.

In order to recognize their identity, participants were classified according to their four activities with respect to the two types of data: button data and signal data.

A. BUTTON DATA

For each activity (standing, sitting, walking, lying down), the behavior of each participant was studied with respect to the Apple TV remote and the Apple watch, in order to understand what made it possible to distinguish between the different participants during the same activity.

The different buttons of the Apple TV remote were renamed as follows:

- Button 1: Siri.
- Button 2: Play/pause.
- Button 3: Volume.
- Button 4: Home.
- Button 5: Touch surface.
- Button 6: Menu.

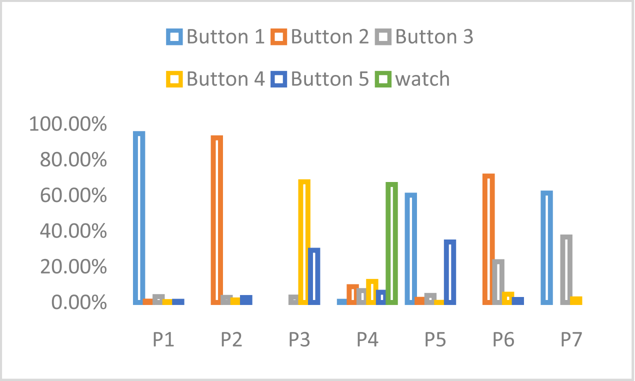

FIGURE 6. **Users' behaviors illustrated by pressing the buttons of the Apple TV remote and Apple watch during sitting activity.**

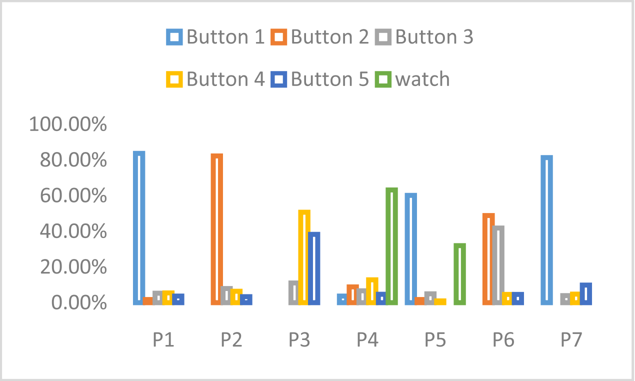

FIGURE 7. **Users' behaviors illustrated by pressing the buttons of the Apple TV remote and Apple watch during lying down activity.**

Figs. 6, 7 and 8 show the behavior of the seven participants during the sitting, lying down and standing activities.

It should be noted that the participants interacted with the two devices in very different ways depending on the activity. In other words, each participant tended to use the Apple TV remote in a certain way with a preference for using certain buttons of the devices. For instance, participants P1 and P2 used Buttons 1 and 2 for more than 90% and 85% of the sitting activity, respectively (Fig.6). They also tended to use the

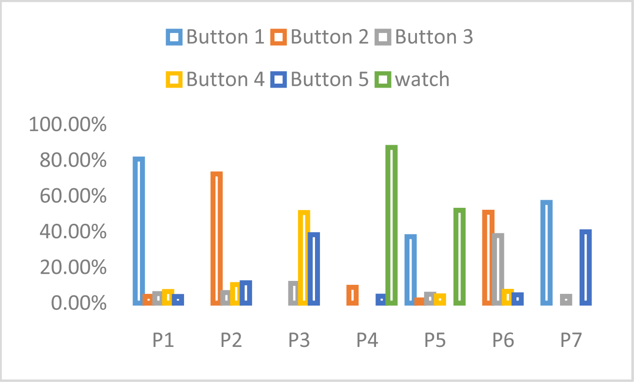

FIGURE 8. Users' behaviors illustrated by pressing the buttons of the Apple TV remote and Apple watch during standing activity.

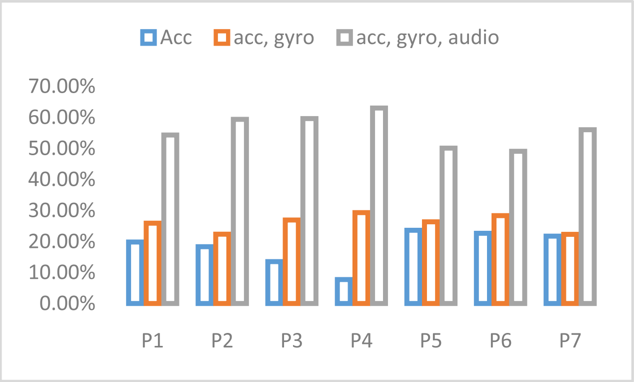

FIGURE 9. Users' behaviors illustrated by their use of the Apple TV remote and Apple watch during the sitting activity.

same buttons during the lying down and standing activities, respectively (as shown in Figs. 7 and 8).

The walking activity was not detected using the buttons; this was studied using only the signal data, and was the activity with the lowest level of recognition.

B. SIGNAL DATA

The same study as in (A) was carried out for the accelerometer, gyroscope and audio data signals. The results are presented in Figures 9 to 12.

Unlike the button data, the behavior of the various participants when using the accelerometer, gyroscope and audio data was approximately similar. For instance, participants P6 and P7 show approximately the same accelerometer, gyroscope and audio data signals for the sitting, walking, laying down and standing activities (Figs. 9 to 12). However,these behaviors tended to be more distinguishable when all the sensor data signals were concatenated. For instance, P6 and P7 used more than 50% and 60% of the concatenated sensor data signals, respectively, during the sitting activity (Fig.9).

The contribution of the data signals to an understanding of participants' behaviors becomes much clearer after merging all the signals. However, the information provided by the audio data made these data signals' contributions more visible.

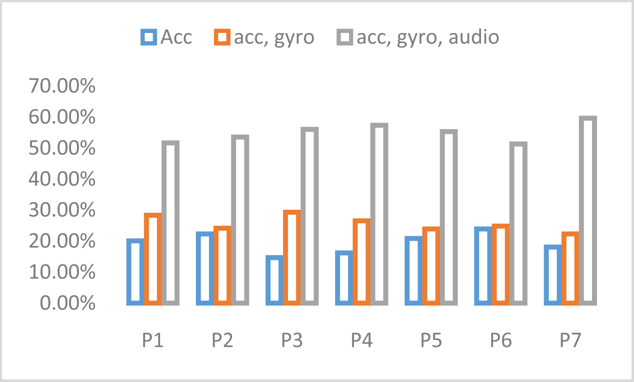

FIGURE 10. Users' behaviors illustrated by their use of the Apple TV remote and Apple watch during the lying down activity.

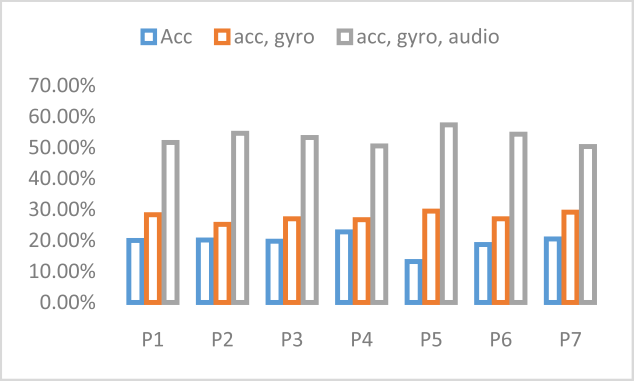

FIGURE 11. Users' behaviors illustrated by their use of the Apple TV remote and the Apple watch during the standing activity.

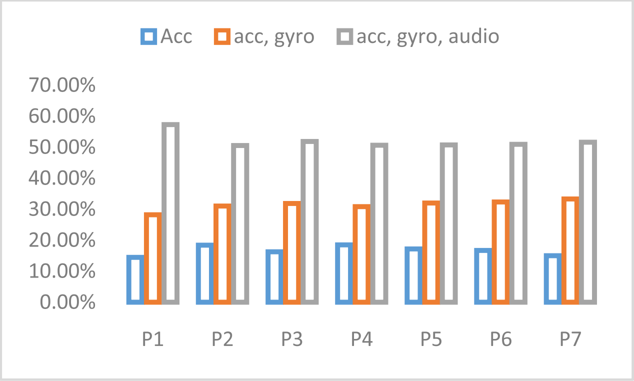

FIGURE 12. Users' behaviors illustrated by their use of the Apple TV remote and Apple watch during the walking activity.

The concatenation of all these data sources yields the classification results shown in Table 2.

User identification for the sitting and lying down activities was better than for the walking and standing activities. For instance, P6 and P7 reached 90.12% and 80.01% accuracy during the sitting and lying down activities, respectively.

In the next section, we will discuss these results and compare these with the state of the art.

TABLE 2. Classification of individuals in relation to the button and signal data using the DNN algorithm.

	Sitting	Lying down	Walking	Standing	Mean
P1/others	89.36%	**91.81%**	69.32%	60.22%	77.67%
P2/others	77.29%	88.78%	66.17%	61.77%	73.50%
P3/others	81.33%	85.95%	61.28%	59.88%	72.11%
P4/others	79.39%	88.18%	58.58%	55.95%	70.52%
P5/others	84.33%	77.98%	60.01%	63.28%	71.4%
P6/P7	**90.12%**	80.01%	61.77%	60.08%	72.99%

VI. DISCUSSION

The recognition of the four activities achieved a good performance of between 92.98% and 95.55% using the Apple watch, and 97.39% accuracy using the Apple TV remote and Apple watch together (see Table 1). This shows that the merging of data signals from two devices can significantly improve the performance of classification.

The information provided by the two devices can improve the accuracy of the classification of the participants. In other words, we aimed to understand how the seven participants behaved during the four types of activities, using these two devices.

A. PARTICIPANT's BEHAVIOURS

Fig.6 clearly shows that during the sitting activity, the participants' behavior was very different from each other. During the sitting activity, participant P1 pressed Button 1 (Siri) during more than 90% of the time and did not press the Apple watch buttons at all. In another words, P1 was mainly interested in voice searches or in using the Apple TV remote with the Siri button.

Participant P2 used Button 2 (the play/pause button) for more than 92% of the sitting activity. A check made of the three IP camera videos showed that P2 was listening to music via YouTube. This explains his excessive use of Button 2. This participant did not use the buttons of the Apple watch.

Participant P3 used only Button 4 (home) and Button 5 (surface), for 67% and 29.3% of the sitting activity, respectively. A check made of the three IP video cameras showed that this participant watched only contents on the Apple TV.

Participant (P4) used the Apple Watch buttons for 66.20% of the sitting activity. This participant also used all the other buttons of the Apple TV but fora shorter time duration. This can be explained by his interest in the Apple Watch rather than the Apple TV contents.

Participant P5 used Button 1 (Siri) and Button 5 (surface) for 60% and 33% of the sitting activity respectively. This can be explained by the fact that this participant searched for voice content and validated it using Button 5.

Participant P6 used Button 2 (play/pause) and Button 3 (volume ±) for 70.88% and 22.84% of the sitting activity, respectively. This participant only played audio content on the Apple TV.

Participant P7 used Button 1 and Button 3 for 61% and 36% of the sitting activity. He carried out voice searches, and listened to what the search returned while increasing or lowering the volume.

Fig.7 shows the participants' behavior during the lying down activity. The participants' behaviors were very different from each other.

Participants P1, P5 and P7 participants used Button 1 for 83%, 60% and 81% of the laying down activity time, respectively. However, the difference between these participants was that P1 tended to use all the buttons at a low rate and did not use the Apple watch buttons. However, P5 used the Apple watch for 63% of the lying down activity. Unlike participant P5, P7 tended to use Buttons 1 and 2. He did not use the Apple watch buttons.

Fig. 8 illustrates the participants' behavior during the standing activity. Participants P1, P5 and P7 used Button 1 (Siri) for 80%, 37% and 56% of the standing activity, respectively. However, the difference between these participants' behaviors was that P1 did not use the Apple watch buttons. Participants P2 and P6 each tended to use Button 2; however, P2 used Button 3 for only 5% of the standing activity, while P6 used the same button for 38% of the standing activity.

In order to understand the participant's behaviors using the signal sensor data of the two devices, we analyzed the signals data from the accelerometers, gyroscopes and audio.

Figs. 9 to 12 show that the merging of data signals provides more than 50% of the information on the activity and the participants'behaviors during the sitting, lying down, standing and walking activities. For instance, Fig.9 shows that the information provided by the accelerometers does not represent more than 30% of the sitting activity. This is therefore a weak method of distinguishing participants for this activity. Thus, the addition of other signals (gyroscopes and audio) provides more than 50% of the behavior information for all participants.

It should be noted that the information data provided from the accelerometers does not contribute more than 20% to each activities for all seven participants, although the merging of the data signals from the accelerometer and gyroscopic sensors with audio data increases the percentages by more than 50%.

This study gives a clear idea of the classification of the seven participants according to their four activities.

During the walking activity, the participants did not use the Apple TV remote. The main information concerning the identification of the seven participants in the walking activity was generated through the signal data (accelerometer, gyroscope and audio).

The DNN-based model was trained using each data activity type in order to determine the identity of the seven participants. The accuracies of user classification in relation to sitting and lying down activities were better than those for standing and walking activities. For instance, unlike the walking and standing activities, for which the participant classification accuracies did not exceed 69%, the participant accuracies for a certain device configuration (Apple TV and

Apple watch) were between 77% and 91% for the sitting and lying down activities.

The difference in terms of classification performance between the sitting/lying down and standing/walking activities was mainly due to the participants' behavior towards the Apple TV remote and Apple watch. More precisely, this was due to the button data of both devices, where participant P1 was identified with 91.81% accuracy during the lying down activity, as shown in Table 2.

Finally, we compared our results with two other basic classification models (SVM, naïve Bayes classifier) using our DNN model [46]–[48]. Table 4 gives the classification accuracies.

TABLE 3. Comparison of classification performance (using the average of the classification accuracy for each activity).

	SVM	C45	NB	DNN
P1/others	62.32%	59.32%	55.06%	**77.67%**
P2/others	60.66%	59.64%	51.32%	73.50%
P3/others	59.6%	55.65%	49.62%	72.11%
P4/others	58.32%	60.04%	49.65%	70.52%
P5/others	61.25%	61.15%	50.51%	71.4%
P6/P7	60.28%	60.9%	51.32%	72.99%

Table 3 gives the classification averages for each participant in relation to all four activities. For instance, P1 has been classified with an average accuracy of 62.32%. This means that P1 was identified with 62.32% accuracy compared to other participants for all activities (sitting, lying down, standing and walking).

It should be noted that our model gives a better classification performance on average. For instance, participant P1 was classified with 77.67% accuracy, taking into account all activities combined. In addition, SVM gives a better classification performance than both of the other classification algorithms (C45 and naive Bayes classifier). However, it is significantly less efficient than our DNN classification model. For instance, participant P1 showed better classification performance than the other participants using SVM and C45.

Our model outperforms all of these models. In addition, our approach is much more robust and precise than the other approaches. For instance, Datta and Manousakis [13] obtained an accuracy of 85%. They did not use activity recognition to identify their users. However, our method for classification of individuals in relation to the button and signal data achieves a performance of up to 91.81% (see Table 2).

VII. CONCLUSION

In this paper, we propose a methodology that can help to identify the user of a smart object, based on the recognition of the participant's activities.

This method is based on the classification of the individual's activities and habits when using the Apple TV remote. Our method uses a classification based on four types of activity (sitting, standing, lying down, and walking) using a DNN model; following this, a second classification is made to classify individuals within each activity type.

Some activities (sitting, lying down) allow for a better classification of participants. For instance, the walking and lying down activities classified participants less effectively.

This work opens new avenues for future work, including the selection of other types of audio descriptors and other smart objects.

ACKNOWLEDGMENTS

The authors would like to thank all the participants in the different experiences. This work was supported by the company IofMe.

REFERENCES

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generat. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[3] G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, "Smart objects as building blocks for the Internet of Things," *IEEE Internet Comput.*, vol. 14, no. 1, pp. 44–51, Jan./Feb. 2010.

[4] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfflé, "Vision and challenges for realising the Internet of Things," Cluster Eur. Res. Projects Internet Things, Eur. Commission, Tech. Rep., 2010.

[5] "Keep your phone safe: How to protect yourself from wireless threats," Tech. Rep., Jun. 2013.

[6] B. Xie and Q. Wu, "Hmm-based tri-training algorithm in human activity recognition with smartphone," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Intell. Syst. (CCIS)*, vol. 1. Oct./Nov. 2012, pp. 109–113.

[7] P. Sarcevic, Z. Kincses, and S. Pletl, "Comparison of different classifiers in movement recognition using WSN-based wrist-mounted sensors," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Apr. 2015, pp. 1–6.

[8] L. Fan, Z. Wang, and H. Wang, "Human activity recognition model based on decision tree," in *Proc. Int. Conf. Adv. Cloud Big Data (CBD)*, Dec. 2013, pp. 64–68.

[9] H.-J. Kim, J. S. Lee, and J.-H. Park, "Dynamic hand gesture recognition using a CNN model with 3D receptive fields," in *Proc. Int. Conf. Neural Netw. Signal Process.*, Jun. 2008, pp. 14–19.

[10] L. Zhang, X. Wu, and D. Luo, "Recognizing human activities from raw accelerometer data using deep neural networks," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 865–870.

[11] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, 2011, pp. 289–296.

[12] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *Proc. IEEE SENSORS*, Oct./Nov. 2016, pp. 1–3.

[13] T. Datta and K. Manousakis, "Using SVM for user profiling for autonomous smartphone authentication," in *Proc. IEEE MIT Undergraduate Res. Technol. Conf. (URTC)*, Nov. 2015, pp. 1–5.

[14] H. G. Kayacik, M. Just, L. Baillie, D. Aspinall, and N. Micallef, "Data driven authentication: On the effectiveness of user behaviour modelling with mobile device sensors," in *Proc. IEEE S&P Symp. 3rd Mobile Secur. Technol. Workshop (MoST)*, May 2014, pp. 1–22.

[15] W.-H. Lee and R. B. Lee, "Multi-sensor authentication to improve smartphone security," in *Proc. 1st Int. Conf. Inf. Syst. Secur. Privacy*, Feb. 2015, pp. 270–280.

[16] J.-S. Wu, W.-C. Lin, C.-T. Lin, and T.-E. Wei, "Smartphone continuous authentication based on keystroke and gesture profiling," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, 2015, pp. 191–197.

[17] W. Meng, D. S. Wong, S. Furnell, and J. Zhou, "Surveying the development of biometric user authentication on mobile phones," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1268–1293, 3rd Quart., 2015.

[18] M. Antal and L. Z. Szabó, "An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices," [Online]. Available: http://www.ms.sapientia.ro/manyi/research/43.pdf

[19] L. Cai and H. Chen,"TouchLogger: Inferring keystrokes on touch screen from smartphone motion," in *Proc. 6th USENIX Conf. Hot Topics Secur.*, 2011.

[20] D. Buschek, A. De Luca, and F. Alt,"Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2015, pp. 1393–1402.

[21] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 136–148, Jan. 2013.

[22] B. Draffin, J. Zhu, and J. Zhang, "KeySens: Passive user authentication through micro-behavior modeling of soft keyboard interaction," in *Proc. 5th Int. Conf., MobiCASE*, 2013, pp. 184–201.

[23] H. Xu, Y. Zhou, and M. R. Lyu, "Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones," in *Proc. Symp. Usable Privacy Secur. (SOUPS)*, 2014, pp. 187–198.

[24] L. Li, X. Zhao, and G. Xue,"Unobservable re-authentication for smartphones," NDSS, The Internet Society, Reston, VA, USA, Tech. Rep., 2013.

[25] F. G. da Silva and E. Galeazzo, "Accelerometer based intelligent system for human movement recognition," in *Proc. 5th IEEE Int. Workshop Adv. Sensors Interfaces (IWASI)*, Jun. 2013, pp. 20–24.

[26] Z.-Y. He and L.-W. Jin, "Activity recognition from acceleration data using AR model representation and SVM," in *Proc. Int. Conf. Mach. Learn.*, vol. 4. 2008, pp. 2245–2250.

[27] Z. He and L. Jin, "Activity recognition from acceleration data based on discrete consine transform and SVM," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2009, pp. 5041–5044.

[28] S. Pirttikangas, K. Fujinami, and T. Nakajima, "Feature selection and activity recognition from wearable sensors," in *Proc. Int. Symp. Ubiquitious Comput. Syst.*, 2006, pp. 516–527.

[29] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2016, pp. 381–388.

[30] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2015, pp. 1488–1492.

[31] M. Zeng *et al.*, "Convolutional neural networks for human activity recognition using mobile sensors," in *Proc. 6th Int. Conf. Mobile Comput., Appl. Services (MobiCASE)*, 2014, pp. 197–205.

[32] S. M. Kia, E. Olivetti, and P. Avesani, "Discrete cosine transform for MEG signal decoding," in *Proc. Int. Workshop Pattern Recognit. Neuroimag. (PRNI)*, 2013, pp. 132–135.

[33] D. Tran and A. Sorokin, "Human activity recognition with metric learning," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 548–561.

[34] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The Munich versatile and fast open-source audio feature extractor," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 1459–1462.

[35] L. Mo, F. Li, Y. Zhu, and A. Huang, "Human physical activity recognition based on computer vision with deep learning model," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2016, pp. 1–6.

[36] H. Yalçın, "Human activity recognition using deep belief networks," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, 2016, pp. 1649–1652.

[37] X. Yin and Q. Chen, "Deep metric learning autoencoder for nonlinear temporal alignment of human motion," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 2160–2166.

[38] K. Nakadai, T. Mizumoto, and K. Nakamura, "Robot-audition-based human-machine interface for a car," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 6129–6136.

[39] S. Jia, T. Lansdall-Welfare, and N. Cristianini, "Gender classification by deep learning on millions of weakly labelled images," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 462–467.

[40] T. Liu, M. Li, S. Zhou, and X. Du, "Sentiment classification via L2-norm deep belief network," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 2489–2492.

[41] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[42] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[43] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[44] X. Cui and V. Goel, "Maximum likelihood nonlinear transformations based on deep neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 11, pp. 2023–2031, Nov. 2016.

[45] J. Ray, B. Thompson, and W. Shen, "Comparing a high and low-level deep neural network implementation for automatic speech recognition," in *Proc. 1st Workshop High Perform. Techn. Comput. Dyn. Lang.*, 2014, pp. 41–46.

[46] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. Int. Conf. Pervasive Comput.*, 2004, pp. 1–17.

[47] S. E. El-Khamy, H. A. Elsayed, and M. M. Rizk, "C45. Classification of OFDM signals using higher order statistics and clustering techniques," in *Proc. 29th Nat. Radio Sci. Conf. (NRSC)*, 2012, pp. 541–549.

[48] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *Proc. 23rd Int. Conf. Archit. Comput. Syst. (ARCS)*, 2010, pp. 1–10.

HAMDI AMROUN received the master's degree in mathematical engineering from the University of Evry Val d'Essonne, the master's degree in data mining from the University of Versailles Saint Quentin en Yvelines, and the Masters of Sciences degree in artificial intelligence from Télécom Paris Sud. He is currently pursuing the Ph.D. degree in computer science in LIMSI-CNRS, University of Paris-Sud.

MEHDI AMMI received the M.A. degree in computer science from the University of Evry Val d'Essonne, the Ph.D. degree in robotics from the University of Orléans, and the Habilitation degree in computer sciences from the University of Paris-Sud. Since 2006, he has been an Assistant Professor of computer sciences with the LIMSI-CNRS, University of Paris-Sud. His research deals with robotics, HCI, and Internet of Things.