A Survey on Human Activity Recognition using Wearable Sensors

Óscar D. Lara and Miguel A. Labrador
Department of Computer Science and Engineering
University of South Florida, Tampa, FL 33620
olarayej@mail.usf.edu, labrador@cse.usf.edu

***Abstract*—Providing accurate and opportune information on people's activities and behaviors is one of the most important tasks in pervasive computing. Innumerable applications can be visualized, for instance, in medical, security, entertainment, and tactical scenarios. Despite human activity recognition (HAR) being an active field for more than a decade, there are still key aspects that, if addressed, would constitute a significant turn in the way people interact with mobile devices. This paper surveys the state of the art in HAR based on wearable sensors. A general architecture is first presented along with a description of the main components of any HAR system. We also propose a two-level taxonomy in accordance to the learning approach (either supervised or semi-supervised) and the response time (either offline or online). Then, the principal issues and challenges are discussed, as well as the main solutions to each one of them. Twenty eight systems are qualitatively evaluated in terms of recognition performance, energy consumption, obtrusiveness, and flexibility, among others. Finally, we present some open problems and ideas that, due to their high relevance, should be addressed in future research.**

***Index Terms*—Human-centric sensing; machine learning; mobile applications; context awareness.**

I. Introduction

During the past decade, there has been an exceptional development of microelectronics and computer systems, enabling sensors and mobile devices with unprecedented characteristics. Their high computational power, small size, and low cost allow people to interact with the devices as part of their daily living. That was the genesis of *Ubiquitous Sensing*, an active research area with the main purpose of extracting knowledge from the data acquired by pervasive sensors [1]. Particularly, the recognition of human activities has become a task of high interest within the field, especially for medical, military, and security applications. For instance, patients with diabetes, obesity, or heart disease are often required to follow a well defined exercise routine as part of their treatment [2]. Therefore, recognizing activities such as *walking*, *running*, or *cycling* becomes quite useful to provide feedback to the caregiver about the patient's behavior. Likewise, patients with dementia and other mental pathologies could be monitored to detect abnormal activities and thereby prevent undesirable consequences [3]. In tactical scenarios, precise information on the soldiers' activities along with their locations and health conditions, is highly beneficial for their performance and safety. Such information is also helpful to support decision making in both combat and training scenarios.

The first works on *human activity recognition* (HAR) date back to the late '90s [4], [5]. However, there are still many issues that motivate the development of new techniques to improve the accuracy under more realistic conditions. Some of these challenges are (1) the selection of the attributes to be measured, (2) the construction of a portable, unobtrusive, and inexpensive data acquisition system, (3) the design of feature extraction and inference methods, (4) the collection of data under realistic conditions, (5) the flexibility to support new users without the need of re-training the system, and (6) the implementation in mobile devices meeting energy and processing requirements [6].

The recognition of human activities has been approached in two different ways, namely using *external* and *wearable* sensors. In the former, the devices are fixed in predetermined points of interest, so the inference of activities entirely depends on the voluntary interaction of the users with the sensors. In the latter, the devices are attached to the user.

Intelligent homes [7]–[10] are a typical example of external sensing. These systems are able to recognize fairly complex activities (e.g., eating, taking a shower, washing dishes, etc.) because they rely on data from a number of sensors placed in target objects which people are supposed to interact with (e.g., stove, faucet, washing machine, etc.). Nonetheless, nothing can be done if the user is out of the reach of the sensors or they perform activities that do not require interaction with them. Additionally, the installation and maintenance of the sensors usually entail high costs.

Cameras have also been employed as external sensors for HAR. In fact, the recognition of activities and gestures from video sequences has been the focus of extensive research [11]–[14]. This is especially suitable for security (e.g, intrusion detection) and interactive applications. A remarkable example, and also commercially available, is the *Kinect* game console [15] developed by Microsoft. It allows the user to interact with the game by means of gestures, without any controller device. Nevertheless, video sequences certainly have some issues in HAR. The first one is *privacy*, as not everyone is willing to be permanently monitored and recorded by cameras. The second one is *pervasiveness* because video recording devices are difficult to attach to target individuals in order to obtain images of their entire body during daily living activities. The monitored individuals should then stay within a perimeter defined by the position and the capabilities of the camera(s). The last issue would be *complexity*, since video processing techniques are relatively expensive, computationally speaking, hindering a real time HAR system to be scalable.

The aforementioned limitations motivate the use of wearable

TABLE I
TYPES OF ACTIVITIES RECOGNIZED BY STATE-OF-THE-ART HAR SYSTEMS.

Group	**Activities**
Ambulation	Walking, running, sitting, standing still, lying, climbing stairs, descending stairs, riding escalator, and riding elevator.
Transportation	Riding a bus, cycling, and driving.
Phone usage	Text messaging, making a call.
Daily activities	Eating, drinking, working at the PC, watching TV, reading, brushing teeth, stretching, scrubbing, and vacuuming
Exercise/fitness	Rowing, lifting weights, spinning, Nordic walking, and doing push ups.
Military	Crawling, kneeling, situation assessment, and opening a door.
Upper body	Chewing, speaking, swallowing, sighing, and moving the head.

sensors in HAR. Most of the measured attributes are related to the user's movement (e.g., using accelerometers or GPS), environmental variables (e.g., temperature and humidity), or physiological signals (e.g., heart rate or electrocardiogram). These data are naturally indexed over the time dimension, allowing us to define the human activity recognition problem as follows:

Definition 1 (HAR problem (HARP)): Given a set $S = \{S_0, ..., S_{k-1}\}$ of k time series, each one from a particular measured attribute, and all defined within time interval $I = [t_\alpha, t_\omega]$, the goal is to find a temporal partition $\langle I_0, ..., I_{r-1}\rangle$ of I, based on the data in S, and a set of labels representing the activity performed during each interval I_j (e.g., sitting, walking, etc.). This implies that time intervals I_j are consecutive, non-empty, non-overlapping, and such that $\bigcup_{j=0}^{r-1} I_j = I$.

This definition is valid assuming that activities are not simultaneous, i.e., a person does not walk and run at the same time. A more general case with overlapping activities will be considered in Section VI. Note that the HARP is not feasible to be solved deterministically. The number of combinations of attribute values and activities can be very large —or even infinite— and finding transition points becomes hard as the duration of each activity is generally unknown. Therefore, machine learning tools are widely used to recognize activities. A relaxed version of the problem is then introduced dividing the time series into fixed length time windows.

Definition 2 (Relaxed HAR problem): Given (1) a set $W = \{W_0, ..., W_{m-1}\}$ of m equally sized time windows, totally or partially labeled, and such that each W_i contains a set of time series $S_i = \{S_{i,0}, ..., S_{i,k-1}\}$ from each of the k measured attributes, and (2) a set $A = \{a_0, ..., a_{n-1}\}$ of activity labels, the goal is to find a mapping function $f : S_i \to A$ that can be evaluated for all possible values of S_i, such that $f(S_i)$ is as similar as possible to the actual activity performed during W_i.

Notice this relaxation introduces some error to the model during *transition windows*, since a person might perform more than one activity within a single time window. However, the number of transitions is expected to be much smaller than the total number of time windows, which makes the relaxation error not significant for most of the applications.

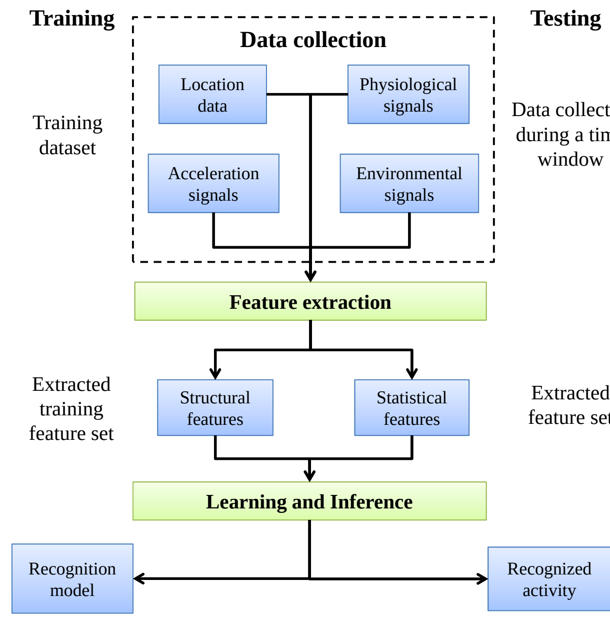

Fig. 1. General data flow for training and testing HAR systems based on wearable sensors.

The design of any HAR system depends on the activities to be recognized. In fact, changing the activity set A immediately turns a given HARP into a completely different problem. From the literature, seven groups of activities can be distinguished. These groups and the individual activities that belong to each group are summarized in Table I.

This paper surveys the state of the art in HAR making use of wearable sensors. Section II presents the general components of any HAR system. Section III introduces the main design issues for recognizing activities and the most important solutions to each one of them. Section IV describes the principal techniques applied in HAR, covering feature extraction and learning methods. Section V shows a qualitative evaluation of state-of-the-art HAR systems. Finally, Section VI introduces some of the most relevant open problems in the field providing directions for future research.

II. GENERAL STRUCTURE OF HAR SYSTEMS

Similar to other machine learning applications, activity recognition requires two stages, i.e., *training* and *testing* (also called *evaluation*). Figure 1 illustrates the common phases involved in these two processes. The training stage initially requires a time series dataset of measured attributes from individuals performing each activity. The time series are split into time windows to apply *feature extraction* thereby filtering relevant information in the raw signals. Later, *learning* methods are used to generate an activity recognition model from the dataset of extracted features. Likewise, for testing, data are collected during a time window, which is used to extract features. Such feature set is evaluated in the priorly trained learning model, generating a predicted activity label.

We have also identified a generic data acquisition architecture for HAR systems, as shown in Figure 2. In the first place, *wearable sensors* are attached to the person's body to measure attributes of interest such as motion [16], location [17], temperature [18], ECG [19], among others. These sensors should communicate with an *integration device* (ID), which can be a cellphone [20], [21], a PDA [18], a laptop [19], [22], or a customized embedded system [23], [24]. The main purpose of the ID is to preprocess the data received from the sensors and, in some cases, send them to an application server for real time monitoring, visualization, and/or analysis [19], [25]. The *communication* protocol might be UDP/IP or TCP/IP, according to the desired level of reliability.

Notice that all of these components are not necessarily implemented in every HAR system. In [26]–[28], the data are collected offline, so there is neither communication nor server processing. Other systems incorporate sensors within the ID [29]–[31], or carry out the inference process directly on it [30], [32]. The presented architecture is rather general and the systems surveyed in this paper are particular instances of it.

III. Design Issues

We have distinguished seven main issues pertaining to human activity recognition, namely, (1) *selection of attributes and sensors*, (2) *obtrusiveness*, (3) *data collection protocol*, (4) *recognition performance*, (5) *energy consumption*, (6) *processing*, and (7) *flexibility*. The main aspects and solutions related to each one of them are analyzed next. In Section V, the systems surveyed in this paper will be evaluated in accordance to these issues.

A. Selection of attributes and sensors

Four groups of attributes are measured using wearable sensors in a HAR context: *environmental attributes*, *acceleration*, *location*, and *physiological signals*.

1) Environmental attributes: These attributes, such as temperature, humidity, audio level, etc., are intended to provide context information describing the individual's surroundings. If the *audio level* and *light intensity* are *fairly low*, for instance, the subject may be *sleeping*. Various existing systems have utilized microphones, light sensors, humidity sensors, and thermometers, among others [18], [24]. Those sensors alone, though, might not provide sufficient information as individuals can perform each activity under diverse contextual conditions in terms of weather, audio loudness, or illumination. Therefore, environmental sensors are generally accompanied by accelerometers and other sensors [3].

2) Acceleration: Triaxial accelerometers are perhaps the most broadly used sensors to recognize ambulation activities (e.g., *walking*, *running*, *lying*, etc.) [26]–[28], [33]–[35]. Accelerometers are inexpensive, require relatively low power [36], and are embedded in most of today's cellular phones. Several papers have reported high recognition accuracy 92.25% [28], 95% [37], 97% [34], and up to 98% [38], under different evaluation methodologies. However, other daily activities such as *eating*, *working at a computer*, or *brushing teeth*, are confusing from the acceleration point of view. For instance, eating might be confused with brushing teeth due to arm motion [24]. The impact of the sensor specifications have also been analyzed. In fact, Maurer et al. [24] studied the behavior of the recognition accuracy as a function of the accelerometer sampling rate (which lies between 10 Hz [3] and 100 Hz [23]). Interestingly, they found that no significant gain in accuracy is achived above 20 Hz for ambulation activities. In addition, the amplitude of the accelerometers varies from $\pm$2g [24], up to $\pm$6g [38] yet $\pm$2g was shown to be sufficient to recognize ambulation activities [24]. The placement of the accelerometer is another important point of discussion: He et al. [28] found that the best place to wear the accelerometer is inside the trousers pocket. Instead, other studies suggest that the accelerometer should be placed in a bag carried by the user [24], on the belt [39], or on the dominant wrist [22]. At the end, the optimal position where to place the accelerometer depends on the application and the type of activities to be recognized.

3) Location: The Global Positioning System (GPS) enables all sort of *location based services*. Current cellular phones are equipped with GPS devices, making this sensor very convenient for context-aware applications, including the recognition of the user's transportation mode [36]. The place where the user is can also be helpful to infer their activity using ontological reasoning [30]. As an example, if a person is at a *park*, they are probably not *brushing their teeth* but might be *running* or *walking*. And, information about places can be easily obtained by means of the Google Places Web Service [40], among other tools. However, GPS devices do not work well indoors and they are relatively expensive in terms of energy consumption, especially in real-time tracking applications [36]. For those reasons, this sensor is usually employed along with accelerometers [30]. Finally, location data has privacy issues because users are not always willing to be tracked. Encryption, obfuscation, and anonymization are some of the techniques available to ensure privacy in location data [41]–[43].

4) Physiological signals: Vital signs data (e.g., heart rate, respiration rate, skin temperature, skin conductivity, ECG, etc.) have also been considered in a few works [3]. Tapia et al. [22] proposed an activity recognition system that combines data from five triaxial accelerometers and a heart rate monitor. However, they concluded that the heart rate is not useful in a HAR context because after performing physically demanding activities (e.g., running) the heart rate remains at a high level for a while, even if the individual is lying or sitting. In a previous study [25] we showed that, by means of structural feature extraction, vital signs can be exploited to improve recognition accuracy. Now, in order to measure physiological signals, additional sensors would be required, thereby increasing the system cost and introducing obtrusiveness [18]. Also, these sensors generally use wireless communication which entails higher energy expenditures.

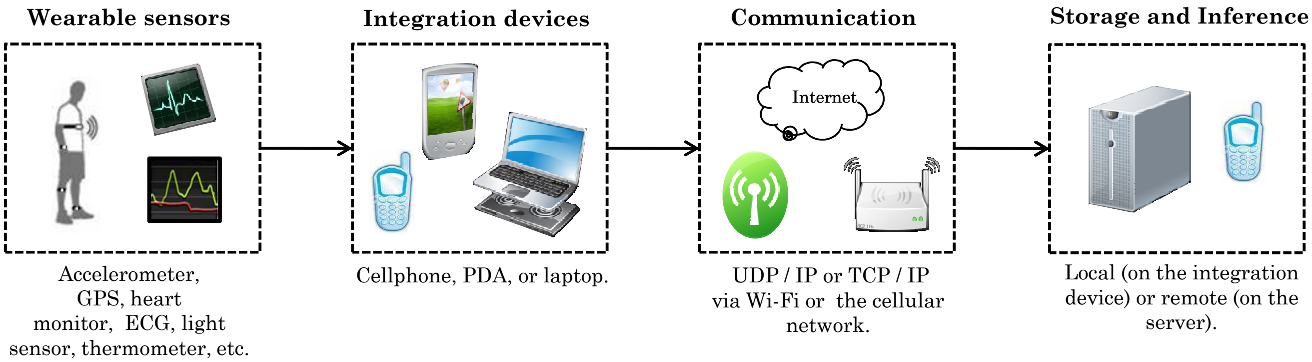

Fig. 2. Generic data acquisition architecture for Human Activity Recognition.

B. Obtrusiveness

To be successful in practice, HAR systems should not require the user to wear many sensors nor interact too often with the application. Furthermore, the more sources of data available, the richer the information that can be extracted from the measured attributes. There are systems which require the user to wear four or more accelerometers [22], [26], [44], or carry a heavy rucksack with recording devices [18]. These configurations may be uncomfortable, invasive, expensive, and hence not suitable for activity recognition. Other systems are able to work with rather unobtrusive hardware. For instance, a sensing platform that can be worn as a sport watch is presented in [24]; Centinela [25] only requires a strap that is placed on the chest and a cellular phone. Finally, the systems introduced in [32], [36] recognize activities with a cellular phone only.

C. Data collection protocol

The procedure followed by the individuals while collecting data is critical in any HAR. In 1999, Foerster et al. [5] demonstrated 95.6% of accuracy for ambulation activities in a controlled data collection experiment, but in a natural environment (i.e., outside of the laboratory), the accuracy dropped to 66%! The number of individuals and their physical characteristics are also crucial factors in any HAR study. A comprehensive study should consider a large number of individuals with diverse characteristics in terms of gender, age, height, weight, and health conditions. This is with the purpose of ensuring flexibility to support new users without the need of collecting additional training data.

D. Recognition performance

The performance of a HAR system depends on several aspects, such as (1) the activity set (2) the quality of the training data, (3) the feature extraction method, and (4) the learning algorithm. In the first place, each set of activities brings a totally different pattern recognition problem. For example, discriminating among *walking*, *running*, and *standing still* [45], turns out to be much easier than incorporating more complex activities such as *watching TV*, *eating*, *ascending*, and *descending* [26]. Secondly, there should be a sufficient amount of training data, which should also be similar to the expected testing data. Finally, a comparative evaluation of several learning methods is desirable as each dataset exhibits distinct characteristics that can be either beneficial or detrimental for a particular method. Such interrelationship among datasets and learning methods can be very hard to analyze theoretically, which accentuates the need of an experimental study. In order to quantitatively understand the recognition performance, some standard metrics are used, e.g., *accuracy, recall, precision, F-measure, Kappa statistic,* and *ROC curves*. These metrics will be discussed in Section IV.

E. Energy consumption

Context-aware applications rely on mobile devices —such as sensors and cellular phones— which are generally energy constrained. In most scenarios, extending the battery life is a desirable feature, especially for medical and military applications that are compelled to deliver critical information. Surprisingly, most HAR schemes do not formally analyze energy expenditures, which are mainly due to processing, communication, and visualization tasks. Communication is often the most expensive operation, so the designer should minimize the amount of transmitted data. In most cases, short range wireless networks (e.g., Bluetooth or Wi-Fi) should be preferred over long range networks (e.g., cellular network or WiMAX) as the former require lower power. Some typical energy saving mechanisms are *data aggregation* and *compression* yet they involve additional computations that may affect the application performance. Another approach is to carry out feature extraction and classification in the integration device, so that raw signals would not have to be continuously sent to the server [30], [46]. This will be discussed in Section III-F. Finally, *since all sensors may not be necessary simultaneously, turning some of them off* or reducing their sampling/transmission rate is very convenient to save energy. For example, if the user's activity is *sitting* or *standing still*, the GPS sensor may be turned off [36].

F. Processing

Another important point of discussion is where the recognition task should be done, whether in the server or in the integration device. On one hand, a server is expected to have huge processing, storage, and energy capabilities, allowing to incorporate more complex methods and models. On the other hand, a HAR system running on a mobile device should substantially reduce energy expenditures, as raw data would not have to be continuously sent to a server for processing. The system would also become more robust and responsive because it would not depend on unreliable wireless communication links, which may be unavailable or error prone; this is particularly important for medical or military applications that require real-time decision making. Finally, a mobile HAR system would be more scalable since the server load would be alleviated by the locally performed feature extraction and classification computations. However, implementing activity recognition in mobile devices becomes challenging because they are still constrained in terms of processing, storage, and energy. Hence, feature extraction and learning methods should be carefully chosen to guarantee a reasonable response time and battery life. For instance, classification algorithms such as Instance Based Learning [31] and Bagging [47] are very expensive in their evaluation phase, which makes them not convenient for HAR.

G. Flexibility

There is an open debate on the design of any activity recognition model. Some authors claim that, as people perform activities in a different manner (due to age, gender, weight, and so on), a *specific* recognition model should be built for each individual [29]. This implies that the system should be retrained for each new user. Other studies [25] rather emphasize the need of a *monolithic* recognition model, flexible enough to work with different users. Consequently, two types of analyses have been proposed to evaluate activity recognition systems: *subject-dependent* and *subject-independent* evaluations [22]. In the first one, a classifier is trained and tested for each individual with his/her own data and the average accuracy for all subjects is computed. In the second one, only one classifier is built for all individuals using cross validation or leave-one-individual-out analysis. It is worth to highlight that, in some cases, it would not be convenient to train the system for each new user, especially when (1) there are too many activities; (2) some activities are not desirable for the subject to carry out (e.g., falling downstairs); or (3) the subject would not cooperate with the data collection process (e.g., patients with dementia and other mental pathologies). On the other hand, an elderly lady would surely walk quite differently than a ten-years-old boy, thereby challenging a single model to recognize activities regardless of the subject's characteristics. A solution to the dichotomy of the monolithic vs. particular recognition model can be addressed by creating groups of users with similar characteristics. Additional design considerations related to this matter will be discussed in Section VI.

IV. Activity Recognition Methods

In Section II, we have seen that, to enable the recognition of human activities, raw data have to first pass through the process of feature extraction. Then, the recognition model is built from the set of feature instances by means of machine learning techniques. Once the model is trained, unseen instances (i.e., time windows) can be evaluated in the recognition model, yielding a prediction on the performed activity. Next, the most noticeable approaches in *feature extraction* and *learning* will be covered.

A. Feature extraction

Human activities are performed during relatively long periods of time (in the order of seconds or minutes) compared to the sensors' sampling rate (which can be up to 250 Hz). Besides, a single sample on a specific time instant (e.g., the Y-axis acceleration is 2.5g, or the heart rate is 130 bpm) does not provide sufficient information to describe the performed activity. Thus, activities need to be recognized in a time window basis rather than in a sample basis. Now, the question is: how do we compare two given time windows? It would be nearly impossible for the signals to be exactly identical, even if they come from the same subject performing the same activity. This is the main motivation for applying *feature extraction* (FE) methodologies to each time window: filtering relevant information and obtaining quantitative measures that allow signals to be compared.

In general, two approaches have been proposed to extract features from time series data: *statistical* and *structural* [48]. Statistical methods, such as the Fourier transform and the Wavelet transform, use quantitative characteristics of the data to extract features, whereas structural approaches take into account the interrelationship among data. The criterion to choose either of these methods is certainly subject to the nature of the given signal.

Figure 3 displays the process to transform the raw time series dataset —which can be from acceleration, environmental variables, or vital signs— into a set of *feature vectors*. Each instance in the processed dataset corresponds to the feature vector extracted from all the signals within a time window. Most of the approaches surveyed in this paper adhere to this mapping.

Next, we will cover the most common FE techniques for each of the measured attributes, i.e., acceleration, environmental signals, and vital signs. GPS data are not considered in this section since they are mostly used to compute the speed [18], [36] or include some knowledge about the place where the activity is being performed [30].

1) Acceleration: Acceleration signals (see Figure 4) are highly fluctuating and oscillatory, which makes it difficult to recognize the underlying patterns using their raw values. Existing HAR systems based on accelerometer data employ statistical feature extraction and, in most of the cases, either time- or frequency-domain features. Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA) have also been applied with promising results [34], as well as autoregressive model coefficients [28]. All these techniques are conceived to handle the high variability inherent to acceleration signals. Table II summarizes the feature extraction

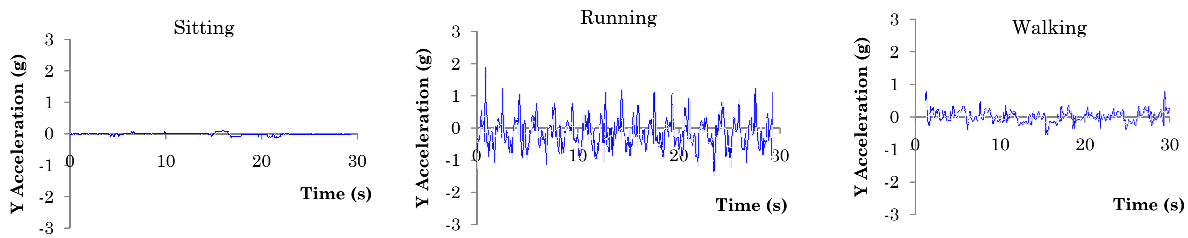

Fig. 4. Acceleration signals for different activities.

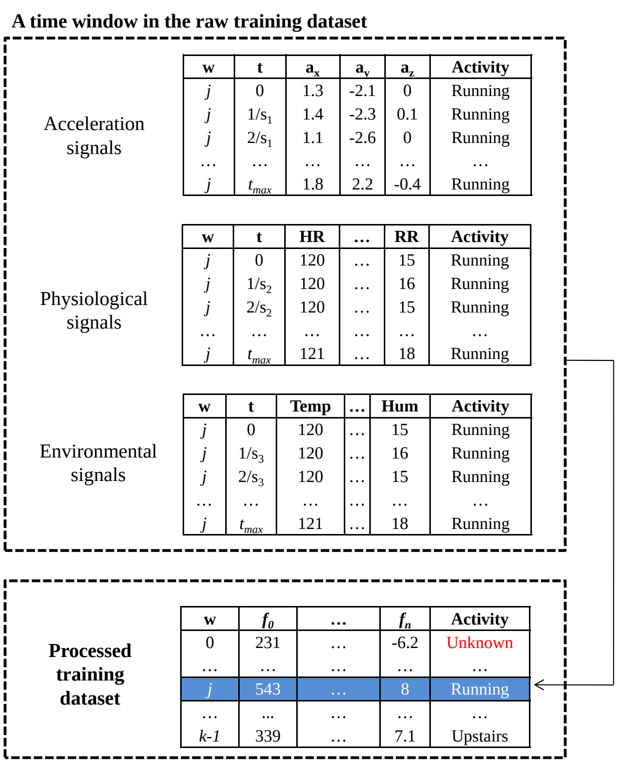

Fig. 3. An example of the mapping from the raw dataset to the feature dataset. w is the window consecutive; s_i is the sampling rate for the group of sensors i, where sensors in the same group have the same sampling rate; f_i is each of the extracted features. Each instance in the processed dataset corresponds to the set of features computed from an entire window in the raw dataset.

methods for acceleration signals. The definition of some of the most widely used features [35] are listed below for a given signal $Y = \{y_1, ..., y_n\}$.

- Central tendency measures such as the *arithmetic mean* $\bar{y}$ and the *root mean square* (RMS) (Equations 1 and 2).
- Dispersion metrics such as the *standard deviation* σ_y, the *variance* σ_y^2, and the *mean absolute deviation* (MAD) (Equations 3, 4, and 5).
- Domain transform measures such as the *energy*, where F_i is the i-th component of the Fourier Transform of Y (Equation 6).

TABLE II
SUMMARY OF FEATURE EXTRACTION METHODS FOR ACCELERATION SIGNALS.

Group	Methods
Time domain	Mean, standard deviation, variance, interquartile range (IQR), mean absolute deviation (MAD), correlation between axes, entropy, and kurtosis [18], [22]–[24], [26], [35], [44].
Frequency domain	Fourier Transform (FT) [26], [35] and Discrete Cosine Transform (DCT) [49].
Others	Principal Component Analysis (PCA) [34], [49], Linear Discriminant Analysis (LDA) [35], Autoregresive Model (AR), and HAAR filters [27].

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (1)$$

$$RMS(Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {y_i}^2} \quad (2)$$

$$\sigma_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (3)$$

$$\sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2 \quad (4)$$

$$MAD(Y) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} |y_i - \bar{y}|} \quad (5)$$

$$Energy(Y) = \frac{\sum_{i=1}^{n} {F_i}^2}{n} \quad (6)$$

2) Environmental variables: Environmental attributes, along with acceleration signals, have been used to enrich context awareness. For instance, the values from air pressure and light intensity are helpful to determine whether the individual is outdoors or indoors [50]. Also, audio signals are useful to conclude that the user is having a conversation rather than listening to music [18]. Table III summarizes the feature extraction methods for environmental attributes.

TABLE III
SUMMARY OF FEATURE EXTRACTION METHODS FOR ENVIRONMENTAL VARIABLES.

Attribute	Features
Altitude	Time-domain [18]
Audio	Speech recognizer [51]
Barometric pressure	Time-doimain and frequency-domain [50]
Humidity	Time-domain [18]
Light	Time-domain [52] and frequency-domain [18]
Temperature	Time-domain [18]

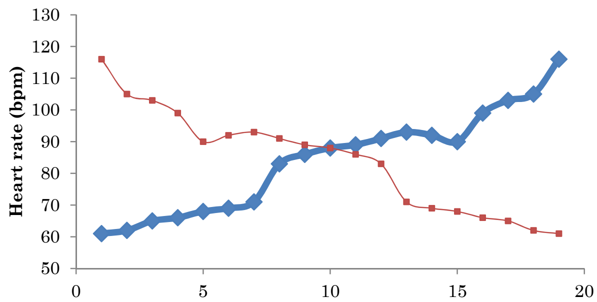

Fig. 5. Heart rate signal for *walking* (bold) and flipped signal (thin).

3) Vital signs: The very first works that explored vital sign data with the aim of recognizing human activities applied statistical feature extraction. In [22], the authors computed the number of heart beats above the resting heart rate value as the only feature. Instead, Parkka et al. [18] calculated time domain features for heart rate, respiration effort, SaO_2, ECG, and skin temperature. Nevertheless, the signal's shape is not described by these features. Consider the situation shown in Figure 5. A heart rate signal $S(t)$ for an individual that was walking is shown with a bold line and the same signal in reverse temporal order, $S'(t)$, is displayed with a thin line. Notice that most time domain and frequency domain features (e.g., mean, variance, and energy) are identical for both signals while they may represent different activities. This is the main motivation for applying structural feature extraction.

In a time series context, *structure detectors* are intended to describe the morphological interrelationship among data. Given a time series $Y(t)$, a structure detector implements a function $f(Y(t)) = \hat{Y}(t)$ such that $\hat{Y}(t)$ represents the structure of $Y(t)$ as an approximation. In order to measure the goodness of fit of $\hat{Y}(t)$ to $Y(t)$, the sum of squared errors (SSE) is calculated as follows:

$$SSE = \sum_t \left(Y(t) - \hat{Y}(t)\right)^2 \tag{7}$$

The extracted features are the parameters of $\hat{Y}(t)$ which, of course, depend on the nature of the function. Table IV contains some typical functions implemented by structure detectors. The authors found that polynomials are the functions that best fit physiological signals such as heart rate, respiration rate, breath amplitude, and skin temperature [25]. The degree of the polynomial should be chosen based upon the number of

TABLE IV
COMMON FUNCTIONS IMPLEMENTED BY STRUCTURE DETECTORS.

Function	Equation	Parameters
Linear	$F(t) = mt + b$	$\{m, b\}$
Polynomial	$F(t) = a_0 + a_1 t + ... + a_{n-1} t^{n-1}$	$\{a_0, ..., a_{n-1}\}$
Exponential	$F(t) = a\|b\|^t + c$	$\{a, b, c\}$
Sinusoidal	$F(t) = a * sin(t + b) + c$	$\{a, b, c\}$

samples to avoid overfitting due to the Runge's phenomenon [53].

4) Selection of the window length: In accordance to Definition 2, dividing the measured time series in time windows is a convenient solution to relax the HAR problem. A key factor is, therefore, the selection of the window length because the computational complexity of any FE method depends on the number of samples. Having rather short windows may enhance the FE performance, but would entail higher overhead due to the recognition algorithm being triggered more frequently. Besides, short time windows may not provide sufficient information to fully describe the performed activity. Conversely, if the windows are too long, there might be more than one activity within a single time window [36]. Different window lengths have been used in the literature: 0.08s [29], 1s [36], 1.5s [54], 3s [55], 5s [49], 7s [56], 12s [25], or up to 30s [22]. Of course, this decision is conditioned to the activities to be recognized and the measured attributes. The heart rate signal, for instance, required 30s time windows in [22]. Instead, for activities such as swallowing, 1.5s time windows were employed.

Time windows can also be either overlapping [22], [25], [26], [57] or disjoint [36], [54], [55], [49]. Overlapping time windows are intended to handle transitions more accurately, although using small non-overlapping time windows, misclassifications due to transitions are negligible.

5) Feature selection: Some features in the processed dataset might contain redundant or irrelevant information that can negatively affect the recognition accuracy. Then, implementing techniques for selecting the most appropriate features is a suggested practice to reduce computations and simplify the learning models. The *Bayesian Information Criterion* (BIC) and the *Minimum Description Length* (MDL) have been widely used for general machine learning problems. In HAR, a common method is the *Minimum Redundancy and Maximum Relevance* (MRMR) [58], utilized in [19]. In that work, the *minimum mutual information between features* is used as criteria for minimum redundancy and the *maximal mutual information between the classes and features* is used as criteria for maximum relevance. In contrast, Maurer et al. [52] applied a *Correlation-based Feature Selection (CFS) approach* [59], taking advantage of the fact that this method is built in WEKA [60]. CFS works under the assumption that features should be highly correlated with the given class but uncorrelated with each other. Iterative approaches have also been evaluated to select features. Since the number of feature subsets is $O(2^n)$, for n features, evaluating all possible subsets is not computationally feasible. Hence, metaheuristic methods such as multiobjective evolutionary algorithms have been employed to explore the space of possible feature subsets [61].

B. Learning

In recent years, the prominent development of sensing devices (e.g., accelerometers, cameras, GPS, etc.) has facilitated the process of collecting attributes related to the individuals and their surroundings. However, most applications require much more than simply gathering measurements from variables of interest. In fact, additional challenges for enabling context awareness involve knowledge discovery since the raw data (e.g., acceleration signals or electrocardiogram) provided by the sensors are often useless. For this purpose, HAR systems make use of *machine learning* tools, which are helpful to build patterns to describe, analyze, and predict data.

In a machine learning context, patterns are to be discovered from a set of given examples or observations denominated *instances*. Such input set is called *training set*. In our specific case, each instance is a feature vector extracted from signals within a time window. The examples in the training set may or may not be labeled, i.e., associated to a known class (e.g., walking, running, etc.). In some cases, labeling data is not feasible because it may require an expert to manually examine the examples and assign a label based upon their experience. This process is usually tedious, expensive, and time consuming in many data mining applications.

There exist two learning approaches, namely *supervised* and *unsupervised* learning, which deal with labeled and unlabeled data, respectively. Since a human activity recognition system should return a label such as walking, sitting, running, etc., most HAR systems work in a supervised fashion. Indeed, it might be very hard to discriminate activities in a completely unsupervised context. Some other systems work in a semi-supervised fashion allowing part of the data to be unlabeled.

1) Supervised learning: Labeling sensed data from individuals performing different activities is a relatively easy task. Some systems [22], [52] store sensor data in a non-volatile medium while a person from the research team supervises the collection process and manually registers activity labels and time stamps. Other systems feature a mobile application that allows the user to select the activity to be performed from a list [25]. In this way, each sample is matched to an activity label, and then stored in the server.

Supervised learning —referred to as *classification* for discrete-class problems— has been a very productive field, bringing about a great number of algorithms. Table V summarizes the most important classifiers in Human Activity Recognition and their description is included below.

TABLE V
CLASSIFICATION ALGORITHMS USED BY STATE-OF-THE-ART HUMAN ACTIVITY RECOGNITION SYSTEMS.

Type	Classifiers	References
Decision tree	C4.5, ID3	[19], [24], [26], [44]
Bayesian	Naïve Bayes and Bayesian Networks	[19], [22], [26], [46]
Instance Based	k-nearest neighbors	[19], [24]
Neural Networks	Multilayer Perceptron	[62]
Domain transform	Support Vector Machines	[28], [33], [34]
Fuzzy Logic	Fuzzy Basis Function, Fuzzy Inference System	[23], [35], [63]
Regression methods	MLR, ALR	[25], [30]
Markov models	Hidden Markov Models, Conditional Random Fields	[64], [65]
Classifier ensembles	Boosting and Bagging	[25], [55]

- *Decision trees* build a hierarchical model in which attributes are mapped to nodes and edges represent the possible attribute values. Each branch from the root to a leaf node is a classification rule. C4.5 is perhaps the most widely used decision tree classifier and is based on the concept of information gain to select which attributes should be placed in the top nodes [66]. Decision trees can be evaluated in $O(\log n)$ for n attributes, and usually generate models that are easy to understand by humans.
- *Bayesian* methods calculate posterior probabilities for each class using estimated conditional probabilities from the training set. The *Bayesian Network* (BN) [67] classifier and *Naïve Bayes* (NB) [68] —which is a specific case of BN— are the principal exponents of this family of classifiers. A key issue in Bayesian Networks is the topology construction, as it is necessary to make assumptions on the independence among features. For instance, the NB classifier assumes that all features are conditionally independent given a class value, yet such assumption does not hold in many cases. As a matter of fact, acceleration signals are highly correlated, as well as physiological signals such as heart rate, respiration rate, and ECG amplitude.
- *Instance based learning* (IBL) [47] methods classify an instance based upon the most *similar* instance(s) in the training set. For that purpose, they define a distance function to measure similarity between each pair of instances. This makes IBL classifiers quite expensive in their evaluation phase as each new instance to be classified needs to be compared to the entire training set. Such high cost in terms of computation and storage, makes IBL models not convenient to be implemented in a mobile device.
- *Support Vector Machines* (SVM) [69] and *Artificial Neural Networks* (ANN) [70] have also been broadly used in HAR although they do not provide a set of rules understandable by humans. Instead, knowledge is hidden within the model, which may hinder the analysis and incorporation of additional reasoning. SVMs rely on kernel functions that project all instances to a higher dimensional space with the aim of finding a linear decision boundary (i.e., a hyperplane) to partition the data. Neural networks replicate the behavior of biological neurons in the human brain, propagating activation signals and encoding knowledge in the network links. Besides, ANNs have been shown to be universal function approximators. The

high computational cost and the need for large amount of training data are two common drawbacks of neural networks.
- *Ensembles of classifiers* combine the output of several classifiers to improve classification accuracy. Some examples are *bagging*, *boosting*, and *stacking*. Classifier ensembles are clearly more expensive, computationally speaking, as they require several models to be trained and evaluated.

2) Semi-supervised learning: Relatively few approaches have implemented activity recognition in a semi-supervised fashion, thus, having part of the data without labels [71]–[75]. In practice, annotating data might be difficult in some scenarios, particularly when the granularity of the activities is very high or the user is not willing to cooperate with the collection process. Since semi-supervised learning is a minority in HAR, there are no standard algorithms or methods, but each system implements its own approach. Section V-C provides more details on the state-of-the-art semi-supervised activity recognition approaches.

3) Evaluation metrics: In general, the selection of the classification algorithm for HAR has been merely supported by empirical evidence. The vast majority of the studies use cross validation with statistical tests to compare classifiers' performance for a particular dataset. The classification results for a particular method can be organized in a *confusion matrix* $M_{n\times n}$ for a classification problem with n classes. This is a matrix such that the element M_{ij} is the number of instances from class i that were actually classified as class j. The following values can be obtained from the confusion matrix in a binary classification problem:

- *True Positives* (TP): The number of positive instances that were classified as positive.
- *True Negatives* (TN): The number of negative instances that were classified as negative.
- *False Positives* (FP): The number of negative instances that were classified as positive.
- *False Negatives* (FN): The number of positive instances that were classified as negative.

The *accuracy* is the most standard metric to summarize the overall classification performance for all classes and it is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

The *precision*, often referred to as *positive predictive value*, is the ratio of correctly classified positive instances to the total number of instances classified as positive:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

The *recall*, also called *true positive rate*, is the ratio of correctly classified positive instances to the total number of positive instances:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

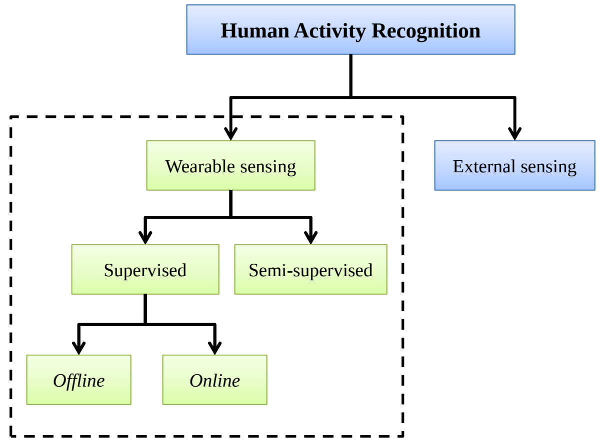

Fig. 6. Taxonomy of Human Activity Recognition Systems.

The *F-measure* combines precision and recall in a single value:

$$F - measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{11}$$

Although defined for binary classification, these metrics can be generalized for a problem with n classes. In such case, an instance could be positive or negative according to a particular class, e.g., positives might be all instances of *running* while negatives would be all instances other than *running*.

4) Machine learning tools: The Waikato Environment for Knowledge Analysis (WEKA) [60] is certainly the best known tool in the machine learning research community. It contains implementations of a number of learning algorithms and it allows to easily evaluate them for a particular dataset using cross validation and random split, among others. WEKA also offers a Java API that facilitates the incorporation of new learning algorithms and evaluation methodologies on top of the pre-existing framework. One of the limitations of current Machine Learning APIs such as WEKA [60] and the Java Data Mining (JDM) platform [76] is that they are not supported by current mobile platforms. In that direction, the authors proposed MECLA [46], a mobile platform for the evaluation of classification algorithms under the Android platform.

V. Evaluation of HAR Systems

In this survey, we have categorized HAR systems that rely on wearable sensors in two levels. The first one has to do with the learning approach, which can be either *supervised* or *semi-supervised*. In the second level, according to the response time, supervised approaches can be either *online* or *offline*. To the best of our knowledge, current semi-supervised systems have been implemented and evaluated offline. Figure 6 summarizes the proposed taxonomy.

On one hand, online schemes provide immediate feedback on the performed activities. On the other hand, offline approaches either need more time to recognize activities due to high computational demands, or are intended for applications that do not require real-time feedback. These two classes of

HAR systems have very different purposes and associated challenges, so they are evaluated separately. The qualitative evaluation encompasses the following aspects:

- Recognized activities (Table I)
- Type of sensors and the measured attributes (Section III-A)
- Integration device
- Level of obtrusiveness, which could be *low*, *medium*, or *high*.
- Type of data collection protocol, which could be either a *controlled* or a *naturalistic* experiment.
- Level of energy consumption, which could be *low*, *medium*, or *high*.
- Classifier flexibility level, which could be either *user-specific* or *monolithic*.
- Feature extraction method(s)
- Learning algorithm(s)
- Overall accuracy for all activities

A. Online HAR systems

Applications of online activity recognition systems can be easily visualized. In healthcare, continuously monitoring patients with physical or mental pathologies becomes crucial for their protection, safety, and recovery. Likewise, interactive games or simulators may enhance user's experience by considering activities and gestures. Table VII summarizes the online state-of-the-art activity recognition approaches. The abbreviations and acronyms are defined in Table VI. The most important works on online HAR are described next.

TABLE VI
LIST OF ABBREVIATIONS AND ACRONYMS.

3DD	3D Deviation of the acceleration signals
ACC	Accelerometers
AMB	Ambulation activities (see Table I)
ANN	Artificial Neural Network
ALR	Additive Logistic Regression classifier
AR	Autoregressive Model Coefficients
AV	Angular velocity
BN	Bayesian Network Classifier
CART	Classification And Regression Tree
DA	Daily activities (see Table I)
DCT	Discrete Cosine Transform
DT	Decision Tree-Based Classifier
DTW	Dynamic Time Warping
ENV	Environmental sensors
FBF	Fuzzy Basis Function
FD	Frequency-domain features
GYR	Gyroscope
HMM	Hidden Markov Models
HRB	Heart Rate Beats above the resting heart rate
HRM	Heart Rate Monitor
HW	Housework activities (see Table I)
KNN	k-Nearest Neighbors classifier
LAB	Laboratory controlled experiment
LDA	Linear Discriminant Analysis
LS	Least Squares algorithm
MIL	Military activities
MNL	Monolithic classifier (subject independent)
NAT	Naturalistic experiment
NB	The Naïve Bayes classifier
NDDF	Normal Density Discriminant Function
N/S	Not Specified
PCA	Principal Component Analysis
PHO	Activities related to phone usage (see Table I)
PR	Polynomial Regression
RFIS	Recurrent Fuzzy Inference System
SD	Subject Dependent evaluation
SFFS	Sequential Forward Feature Selection
SI	Subject Independent evaluation
SMA	Signal Magnitude Area
SMCRF	Semi-Markovian Conditional Random Field
SPC	User-specific classifier (subject dependent)
SPI	Spiroergometry
TA	Tilt Angle
TD	Time-domain features
TF	Transient Features [25]
TR	Transitions between activities
UB	Upper body activities (see Table I)
VS	Vital sign sensors

1) eWatch: Maurer et al. [24] introduced *eWatch* as an online activity recognition system which embeds sensors and a microcontroller within a device that can be worn as a sport watch. Four sensors are included, namely an accelerometer, a light sensor, a thermometer, and a microphone. These are passive sensors and, as they are embedded in the device, no wireless communication is needed; thus, eWatch is very energy efficient. Using a C4.5 decision tree and time-domain feature extraction, the overall accuracy was up to 92.5% for six ambulation activities, although they achieved less than 70% for activities such as *descending* and *ascending*. The execution time for feature extraction and classification is less than 0.3 ms, which makes the system very responsive. However, in eWatch, data were collected under controlled conditions, i.e., a lead experimenter supervised and gave specific guidelines to the subjects on how to perform the activities [24]. In Section III-C, we reviewed the disadvantages of this approach.

2) Vigilante: The authors proposed Vigilante [46], a mobile application for real-time human activity recognition under the Android platform. The Zephyr's BioHarness BT [77] chest sensor strap was used to measure acceleration and physiological signals such as heart rate, respiration rate, breath waveform amplitude, and skin temperature, among others. In that work, we proposed MECLA, a library for the mobile evaluation of classification algorithms, which can also be utilized in further pattern recognition applications. Statistical time- and frequency-domain features were extracted from acceleration signals while polynomial regression was applied to physiological signals via *least squares*. The C4.5 decision tree classifier recognized three ambulation activities with an overall accuracy of 92.6%. The application can run for up to 12.5 continuous hours with a response time of no more than 8% of the window length. Different users with diverse characteristics participated in training and testing phases, ensuring flexibility to support new users without the need to re-train the system. Unlike other approaches, Vigilante was evaluated completely online to provide more realistic results. Vigilante is moderately energy efficient because it requires permanent Bluetooth communication between the sensor strap and the phone.

3) Tapia et al.: This system recognizes 17 ambulation and gymnasium activities such as *lifting weights*, *rowing*, *doing push ups*, etc., with different intensities (a total of 30

TABLE VII
SUMMARY OF STATE-OF-THE-ART IN ONLINE HUMAN ACTIVITY RECOGNITION SYSTEMS.

Reference	**Activities**	**Sensors**	**ID**	**Obtrusive**	**Experiment**	**Energy**	**Flexibility**	**Processing**	**Features**	**Learning**	**Accuracy**
Ermes [44]	AMB (5)	ACC (wrist, ankle, chest)	PDA	High	N/S	High	SPC	High	TD, FD	DT	94%
eWatch [24]	AMB (6)	ACC, ENV (wrist)	Custom	Low	LAB	Low	MNL	Low	TD, FD	C4.5, NB	94%
Tapia [22]	EXR (30)	ACC (5 places), HRM	Laptop	High	LAB	High	Both	High	TD, FD, HB	C4.5, NB	86% (SD), 56% (SI)
Vigilante [46]	AMB (3)	ACC and VS (chest)	Phone	Medium	NAT	Medium	MNL	Low	TD, FD, PR, TF	C4.5	92.6%
Kao [23]	AMB, DA (7)	ACC (wrist)	Custom	Low	N/S	Medium	MNL	Low	TD, LDA	FBF	94.71%
Brezmes [31]	AMB (5)	ACC (phone)	Phone	Low	N/S	Low	SPC	High	TD, FD	KNN	80%
COSAR [30]	AMB, DA (10)	ACC (watch, phone), GPS	Phone	Low	NAT	Medium	MNL	Mediun	TD	COSAR	93%
ActiServ [29], [32]	AMB, PHO (11)	ACC (phone)	Phone	Low	N/S	Low	SPC	High	$\bar{y}, \sigma_y^2$	RFIS	71% - 98%

activities). A comprehensive study was carried out, including 21 participants and both subject-dependent and subject-independent studies. The average classification accuracy was reported to be 94.6% for subject-dependent analysis whereas a 56% of accuracy was reached in the subject-independent evaluation. If intensities are not considered, the overall subject-independent accuracy is 80.6%. This system works with very obtrusive hardware, i.e., five accelerometers were placed on the user's dominant arm and wrist, hip, thigh, and ankle, as well as a heart rate monitor on the chest. Besides, all these sensors require wireless communication, involving high energy consumption. Finally, the integration device is a laptop, which allows for better processing capabilities, but prevents portability and pervasiveness.

4) ActiServ: In 2010, Berchtold et al. introduced *ActiServ* as an activity recognition service for mobile phones [29], [32]. The system was implemented on the Neo FreeRunner phone. They make use of a fuzzy inference system to classify ambulation and phone activities based on the signals given by the phone's accelerometer only. This makes ActiServ a very energy efficient and portable system. The overall accuracy varies between 71% and 97%. However, in order to reach the top accuracy level, the system requires a runtime duration in the order of days! When the algorithms are executed to meet a real-time response time, the accuracy drops to 71%. ActiServ can also reach up to 90% after personalization, in other words, a subject-dependent analysis. From the reported confusion matrices, the activity labeled as *walking* was often confused with *cycling*, while *standing* and *sitting* could not be differentiated when the cellphone's orientation was changed.

5) COSAR: Riboni et al. [30] presented COSAR, a framework for context-aware activity recognition using statistical and ontological reasoning under the Android platform. The system recognizes ambulation activities as well as *brushing teeth*, *strolling*, and *writing on a blackboard*. COSAR gathers data from two accelerometers, one in the phone and another on the individual's wrist, as well as from the cellphone's GPS. Since COSAR makes use of the GPS sensor, it was catalogued as a moderately energy efficient system. COSAR uses an interesting concept of *potential activity matrix* to filter activities based upon the user's location. For instance, if the individual is in the *kitchen*, he or she is probably not *cycling*. Another contribution is the statistical classification of activities with a historical variant. For example, if the predictions for the last five time windows were {*jogging*, *jogging*, *walking*, *jogging*, *jogging*}, the third window was likely a misclassification (e.g., due to the user performing some atypical movement) and the algorithm should automatically correct it. The overall accuracy was roughly 93% though, in some cases, *standing still* was confused with *writing on a blackboard*, as well as *hiking up* with *hiking down*.

6) Kao: Kao et al. [23] presented a portable device for online activity detection. A triaxal accelerometer is placed on the user's dominant wrist, sampling at 100 Hz. They apply time domain features and the *Linear Discriminant Analysis (LDA)* to reduce the dimension of the feature space. Then, a *Fuzzy Basis Function* learner —which uses fuzzy *If-Then rules*— classifies the activities. An overall accuracy of 94.71% was reached for seven activities: *brushing teeth*, *hitting*, *knocking*, *working at a PC*, *running*, *walking*, and *swinging*. The system reports an average response time of less than 10 ms, which support its feasibility. All the computations are done in an embedded system that should be carried by the user as an additional device. This has some disadvantages with respect to a mobile phone in terms of portability, comfort, and cost. Moreover, the size of the time window was chosen to be 160 ms. Given the nature of the recognized activities, this excessive granularity causes accidental movements when *swinging* or *knocking* to be confused with *running*, for instance. Such a small window length also (1) induces more overhead due to the classification algorithm being triggered very often and (2) is not beneficial for the feature extraction performance as time domain features require $O(n)$ computations.

7) Other approaches: The system proposed by Brezmes et al. [31] features a mobile application for HAR under the Nokia platform. They used the k-nearest neighbors classifier, which is computationally expensive and not scalable for mobile phones as it needs the entire training set —which can be fairly large— to be stored in the device. Besides, their system requires each new user to collect additional training data in order to obtain accurate results. Ermes et al. [44] developed an online system that reaches 94% overall average accuracy but they only applied a subject-dependent evaluation. Besides, their data were collected from only three subjects, which inhibits flexibility to support new users.

8) Discussion: As we have seen, each online HAR system surveyed has its own benefits and drawbacks. Thus, the selection of a particular approach for a real case study depends on the application requirements. If portability and obtrusiveness are the key issues, eWatch would be an appropriate option for ambulation activities. But if a broader set of activities needs to be recognized, COSAR should be considered, although it entails higher energy expenditures due to Bluetooth communication and the GPS sensor. The system proposed by Tapia et al. would be a better choice to monitor exercise habits yet it may be too obtrusive. Overall, most systems exhibit similar accuracy levels (more than 92%) but since each system works with a specific dataset and activity set, there is no significant evidence to argue that a system is more accurate than the others. Vigilante is the only approach that collects vital sign information, which opens a broader spectrum of applications for healthcare purposes. In addition, COSAR and Vigilante work under the Android platform —reported as best-selling smartphone platform in 2010 by Canalys [78]— facilitating the deployment in current cellular phones. The system cost is also an important aspect, especially when the application's aim is being scaled to hundreds or thousands of users. Vigilante, COSAR, eWatch, and the work of Kao et al. require specialized hardware such as sensors and embedded computers whereas ActiServ and the system proposed by Brezmes et al. only need a conventional cellular phone.

B. Supervised Offline Systems

There are cases in which the user does not need to receive immediate feedback. For example, applications that analyze exercise and diet habits in patients with heart disease, diabetes, or obesity, or applications that estimate the number of calories burned after an exercise routine [79], [80] can work on an offline basis. Another example of an offline HAR system is an application to discover commercial patterns for advertisement. For instance, if an individual performs exercise activities very frequently, they could be advertised on sport wear items. In all these cases, gathered data can be analyzed on a daily —or even weekly— basis to draw conclusions on the person's behavior. Table VIII summarizes state-of-the-art works in supervised offline human activity recognition based on wearable sensors. The most relevant approaches are described next.

1) Parkka: The work of Parkka et al. [18] considers seven activities: *lying, rowing, riding a bike, standing still, running, walking*, and *Nordic walking*. Twenty two signals were measured, including acceleration, vital signs, and environmental variables. This requires a number of sensors on the individual's chest, wrist, finger, forehead, shoulder, upper back, and armpit. The integration device is a compact computer placed in a 5 kg rucksack. Therefore, we have classified this system as highly obtrusive. Time- and frequency-domain features were extracted from most signals while a speech recognizer [51] was applied to the audio signal. This entails not only high processing demands but also privacy issues due to continuous recording of the user's speech. Three classification methods were evaluated, namely an *automatically generated decision tree*, a *custom decision tree* which introduces domain knowledge and visual inspection of the signals, as well as an *artificial neural network*. The results indicate that the best accuracy was 86%, given by the first method, though activities such as as *rowing*, *walking*, and *Nordic walking* were not accurately discriminated. Parkka et al. mentioned that one of the causes of such misclassification is the lack of synchronization between the activity performances and annotations.

2) Bao: With more than 700 citations [81], the work of Bao and Intelle in 2004 [26] brought significant contributions to the field of activity recognition. The system recognizes 20 activities, including ambulation and daily activities such as *scrubbing, vacuuming, watching TV*, and *working at the PC*. All the data were labeled by the user in a naturalistic environment. Five bi-axial accelerometers were initially placed on the user's knee, ankle, arm, and hip, yet they concluded that with only two accelerometers —on the hip and wrist— the recognition accuracy is not significantly diminished (in about a 5%). Using time- and frequency-domain features along with the C4.5 decision tree classifier, the overall accuracy was 84%. Ambulation activities were recognized very accurately (with up to 95% of accuracy) but activities such as *stretching*, *scrubbing*, *riding escalator* and *riding elevator* were often confused. The inclusion of location information is suggested to overcome this issues. Such idea was later adopted and implemented by other systems [30].

3) Khan: The system proposed by Khan et al. [38] not only recognizes ambulation activities, but also transitions among them, e.g., *sitting to walking*, *sitting to lying*, and so forth. An accelerometer was placed on the individual's chest, sampling at 20Hz, and sending data to a computer via Bluetooth for storage. Three groups of features were extracted from the acceleration signals: (1) *autoregressive model coefficients*, (2) the *Tilt Angle* (TA), defined as the angle between the positive Z-axis and the gravitational vector g, as well as the (3) *Signal Magnitude Area* (SMA), which is the summation of the absolute values of all three signals. Linear Discriminant Analysis is used to reduce the dimensionality of the feature vector and an Artificial Neural Network classified activities and transitions with a 97.76% subject independent accuracy. The results indicate that the TA plays a key role in the improvement of the recognition accuracy. This is expected because the sensor inclination values are clearly different for lying and standing, in view of the sensor being placed on the chest.

4) Zhu: The system proposed by Zhu and Sheng [64] uses *Hidden Markov Models* (HMM) to recognize ambulation activities. Two accelerometers, placed on the subject's wrist and waist, are connected to a PDA via serial port. The PDA sends the raw data via Bluetooth to a computer which processes the data. This configuration is obtrusive and uncomfortable as the user has to wear wired links that may interfere with the normal course of activities. The extracted features are the angular velocity and the 3D deviation of the acceleration signals. The classification of activities operates in two stages. In the first place, an Artificial Neural Network discriminates among stationary (e.g, sitting and standing) and non-stationary activities (e.g., walking and running). Then, a HMM receives the ANN's output and generates a specific activity prediction. An important issue related to this system is that all the data were collected from one single individual, which does not permit to draw strong conclusions on the system performance.

5) Centinela: The authors proposed Centinela, a system that combines acceleration data with vital signs to achieve accurate activity recognition. Centinela recognizes five ambulation activities and includes a portable and unobtrusive real-time data collection platform, which only requires a single sensing device and a mobile phone. Time- and frequency-domain features are extracted from acceleration signals while polynomial regression and transient features [25] are applied to physiological signals. After evaluating eight different classifiers and three different time window sizes, Centinela achieves up to 95.7% overall accuracy. The results also indicate that physiological signals are useful to discriminate among certain activities. Indeed, Centinela achieves 100% accuracy for activities such as *running* and *sitting*, and slightly improves the classification accuracy for *ascending* compared to the cases that utilize acceleration data only. However, Centinela relies on classifier ensembles which entail higher computational costs. And, the presented mobile application runs under the J2ME platform [82] which is falling into disuse. To overcome these issues, the authors proposed Vigilante (see Section V-A2).

TABLE VIII
SUMMARY OF STATE-OF-THE-ART IN OFFLINE HUMAN ACTIVITY RECOGNITION SYSTEMS.

Reference	**Activities**	**Sensors**	**Obtrusive**	**ID**	**Experi-ment**	**Flexi-bility**	**Features**	**Learning**	**Accuracy**
Bao [26]	AMB, DA (20)	ACC (wrist, ankle, thigh, elbow, hip)	High	None	NAT	MNL	TD, FD	KNN, C4.5, NB	84%
Hanai [27]	AMB (5)	ACC (chest)	Low	Laptop	N/S	MNL	HAAR filters	C4.5	93.91%
Parkka [18]	AMB, DA (9)	ACC, ENV, VS (22 signals)	High	PC	NAT	MNL	TD, FD	DR, KNN	86%
He [28]	AMB (4)	ACC	Low	PC	N/S	MNL	AR	SVM	92.25%
He [34]	AMB (4)	ACC (trousers pocket)	Low	PC	N/S	MNL	DCT, PCA	SVM	97.51%
Zhu [64]	AMB, TR (12)	ACC (wrist, waist)	High	PC	N/S	SPC	AV, 3DD	HMM	90%
Altun [49]	AMB (19)	ACC, GYR (chest, arms, legs)	High	None	NAT	MNL	PCA, SFFS	BN, LS, KNN, DTW, ANN	87% - 99%
Cheng [54]	UB (11)	Electrodes (neck, chest, leg, wrist)	High	PC	LAB	MNL	TD	LDA	77% %
McGlynn [56]	DA (5)	ACC (thigh, hip, wrist)	Low	None	N/S	SPC	DTW	DTW ensemble	84.3%
Pham [57]	AMB, DA (4)	ACC (jacket)	Medium	N/S	N/S	Both	Relative Energy	NB, HMM	97% (SD), 95% (SI)
Vinh [65]	AMB, DA (21)	ACC (wrist, hip)	Medium	N/S	N/S	N/S	TD	SMCRF	88.38%
Centinela [25]	AMB (5)	ACC and VS (chest)	Medium	Cellphone	NAT	MNL	TD, FD, PR, TF	ALR, Bagging, C4.5, NB, BN	95.7%
Khan [38]	AMB, TR (15)	ACC (chest)	Medium	Computer	NAT	MNL	AR, SMA, TA, LDA	ANN	97.9%
Jatoba [19]	AMB (6)	ACC, SPI	High	Tablet	LAB	Both	TD / FD	CART, KNN	86% (SI), 95% (SD)
Chen [35]	AMB, DA, HW (8)	ACC (2 wrists)	Medium	N/S	LAB	MNL	TD, FD	FBF	93%
Minnen [55]	AMB, MIL (14)	ACC (6 places)	High	Laptop	Both	SPC	TD, FD	Boosting	90%

6) Other approaches: In 2002, Randel et al. [62] introduced a system to recognize ambulation activities which calculates the *Root Mean Square* (RMS) from acceleration signals and makes use of a *Backpropagation Neural Network* for classification. The overall accuracy was 95% using user-specific training but no details are provided regarding the characteristics of the subjects, the data collection protocol, and the confusion matrix. The system proposed in [27] uses HAAR filters to extract features and the C4.5 algorithm for classification purposes. HAAR filters are intended to reduce the feature extraction computations, compared to traditional TD and FD features. However, the study only collected data from four individuals with unknown physical characteristics, which might be insufficient to provide flexible recognition of activities on new users. He et al. [28], [33], [34] achieved up to 97% of accuracy but only considered four activities: *running*, *being still*, *jumping*, and *walking*. These activities are quite different in nature, which considerably reduces the level of uncertainty thereby enabling higher accuracy. Chen et al. [35] introduces an interesting Dynamic LDA approach to add or remove activity classes and training data online, i.e., the classifier does not have to be re-trained from scratch. With a Fuzzy Basis Function classifier, they reach 93% of accuracy for eight ambulation and daily activities. Nonetheless, all the data were collected inside the laboratory, under controlled conditions. Finally, Vinh et al. [65] use semi-Markovian conditional random fields to recognize not only activities but routines such as *dinner*, *commuting*, *lunch*, and *office*. These routines are composed by sequences of subsets of activities from a total set of 20 activities. Their results indicate 88.38% of accuracy (calculated by the authors from the reported recall tables).

7) Discussion: Unlike online systems, offline HAR are not dramatically affected by processing and storage issues because the required computations could be done in a server with huge computational and storage capabilities. Additionally, we do not elaborate on energy expenditures as a number of systems require neither integration devices nor wireless communication so the application lifetime would only depend on the sensor specifications.

Ambulation activities are recognized very accurately by [25], [38], [83]. These systems place an accelerometer on the subject's chest, which is helpful to avoid ambiguities due to abrupt corporal movements that arise when the sensor is on the wrist or hip [52]. Other daily activities such as *dressing*, *preparing food*, *using the bathroom*, *using the PC*, and *using a phone* are considered in [56]. This introduces additional challenges given that, in reality, an individual could use the phone while walking, sitting, or lying, thereby exhibiting different acceleration patterns. Similarly, in [26], activities such as *eating*, *reading*, *walking*, and *climbing stairs* could happen concurrently yet no analysis is presented to address that matter. Section VI provides insights to the problem of recognizing concurrent activities.

Unobtrusiveness is a desirable feature of any HAR system but having more sensors enables the recognition of a broader set of activities. The scheme presented by Cheng et al. [54] recognizes head movements and activities such as *swallowing*, *chewing*, and *speaking* but requires obtrusive sensors on the throat, chest, and wrist, connected via wired links. In tactical scenarios, this should not be a problem considering that a soldier is accustomed to carry all sort of equipment (e.g., sensors, cameras, weapons, and so forth). Yet, in healthcare applications involving elderly people or patients with heart disease, obtrusive sensors are not convenient.

The studies presented in [19], [35], [54] are based on data collected under controlled conditions while the works in [28], [33], [34], [56], [57], [65] do not specify the data collection procedure. This is a critical issue since a laboratory environment affects the normal development of human activities [5], [26]. The number of subjects also plays a significant role in the validity of any HAR study. In [55], [64] only one individual collected data while in [27], data were collected from four individuals. Collecting data from a small number of people might be insufficient to provide flexible recognition of activities on new users.

C. Semi-supervised approaches

The systems studied so far rely on large amounts of labeled training data. Nonetheless, in some cases, labeling all instances may not be feasible. For instance, to ensure a naturalistic data collection procedure, it is recommended for users to perform activities without the participation of researchers. If the user cannot be trusted or the activities change very often, some labels could be missed. These unlabeled data can still be useful to train a recognition model by means of semi-supervised learning. Some of the most important works in this field are described next.

1) Multi-graphs: Stikic et al. [71], [72] developed a multi-graph-based semi-supervised learning technique which propagates labels through a graph that contains both labeled and unlabeled data. Each node of the graph corresponds to an instance while every edge encodes the similarities between a pair of nodes as a probability value. The topology of the graph is given by the k-nearest neighbors in the feature space. A probability matrix Z is estimated using both Euclidean distance in the feature space and temporal similarity [71]. Once the labels have been propagated throughout the graph (i.e., all instances are labeled), classification is carried out with a Support Vector Machine classifier that relies on a Gaussian radial basis function kernel. The classifier also used the probability matrix Z to introduce knowledge on the level of confidence of each label. The overall accuracy was up to 89.1% and 96.5% after evaluating two public datasets and having labels for only 2.5% of the training data.

2) En-co-training: A well-known method in semi-supervised learning is *co-training*, proposed by Blum and Mitchel in 1998 [84]. This approach requires the training set to have two sufficient and redundant attribute subsets, condition that does not always hold in a HAR context. Guan et al. [74] proposed *en-co-training*, an extension of co-training which does not have the limitations of its predecessor. The system was tested with ten ambulation activities and compared to three other fully supervised classifiers (the k-nearest neighbors, naïve Bayes and a decision tree). The maximum error rate improvement reached by en-co-training was from 17% to 14% —when 90% of the training data were not labeled. If 20% or more of the training data are labeled, the error rate difference between en-co-training and the best fully supervised classifier does not exceed 1.3%.

3) Ali: Ali et al. [73] implemented a Multiple Eigenspaces (MES) technique based on the Principal Component Analysis combined with Hidden Markov Models. The system is designed to recognize finger gestures with a laparoscopic gripper tool. The individuals wore a sensor glove with two bi-axial accelerometers sampling at 50Hz. Five different rotation and translation movements from the individual's hand were recognized with up to 80% of accuracy. This system becomes hard to analyze since no details are provided on the amount of labeled data nor the evaluation procedure.

4) Huynh: Huynh et al. [75] combined Multiple Eigenspaces with Support Vector Machines to recognize eight ambulation and daily activities. Eleven accelerometers were placed on individuals' ankles, knees, elbows, shoulders, wrists, and hip. The amount of labeled training data varied from 5% to 80% and the overall accuracy was between 88% to 64%, respectively. Their approach also outperformed the fully supervised naïve Bayes algorithm, which was used as a baseline. Still, activities such as *shaking hands*, *ascending stairs* and *descending stairs* were often confused.

5) Discussion: The next step in semi-supervised learning HAR would be their implementation online, opening the possibility to use the data collected in production stage —which are unlabeled— to improve the recognition performance. Nevertheless, implementing this approach becomes challenging in terms of computational complexity. This is because most semi-supervised HAR approaches first estimate the labels of all instances in the training set and then apply a conventional supervised learning algorithm. And, the label estimation process is often computationally expensive; for instance, in [71], [72], a graph with one node per instance has to be built. In their experiments, the resulting graphs consisted of up to 16875 nodes, causing the computation of the probability matrix to be highly demanding in regards to processing and storage. Other approaches do not seem to be ready for real scenarios: En-co-training [74] did not report substantial improvement in the classification accuracy. The system proposed by Ali et al. [73] was intended for a very specific purpose but not suitable for recognizing daily activities thereby limiting its applicability to context-aware applications. Finally, the system proposed in [75] required eleven sensors, which introduces high obtrusiveness. Overall, we believe that the field of semi-supervised activity recognition has not reached maturity and needs additional contributions to overcome the aforementioned issues.

VI. Future research considerations

In order to realize the full potential in HAR systems, some topics need further investigation. Next, a list of those topics is included.

- **Activity recognition datasets:** The quantitative comparison of HAR approaches has been hindered by the fact that each system works with a different dataset. While in research areas such as *data mining*, there exist standard datasets to validate the effectiveness of a new method, this is not the case in activity recognition. Each research group collects data from different individuals, uses a different activity set, and utilizes a different evaluation methodology. In that direction, we have included various datasets publicly open to the research community which can be used as benchmarks to evaluate new approaches. Several universities and institutions have published their datasets in [85]–[88]. Another dataset is provided by the 2011 Activity Recognition Challenge [89], in which researchers worldwide were invited to participate.

- **Composite activities:** The activities we have seen so far are quite simple. In fact, many of them could be part of more complex routines or behaviors. Imagine, for example, the problem of automatically recognizing when a user is *playing tennis*. Such activity is composed by several instances of *walking*, *running*, and *sitting*, among others, with certain logical sequence and duration. The recognition of these *composite* activities from a set of *atomic* activities would surely enrich context awareness but, at the same time, brings additional uncertainty. Blanke et al. [90] provide an overview on this topic and propose a solution through several layers of inference.

- **Concurrent and overlapping activities:** The assumption that an individual only performs one activity at a time is true for basic ambulation activities (e.g., walking, running, lying, etc.). In general, human activities are rather overlapping and concurrent. A person could be *walking* while *brushing their teeth*, or *watching TV* while *having lunch*. Since only few works have been reported in this area, we foresee great research opportunities in this field. The interested reader might refer to the article of Helaoui et al. [91] for further information.

- **Multiattribute classification:** The purpose of a HAR system is, of course, providing feedback on the user's activity. But, context awareness may be enriched by also recognizing user's personal attributes. A case study could be a system that not only recognizes an individual *running*, but also identifies them as a female between 30 and 40 years old. We hypothesize that vital signs may have an important role in the determination of these attributes. To the best of our knowledge, there is no previous work on this topic.

- **Cost-sensitive classification:** Imagine an activity recognition system monitoring a patient with heart disease who cannot make significant physical effort. The system should never predict that the individual is *sitting* when they are actually *running*. But confusions between activities such as *waking* and *sitting* might be tolerable in this scenario. *Cost-sensitive classification* works exactly in that direction, maintaining a cost matrix C where the value C_{ij} is the cost of predicting activity i given that the actual activity is j. The values in this matrix depend on the specific application. In prediction time, the classifier can be easily adapted to output the activity class with the smallest misclassification cost. Also, in training time, the proportion of instances can be increased for the most expensive classes, forcing the learning algorithm to classify them more accurately. Additional information on cost-sensitive classification is available in [3], [47], [92].

- **Crowd-HAR:** The recognition of human activities has been somehow individualized, i.e., the majority of the systems predict activities in a single user. Although information from social networks has been shown effective to recognize human behaviors [93], recognizing collective activity patterns can be taken one step further. If we could gather activity patterns from a significant sample of people in certain area (e.g., a city, a state, or a country), that information could be used to estimate levels of sedentarism, exercise habits, and even health conditions in a target population. Furthermore, this sort of *participatory-human-centric* application would not require an economic incentive method. The users would be willing to participate in the system as long as they receive information on their health conditions and exercise performance, for example. Such data from thousands or millions of users may also be used to feed classification algorithms thereby enhancing their overall accuracy.

- **Predicting future activities:** Previous works have not only estimated activities but also behavior routines [65]. Based on this information, the system could predict what the user is about to do. This becomes especially useful for certain applications such as those based on advertisements. For instance, if the user is going to *have lunch*, he or she may receive advertisement on restaurants nearby.

- **Classifier flexibility:** People certainly perform activities in a different manner due to particular physical characteristics. Thus, acceleration signals measured from a child versus an elderly person are expected to be quite different. As we have seen in Section III-G, a human activity recognition model might be either *monolithic* or *user-specific*, each one having its own benefits and drawbacks. A middle ground to make the most of both worlds might be creating *group-specific* classifiers, clustering individuals with similar characteristics such as age, weight, gender, health conditions, among others. Then, our hypothesis is that a HAR system would be more effective having a recognition model for *overweight young men*, one for *normal male children*, another one for *female elderly*, and so forth.

VII. Conclusions

This paper surveys the state-of-the-art in human activity recognition based on wearable sensors. A two-level taxonomy is introduced that organizes HAR systems according to their response time and learning scheme. Twenty eight systems are qualitatively compared in regards to response time, learning approach, obtrusiveness, flexibility, recognition accuracy, and other important design issues. The fundamentals of feature extraction and machine learning are also included, as they are important components of every HAR system. Finally, various ideas are proposed for future research to extend this field to more realistic and pervasive scenarios.

References

[1] A. J. Perez, M. A. Labrador, and S. J. Barbeau, "G-sense: A scalable architecture for global sensing and monitoring," *IEEE Network*, vol. 24, no. 4, pp. 57–64, 2010.

[2] Y. Jia, "Diatetic and exercise therapy against diabetes mellitus," in *Second International Conference on Intelligent Networks and Intelligent Systems*, pp. 693–696, 2009.

[3] J. Yin, Q. Yang, and J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[4] O. X. Schlmilch, B. Witzschel, M. Cantor, E. Kahl, R. Mehmke, and C. Runge, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Computers in Human Behavior*, vol. 15, no. 5, pp. 571–583, 1999.

[5] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Computers in Human Behavior*, vol. 15, no. 5, pp. 571–583, 1999.

[6] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *Pervasive Computing, IEEE*, vol. 9, no. 1, pp. 48–53, 2010.

[7] T. van Kasteren, G. Englebienne, and B. Krse, "An activity monitoring system for elderly care using generative and discriminative models," *Journal on Personal and Ubiquitous Computing*, 2010.

[8] A. Tolstikov, X. Hong, J. Biswas, C. Nugent, L. Chen, and G. Parente, "Comparison of fusion methods based on dst and dbn in human activity recognition," *Journal of Control Theory and Applications*, vol. 9, pp. 18–27, 2011.

[9] J. Yang, J. Lee, and J. Choi, "Activity recognition based on rfid object usage for smart mobile devices," *Journal of Computer Science and Technology*, vol. 26, pp. 239–246, 2011.

[10] J. Sarkar, L. Vinh, Y.-K. Lee, and S. Lee, "Gpars: a general-purpose activity recognition system," *Applied Intelligence*, pp. 1–18, 2010.

[11] P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

[12] J. Candamo, M. Shreve, D. Goldgof, D. Sapper, and R. Kasturi, "Understanding transit scenes: A survey on human behavior-recognition algorithms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 206–224, 2010.

[13] C. N. Joseph, S. Kokulakumaran, K. Srijeyanthan, A. Thusyanthan, C. Gunasekara, and C. Gamage, "A framework for whole-body gesture recognition from video feeds," in *International Conference on Industrial and Information Systems (ICIIS)*, pp. 430–435, 2010.

[14] M. Ahad, J. Tan, H. Kim, and S. Ishikawa, "Human activity recognition: Various paradigms," in *International Conference on Control, Automation and Systems*, pp. 1896–1901, 2008.

[15] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, Alex, Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[16] J. Iglesias, J. Cano, A. M. Bernardos, and J. R. Casar, "A ubiquitous activity-monitor to prevent sedentariness," in *IEEE Conference on Pervasive Computing and Communications*, 2011.

[17] D. Choujaa and N. Dulay, "Tracme: Temporal activity recognition using mobile phone data," in *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, vol. 1, pp. 119–126, 2008.

[18] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 119–128, 2006.

[19] L. C. Jatoba, U. Grossmann, C. Kunze, J. Ottenbacher, and W. Stork, "Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity," in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5250–5253, 2008.

[20] T. Brezmes, J.-L. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone," in *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, vol. 5518, pp. 796–799, Springer Berlin / Heidelberg, 2009.

[21] K. Oh, H.-S. Park, and S.-B. Cho, "A mobile context sharing system using activity and emotion recognition with bayesian networks," in *International Conference on Ubiquitous Intelligence Computing*, pp. 244–249, 2010.

[22] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart monitor," in *International Symposium on Wearable Computers*, 2007.

[23] T.-P. Kao, C.-W. Lin, and J.-S. Wang, "Development of a portable activity detector for daily activity recognition," in *IEEE International Symposium on Industrial Electronics*, pp. 115–120, 2009.

[24] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *International Workshop on Wearable and Implantable Body Sensor Networks*, (Washington, DC, USA), IEEE Computer Society, 2006.

[25] O. D. Lara, A. J. Perez, M. A. Labrador, and J. D. Posada, "Centinela: A human activity recognition system based on acceleration and vital sign data," *Journal on Pervasive and Mobile Computing*, 2011.

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

[27] Y. Hanai, J. Nishimura, and T. Kuroda, "Haar-like filtering for human activity recognition using 3d accelerometer," in *IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, pp. 675–678, 2009.

[28] Z.-Y. He and L.-W. Jin, "Activity recognition from acceleration data using ar model representation and svm," in *International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 2245–2250, 2008.

[29] M. Berchtold, M. Budde, H. Schmidtke, and M. Beigl, "An extensible modular recognition concept that makes activity recognition practical," in *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, pp. 400–409, Springer Berlin / Heidelberg, 2010.

[30] D. Riboni and C. Bettini, "Cosar: hybrid reasoning for context-aware activity recognition," *Personal and Ubiquitous Computing*, vol. 15, pp. 271–289, 2011.

[31] T. Brezmes, J.-L. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone," in *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, vol. 5518 of *Lecture Notes in Computer Science*, pp. 796–799, Springer Berlin / Heidelberg, 2009.

[32] M. Berchtold, M. Budde, D. Gordon, H. Schmidtke, and M. Beigl, "Actiserv: Activity recognition service for mobile phones," in *International Symposium on Wearable Computers*, pp. 1–8, 2010.

[33] Z. He, Z. Liu, L. Jin, L.-X. Zhen, and J.-C. Huang, "Weightlessness feature; a novel feature for single tri-axial accelerometer based activity recognition," in *19th International Conference on Pattern Recognition*, pp. 1–4, 2008.

[34] Z. He and L. Jin, "Activity recognition from acceleration data based on discrete consine transform and svm," in *IEEE International Conference on Systems, Man and Cybernetics*, pp. 5041–5044, 2009.

[35] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, Gwo-Yun=Lee, and J.-S. Wang, "Online classifier construction algorithm for human activity detection using a tri-axial accelerometer," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 849–860, 2008.

[36] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes," *ACM Transactions on Sensor Networks*, vol. 6, no. 2, pp. 1–27, 2010.

[37] A. Khan, Y. Lee, and S. Lee, "Accelerometer's position free human activity recognition using a hierarchical recognition model," in *IEEE International Conference on e-Health Networking Applications and Services (Healthcom)*, pp. 296–301, 2010.

[38] A. Khan, Y.-K. Lee, S. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1166–1172, 2010.

[39] K. Frank, M. Rockl, M. Nadales, P. Robertson, and T. Pfeifer, "Comparison of exact static and dynamic bayesian context inference methods for activity recognition," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 189–195, 2010.

[40] "Google Places API," http://code.google.com/apis/maps/documentation/places/.

[41] K. L. Huang, S. S. Kanhere, and W. Hu, "Preserving privacy in participatory sensing systems," *Computer Communications*, vol. 33, no. 11, pp. 1266–1280, 2010.

[42] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *Journal of Systems and Software*, vol. 84, no. 11, pp. 1928–1946, 2011.

[43] I. J. Vergara-Laurens and M. A. Labrador, "Preserving privacy while reducing power consumption and information loss in lbs and participatory sensing applications," in *IEEE GLOBECOM*, 2011.

[44] M. Ermes, J. Parkka, and L. Cluitmans, "Advancing from offline to online activity recognition with wearable sensors," in *Engineering in Medicine and Biology Society. 30th Annual International Conference of the IEEE*, pp. 4451–4454, 2008.

[45] B. Longstaff, S. Reddy, and D. Estrin, "Improving activity classification for health applications on mobile devices using active and semi-supervised learning," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pp. 1–7, 2010.

[46] O. D. Lara and M. A. Labrador, "A mobile platform for real time human activity recognition," in *IEEE Conference on Consumer Communications and Networks*, 2012.

[47] I. H. Witten and E. Frank, *Data Mining, Practical Machine Learning Tools and Techniques*. Elsevier, 2 ed., 2005.

[48] R. T. Olszewski, C. Faloutsos, and D. B. Dot, *Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data*. 2001.

[49] K. Altun and B. Barshan, "Human activity recognition using inertial/magnetic sensor units," in *Human Behavior Understanding*, Lecture Notes in Computer Science, pp. 38–51, Springer Berlin / Heidelberg, 2010.

[50] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, A. Rahimi, A. Rea, G. Bordello, B. Hemingway, P. Klasnja, K. Koscher, J. Landay, J. Lester, D. Wyatt, and D. Haehnel, "The mobile sensing platform: An embedded activity recognition system," *Pervasive Computing, IEEE*, vol. 7, no. 2, pp. 32–41, 2008.

[51] J. Penttila, J. Peltola, and T. Seppanen, "A speech/music discriminator based audio browser with a degree of certanity measure," in *International Workshop Information Retrieval*, p. 125131, 2001.

[52] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *International Workshop on Wearable and Implantable Body Sensor Networks*, pp. 113–116, 2006.

[53] G. Dahlquist and ke Bjrck, *Numerical methods in scientific computing. vol 1*. SIAM, 2008.

[54] J. Cheng, O. Amft, and P. Lukowicz, "Active capacitive sensing: Exploring a new wearable sensing modality for activity recognition," in *Pervasive Computing*, vol. 6030 of *Lecture Notes in Computer Science*, pp. 319–336, Springer Berlin / Heidelberg, 2010.

[55] D. Minnen, T. Westeyn, D. Ashbrook, P. Presti, and T. Starner, "Recognizing soldier activities in the field," in *4th International Workshop on Wearable and Implantable Body Sensor Networks*, vol. 13, pp. 236–241, Springer Berlin Heidelberg, 2007.

[56] D. McGlynn and M. G. Madden, "An ensemble dynamic time warping classifier with application to activity recognition," in *Research and De-*

velopment in Intelligent Systems XXVII, pp. 339–352, Springer London, 2011.

[57] N. Pham and T. Abdelzaher, "Robust dynamic human activity recognition based on relative energy allocation," in *Distributed Computing in Sensor Systems*, vol. 5067 of *Lecture Notes in Computer Science*, pp. 525–530, Springer Berlin / Heidelberg, 2008.

[58] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[59] M. A. Hall, *Correlation-based Feature Selection for Machine Learning*. PhD. Dissertation, The University of Waikato, New Zealand, 1999.

[60] "The Waikato Environment for Knowledge Analysis," http://www.cs.waikato.ac.nz/ml/weka/.

[61] R. Cilla, M. Patricio, A. Berlanga, and J. Molina, "Creating human activity recognition systems using pareto-based multiobjective optimization," in *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2009.

[62] C. Randell and H. Muller, "Context awareness by analysing accelerometer data," in *The Fourth International Symposium on Wearable Computers*, pp. 175–176, 2000.

[63] M. Berchtold, M. Budde, D. Gordon, H. Schmidtke, and M. Beigl, "Actiserv: Activity recognition service for mobile phones," pp. 1–8, 2010.

[64] C. Zhu and W. Sheng, "Human daily activity recognition in robot-assisted living using multi-sensor fusion," in *IEEE International Conference on Robotics and Automation*, pp. 2154–2159, 2009.

[65] L. Vinh, S. Lee, H. Le, H. Ngo, H. Kim, M. Han, and Y.-K. Lee, "Semi-markov conditional random fields for accelerometer-based activity recognition," *Applied Intelligence*, vol. 35, pp. 226–241, 2011.

[66] J. Quinlan, *C4.5: programs for machine learning*. Morgan Kaufmann series in machine learning, Morgan Kaufmann Publishers, 1993.

[67] P. Antal, "Construction of a classifier with prior domain knowledge formalised as bayesian network," in *24th Annual Conference of the IEEE Industrial Electronics Society*, vol. 4, pp. 2527–2531, 1998.

[68] H. Zhang, "The Optimality of Naive Bayes.," in *FLAIRS Conference*, AAAI Press, 2004.

[69] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[70] S. Gallant, "Perceptron-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 1, no. 2, pp. 179–191, 1990.

[71] M. Stikic, D. Larlus, and B. Schiele, "Multi-graph based semi-supervised learning for activity recognition," in *International Symposium on Wearable Computers*, pp. 85–92, 2009.

[72] M. Stikic, D. Larlus, S. Ebert, and B. Schiele, "Weakly supervised recognition of daily life activities with wearable sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2521–2537, 2011.

[73] A. Ali, R. King, and G.-Z. Yang, "Semi-supervised segmentation for activity recognition with multiple eigenspaces," in *International Summer School and Symposium on Medical Devices and Biosensors*, pp. 314–317, 2008.

[74] D. Guan, W. Yuan, Y.-K. Lee, A. Gavrilov, and S. Lee, "Activity recognition based on semi-supervised learning," in *13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications*, pp. 469–475, 2007.

[75] T. Huynh and B. Schiele, "Towards less supervision in activity recognition from wearable sensors," in *10th IEEE International Symposium on Wearable Computers*, pp. 3–10, 2006.

[76] "The Java Data Mining Platform," http://www.jdmp.org/.

[77] "Zephyr Bioharness BT Website," http://www.zephyr-technology.com/bioharness-bt.html.

[78] "Google's Android becomes the world's leading smart phone platform," http://www.canalys.com/newsroom/google%E2%80%99s-android-becomes-world%E2%80%99s-leading-smart-phone-platform.

[79] D. Arvidsson, F. Slinde, and L. Hulthén, "Free-living energy expenditure in children using multi-sensor activity monitors," *Clinical nutrition (Edinburgh, Scotland)*, vol. 28:3, s. 305, 2009.

[80] A. Teller, "A platform for wearable physiological computing," *Interacting with Computers*, pp. 917–937, 2004.

[81] "Google Scholar," http://scholar.google.com/.

[82] "Sprint Java ADP," http://developer.sprint.com/site/global/home/p_home.jsp.

[83] A. Khan, Y. Lee, and S. Lee, "Accelerometer's position free human activity recognition using a hierarchical recognition model," in *12th IEEE International Conference on e-Health Networking Applications and Services (Healthcom)*, pp. 296–301, 2010.

[84] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Eleventh Annual Conference on Computational Learning Theory*, pp. 92–100, ACM, 1998.

[85] "Datasets for Human Activity Recognition from Tim Van Kasteren's website," https://sites.google.com/site/tim0306/datasets.

[86] "Datasets for Human Activity Recognition from MIT Media Lab," http://architecture.mit.edu/house_n/data/PlaceLab/PlaceLab.htm.

[87] "Datasets for Human Activity Recognition from Tanzeem Choudhury's website," http://www.cs.dartmouth.edu/~tanzeem/teaching/CS188-Fall08/dataset.html.

[88] "Datasets for Human Activity Recognition from ETH's Wearable Computing Lab," http://www.wearable.ethz.ch/resources/Dataset.

[89] "2011 Activity Recognition Challenge," www.oportunity-project.eu/challenge.

[90] U. Blanke and B. Schiele, "Remember and transfer what you have learned - recognizing composite activities based on activity spotting," in *International Symposium on Wearable Computers*, pp. 1–8, 2010.

[91] R. Helaoui, M. Niepert, and H. Stuckenschmidt, "Recognizing interleaved and concurrent activities: A statistical-relational approach," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–9, 2011.

[92] C. Elkan, "The foundations of cost-sensitive learning," in *Seventeenth International Joint Conference on Artificial Intelligence*, pp. 973–978, 2001.

[93] N. Lane, Y. Xu, H. Lu, A. Campbell, T. Choudhury, and S. Eisenman, "Exploiting social networks for large-scale human behavior modeling," *Pervasive Computing, IEEE*, vol. 10, no. 4, pp. 45–53, 2011.