Expert Systems with Applications 39 (2012) 8013–8021

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Daily living activity recognition based on statistical feature quality group selection

Oresti Banos *, Miguel Damas, Hector Pomares, Alberto Prieto, Ignacio Rojas

University of Granada, Higher Technical School of Computer Sciences and Telecommunications Engineering, Department of Computer Architecture and Computer Technology, C/Periodista Daniel Saucedo Aranda s/n, E-18071 Granada, Spain

ARTICLE INFO

Keywords:
Activity recognition
Ambulatory monitoring
E-health
E-monitoring
Feature selection
Knowledge extraction

ABSTRACT

The benefits arising from proactive conduct and subject-specialized healthcare have driven e-health and e-monitoring into the forefront of research, in which the recognition of motion, postures and physical exercise is one of the main subjects. We propose here a multidisciplinary method for the recognition of physical activity with the emphasis on feature extraction and selection processes, which are considered to be the most critical stages in identifying the main unknown activity discriminant elements. Efficient feature selection processes are particularly necessary when dealing with huge training datasets in a multidimensional space, where conventional feature selection procedures based on wrapper methods or 'branch and bound' are highly expensive in computational terms. We propose an alternative filter method using a feature quality group ranking via a couple of two statistical criteria. Satisfactory results are achieved in both laboratory and semi-naturalistic activity living datasets for real problems using several classification models, thus proving that any body sensor location can be suitable to define a simple one-feature-based recognition system, with particularly remarkable accuracy and applicability in the case of the wrist.

1. Introduction

The percentage of EU citizens aged 65 years or over is projected to increase from 17.1% in 2008 to 30.0% in 2060. In particular, the number of 65 years old is projected to rise from 84.6 million to 151.5 million, while the number of people aged 80 or over is projected to almost triple from 21.8 million to 61.4 million (*EUROSTAT: New European Population projections* 2008–2060). It has been calculated that the purely demographic effect of an ageing population will push up health-care spending by between 1% and 2% of the gross domestic product (GDP) of most member states. At first sight this may not appear to be very much when extended over several decades, but on average it would in fact amount to approximately a 25% increase in spending on health care, as a share of GDP, in the next 50 years (European Economy Commission, 2006). The effective incorporation of technology into health-care systems could therefore be decisive in helping to decrease overall public spending on health. One of these emerging health-care systems is daily living physical activity recognition.

Daily living physical activity recognition is currently being applied in chronic disease management (Amft & Tröster, 2008; Zwartjes, Heida, van Vugt, Geelen, & Veltink, 2010), rehabilitation systems (Sazonov, Fulk, Sazonova, & Schuckers, 2009) and disease prevention (Sazonov, Fulk, Hill, Schutz, & Browning, 2011; Warren et al., 2010), as well as being a personal indicator to health status (Arcelus et al., 2009). One of the principal subjects of the health-related applications being mooted is the monitoring of the elderly. For example, falls represent one of the major risks and obstacles to old people's independence (Najafi, Aminian, Loew, Blanc, & Robert, 2002; Yu, 2008). This risk is increased when some kind of degenerative disease affects them. Most Alzheimer's patients, for example, spend a long time every day either sitting or lying down since they would otherwise need continuous vigilance and attention to avoid a fall.

The registration of daily events, an important task in anticipating and/or detecting anomalous behavior patterns and a primary step towards carrying out proactive management and personalized treatment, is normally poorly accomplished by patients' families, healthcare units or auxiliary assistants because of limitations in time and resources. Automatic activity-recognition systems could allow us to conduct a completely detailed monitoring and assessment of the individual, thus significantly reducing current human supervision requirements.

The primary difficulty in activity recognition lies in designing a system the reliability of which is independent of the person carrying out the exercise or the particular style of execution of the activity in question. Complexity is further increased by distortion elements related to system monitoring and processing, along with the random character of the execution. Most studies to date have been based on laboratory data (i.e., involving direct

* Corresponding author.
E-mail address: oresti@atc.ugr.es (O. Banos).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.164

supervision by the researcher) and have achieved successful recognition of the most prevalent everyday activities (lying, sitting, standing and walking: Aminian et al., 1999; Karantonis, Narayanan, Mathie, Lovell, & Celler, 2006; Maurer, Smailagic, Siewiorek, & Deisher, 2006; Ravi, Dandekar, Mysore, & Littman, 2005). Nonetheless, the apparently good recognition results obtained during supervised experiences cannot be extrapolated to habitual real-life conditions (Könönen, Mäntyjärvi, Similä, Pärkkä, & Ermes, 2010).

The ideal scenario would be a naturalistic monitoring context consisting of a scenario with no intervention on the researcher's part and without the subject's cognitive knowledge about the exercise conducted, but unfortunately this is currently unfeasible. Some studies have applied a so-called semi-naturalistic approach (Bao & Intille, 2004; Ermes, Parkka, Mantyjarvi, & Korhonen, 2008; Foerster, Smeja, & Fahrenberg, 1999; Pirttikangas, Fujinami, & Nakajima, 2006; Uiterwaal, Glerum, Busser, & van Lummel, 1998), an intermediate between laboratory and naturalistic monitoring based on the inference of the hidden activity through the proposal of a related exercise, thus minimizing the subject's awareness of the true nature of the data being collected. This approximation is somewhat more realistic than laboratory experimental setups.

The classic method for activity identification is based on three main stages: feature extraction (e.g., statistical features (Baek, Lee, Park, & Yun, 2004; Maurer et al., 2006; Ravi et al., 2005), wavelet coefficients (Nyan, Tay, Seah, & Sitoh, 2006; Preece, Goulermas, Kenney, & Howard, 2009; Preece et al., 2009) or other custom-defined coefficients (He, Liu, Jin, Zhen, & Huang, 2008; Mathie, Coster, Lovell, & Celler, 2003)), feature selection (e.g., principal or independent component analysis (Mantyjarvi, Himberg, & Seppanen, 2001), forward–backward selection (Pirttikangas et al., 2006), correlation (Maurer et al., 2006), etc.) and classification (primarily supervised learning approaches such as artificial neural networks (Engin et al., 2007; Parkka et al., 2006; Zhang et al., 2005), support vector machines (Begg & Kamruzzaman, 2005; Parera, Angulo, Rodríguez-Molinero, & Cabestany, 2009; Sazonov et al., 2009), Bayesian classifiers (Bao & Intille, 2004; Wu, Osuntogun, Choudhury, Philipose, & Rehg, 2007) and hidden Markov models (Minnen, Starner, Essa, & Isbell, 2006; Sazonov et al., 2011), among others). For a detailed review of classification techniques used in activity recognition the reader is referred to Preece, Goulermas, Kenney, and Howard (2009) and Preece et al. (2009).

Evidently, all these stages are important, but in this work we want to emphasize the importance of selecting the most interesting features to improve the efficiency of the subsequent pattern recognition systems, especially bearing in mind the rather discouraging results obtained with semi-naturalistic data. It is well known that a large number of features are directly translated into numerous classifier parameters, so keeping the number of features as small as possible is in line with our desire to design classifiers with good generalization capabilities, the best scenario being a knowledge inference system defined by just a few features. Consequently, we propose here an automatic method to extract a subset of the most important features to be used in activity recognition, which is especially suitable for looking for optimum single-feature classifiers with multiclass absolute discrimination capability.

The rest of the paper is organized as follows: Section 2 contains a description of the experimental setup, preprocessing process, features extracted from the data and the proposed rank-based feature selection method. Section 3 presents the results obtained, including a comparison of the performance of several different approaches. These results are subsequently discussed in Section 4 and our final conclusions are summarized in Section 5.

2. Methods

2.1. Experimental setup

Our experimental setup starts from a set of signals corresponding to acceleration values measured by a group of sensors (accelerometers) attached to different strategic parts of the body (hip, wrist, arm, ankle and thigh) for several daily activities[1] following both laboratory and semi-naturalistic monitoring schemes (Bao & Intille, 2004). Our study is focused on the four most common physical activities that are of particular relevance to health-care applications: *walking*, *sitting and relaxing*, *standing still* and *running* (Fig. 1). Although other daily living activities may be chosen, we have specifically considered these four for the pairwise similarities between walking/running and sitting/standing, both with respect to the way they are performed and the energy they entail, although this assumption may be distorted under natural circumstances.

2.2. Signal processing

The initial information provided by the sensors has some artifacts and noise associated to the data acquisition process. Bearing in mind that a 20 Hz sampling is sufficient to assess habitual daily physical activity (Bouten, Koekkoek, Verduin, Kodde, & Janssen, 1997; Mathie, Coster, Lovell, & Celler, 2004), a low-pass elliptic filter with 20 Hz cutoff frequency, followed by a 0.5 Hz cutoff frequency high-pass elliptic filter are applied to respectively remove the high frequency noise and the gravitational acceleration component from the signal (Fahrenberg, Foerster, Smeja, & Müller, 1997). Other proposals such as mean/median or wavelet-based filtering (Najafi et al., 2002) could be assessed for signal enhancement, but we will consider them in the next feature extraction phase.

2.3. Feature extraction

It is common in works concerning activity recognition to use a reduced feature set to characterize the monitored signals, mainly composed of statistical, time-frequency and heuristic features. The validity of this approach has been demonstrated in laboratory-context experiments, but due to the difficulty of precise knowledge inference concerning semi-naturalistic monitoring, a wider analysis is needed to reveal any unidentified powerful discriminant features, even those lacking obvious physical interpretability.

Thus we generated a parameter set comprising 861 features corresponding to a combination of statistical functions such as median, kurtosis, mode, range and so on, and magnitudes or functions obtained from a domain transformation of the original data such as energy spectral density, spectral coherence and wavelet coefficients ("*a*1 to *a*5" and "*d*1 to *d*5" Daubechies levels of decomposition) among others, for both signal axes. "Fisher asymmetry coefficient of the *X* axis signal histogram", "*Y* axis signal energy spectral density maximum" or "*X* axis–*Y* axis cross correlation harmonic mean" are possible examples of features obtained from the complete set defined (Table 1). Several of these features have been tested in previous works primarily on time and frequency domain (for example, amplitude peak (Laerhoven & Gellersen, 2004), arithmetic mean (Lee & Mase, 2002; Wang, Yang, Chen, Chen, & Qinfeng Zhang, 2005), variance or standard deviation (Heinz et al., 2003; Kern, Schiele, & Schmidt, 2003), energy and correlation between axes (Bao & Intille, 2004; Ravi et al., 2005), etc.), but many of them are unprecedented in this context. Features are extracted from the

[1] Database facilitated in Bao and Intille (2004) by Prof. Stephen Intille (Massachusetts Institute of Technology).

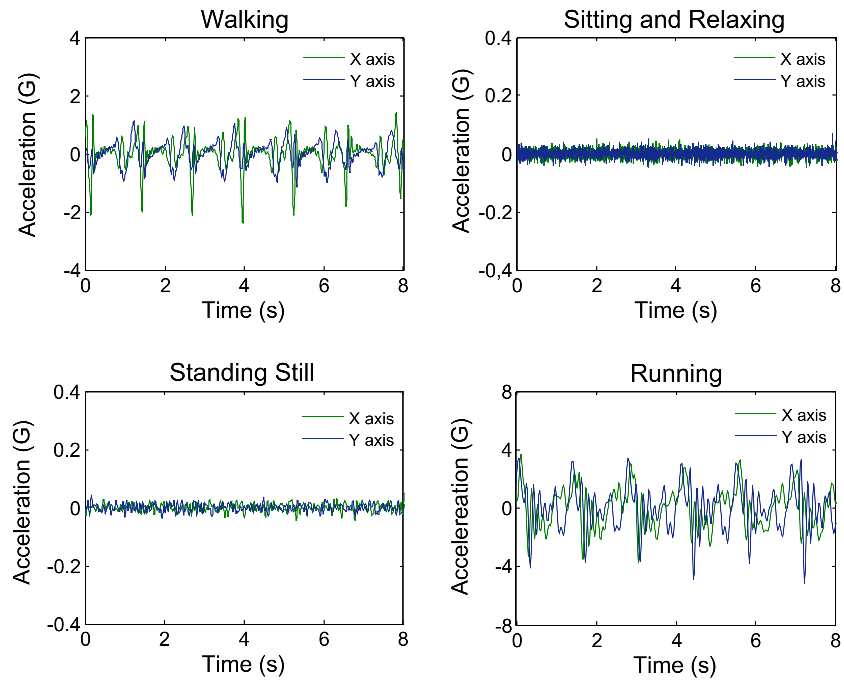

Fig. 1. Signals corresponding to four common daily physical activities (original ankle accelerometer filtered data). Noteworthy is the similarity of the signal patterns between "walking" and "running" (dynamic activities), and "sitting and relaxing" and "standing still" (static activities) respectively.

Table 1
Feature set generation functions.

Magnitudes	Statistical functions
Amplitude (AMP)	4th and 5th central statistical moments (γ_4 and γ_5)
Autocorrelation function (ACF)	Energy (e)
Cepstrum (CEPS)	Arithmetic/Harmonic/Geometric/Trimmed mean (μ_a, μ_h, μ_g, μ_t)
Cross correlation function (XCORRF)	Entropy (Ω)
Energy spectral density (ESD)	Fisher asymmetry coefficient (γ_1)
Spectral coherence (SC)	Maximum/position of (max/p_{max})
Spectrum amplitude/phase (SA/SP)	Median ($\mu_{1/2}$)
Histogram (HIST)	Minimum/position of (min/p_{min})
Historical data lags (HLAGS)	Mode (m)
Minimum phase reconstruction (MPR)	Kurtosis (γ_2)
Daubechies wavelet decomposition (WAV)	Data range (rang)
	Total harmonic deviation (thd)
	Variance (σ^2)
	Zero crossing counts (zcc)

data according to a windowing scheme similar to that proposed in (Bao & Intille, 2004).

At this stage, it is necessary to rely on a feature selection process with the responsibility of deciding which features are the most important ones to ascertain the kind of activity the person is carrying out. The method designed to accomplish this task is described in the next subsection.

2.4. Feature selection: proposed ranking quality group technique based on discrimination and robustness (RQG-DR)

Obtaining a specific group of variables from a large initial set is not a trivial task because of the huge number of possible feature combinations. In our experimental setup the sample space is represented by $n = 861$ features, so brute force techniques such as 'branch and bound' ($O(2^n)$ convergence, which implies $2^{861} \approx 1.5 \times 10^{259}$ possible combinations) or wrapper methods are impractical. In this section, we present an alternative filter method based on the concepts of discrimination and robustness of a set of features.

Let us define the *sample range* of a class for a given feature as the set of values included between the maximum and minimum values (both inclusive) that the feature or variable takes for this class. In Fig. 2a and b, numerical values from two specific features are represented for each class. The boxes associated with every sample set represent the statistical sample range, which is based on the data distribution.

Given a feature defined by a group of samples for each class, we rank its discriminant capability with respect to a particular class according to the overlapping probability between this class and the others. This is calculated by computing the number of samples from the analyzed class that also fall inside the sample ranges of other classes. Given N classes and Q_n samples belonging to class n, the *overlapping probability* of class k for a specific feature f can be defined as:

$$p_f(k) = \frac{1}{N-1}\sum_{n \neq k}^{N} \frac{m(k,n)}{Q_n} \quad \forall n, k = 1, \ldots, N \tag{1}$$

where $m(k,n)$ is the number of samples from the class k inside the sample range of class n. To make this more easily understandable, two examples (both with $N = 4$ and $Q_n = 8$) for two specific features are shown in Fig. 2. In the first case (Fig. 2a) the overlapping probability for the class *walking* (identified as $k = 1$ for simplicity) for the feature f_1 = *X axis signal wavelet coef. (d1) zero crossing counts* is $p_{f1}(1) = ⅓(0/8 + 0/8 + 6/8) = 0.25$ since there are six samples from the class *running* in the data range defined for class *walking* and 0 from the rest of the classes. Similarly, it is easy to see that the overlapping probability is 0, 0.08 and 0.33, for *sitting*, *standing* and *running* respectively. Therefore, this feature is able to distinguish between the activity *walking* and the activities *standing still* or *sitting and relaxing*, but it could be mistaken with an approximately 25% probability for one of the other activities. If the statistical sample range of the very same feature is considered (i.e., without outliers, marked with a '+' sign) the overlapping probability gets reduced to approximately 4%: $p_{f1}(1) = ⅓(0/8 + 0/8 + 1/8) = 0.04$.

We want to emphasize this detail because the inclusion of outliers can result in some possible discriminant features being rejected, but looking for an extremely robust system, the best features may probably be those which offer an important discriminant capability even when considering outliers. This is especially important for semi-naturalistic data because of the great variety of unpredictable possible actions and movements associated to every base activity. Fig. 2b illustrates an example of a completely discriminant feature for all the classes, since no overlapping appears between any of them. From a statistical point of view, a possible outlier is determined for the class *running*, and even so no overlapping with other classes is presented, which makes it a more robust feature.

A thresholding process is now carried out to identify whether the feature analyzed is discriminative or not. Given any specific feature, if the overlapping probability for every class analyzed exceeds a pre-specified threshold, the feature will be considered as not being discriminant for this class. This threshold is defined as the *overlapping threshold* (o_{th}) and it can take values from 0 (the most restrictive case, for no overlapping between classes) to 1 (the most relaxed, when every sample from a class is inside the others). Therefore, using the above definition of $p_f(k)$ we can say that:

$$\begin{aligned} &f \text{ discriminates class } k \text{ if } p_f(k) \leqslant o_{th} \\ &f \text{ does not discriminate class } k \text{ if } p_f(k) > o_{th} \end{aligned} \qquad (2)$$

Continuing the example of Fig. 2, it is clear that if the most restrictive threshold is used, the feature *X axis signal wavelet coef. (d1) zero crossing counts* will be considered as discriminant only for the class *sitting* (associated to the wrist sensor), while the feature *Y axis wavelet coef. (d2) range* will be recognized as completely discriminatory for all the activities (for the hip accelerometer).

Fig. 3 shows the evolution of the discriminant feature set as the overlapping probability is relaxed for both laboratory and semi-naturalistic data. By considering the overlapping probability as a relative measure of the class confusion error, the larger the overlapping threshold, the fewer discriminant capabilities the chosen features have. Fig. 3a illustrates that the number of features that can distinguish the current activity from the rest increases with o_{th} since samples from other classes are positively accepted within the sample range, but at the same time allowing for an increasing error probability. Moreover, it is important to underline the importance of the monitoring philosophy in this context. In Fig. 3b less significant discriminant features are available for most of the classes for the zero overlapping threshold, which is especially notable for the static activities. This is related to the diversity of movements which are typically appended in unsupervised monitoring to every canonical activity (i.e., lab activity), a consequence of the natural behavior of the people engaged in those daily living tasks. As can be expected, activities requiring low acceleration values (*sitting/standing*) are more affected by unexpected actions due to any isolated energetic movement which is *a priori* not allowed for in their natural description. This is clearly less important for intrinsic dynamic activities (*walking/running*) where all those unexpected movements can be more easily masked.

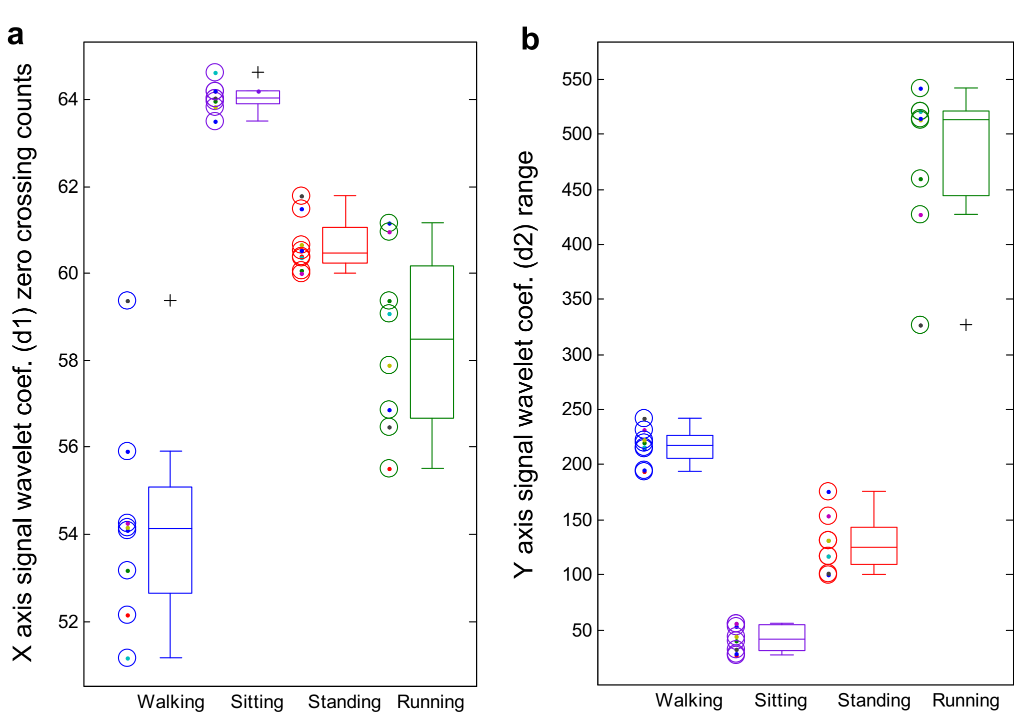

Fig. 2. Feature values extracted for the (a) wrist and (b) hip accelerometer respectively. Every circle represents the value (sample) of the feature for a specific subject (8 in all, an example subset of the 20 subjects considered in the study) and the associated box represents the statistic distribution of the sample set (the central mark is the median, the edges of the box are the 25th and 75th percentiles, and outliers are plotted individually using a '+' marker).

Apart from the discriminant capacity of a feature or a set of features, a second characteristic is now defined which takes into account the usability of this set of features in different information contexts or sensors. For instance, a specific measure taken from the ankle accelerometer, for instance, can be very discriminative when distinguishing between the activities *walking* and *standing still*, but this very same measure may not be that reliable when taken from the thigh accelerometer. There may be some measures with the same discriminant capability between those activities that are not so dependent on the exact location of the sensor or, at least, that are still reliable when taken from a larger number of sensors. The feature selection method proposed allows us to have a measure of this generalization capability. We will denote this measure as the *robustness* criterion of a set of features. In short, the discriminant capacity (DC) of a feature provides information about how useful this feature is for a given sensor, and the robustness (R) of a feature is concerned about how this discriminant capacity depends on where the sensor is located.

Combining both criteria we obtain a quality ranking procedure capable of grouping features into different stages. Features are first ordered by taking into account the number of classes they can discriminate (DC). They are then subranked according to the number of sensors with which these features are discriminant (R) for the same DC.

To assess the effectiveness of the method, we made a comparison with other feature selectors based on the Bhattacharyya distance, Entropy, ROC, T-test and Mann–Whitney–Wilcoxon test. A specific description of the realization implemented for those feature selectors can be found in (Liu, 1998; Theodoridis & Koutroumbas, 2009).

2.5. Classification

Different classification schemes have been used in works concerning activity recognition, among which we have taken into account in our approach decision trees (DT), Naive Bayes (NB) and support vector machines (SVM).

Decision trees have proved to perform excellently in activity recognition in some papers (Bao & Intille, 2004; Maurer et al., 2006; Parkka et al., 2006), although they were less accurate in others (Ermes et al., 2008). DT algorithms examine the discriminatory ability of the features one at a time, creating a set of rules that ultimately lead to a complete classification system. A C4.5 implementation (Duda, Hart, & Stork, 2001), which is one of the most successful DT algorithms, has been used in this work.

As the aim of this study is to find the minimum number of features (in the best case, only one) for a given classification accuracy, NB (Theodoridis & Koutroumbas, 2009) may be an appropriate approach as long as stochastical independence is guaranteed, which is more attainable if few features are used.

SVM (Cristianini & Taylor, 2000; Vapnik, 1998) is a very popular technique in machine learning problems. This method has not been extensively used in activity recognition studies but, bearing in mind the remarkable accuracy rates obtained in other contexts, and the increasing use of SVM in recent years, we have taken it into account in our work. In particular, we used an RBF kernel with hyper-parameters γ and C automatically tuned using a grid search technique for each classifier. The multiclass extension was achieved using a 'one versus all' approach in this context.

2.6. Experiments

The first study undertaken consisted of assessing the proposed method applied to laboratory and semi-naturalistic data separately. Consequently, the best-ranked features for each environmental context (laboratory and semi-naturalistic) were used to define the different inference knowledge systems, analyzing the recognition performance in every case.

In (Bao & Intille, 2004) both monitoring datasets were used together, using cross-validation only with the semi-naturalistic data. This is a valid approach to effect an overall interpretation of activities in both contexts, but we considered that specific capabilities of the resulting system might remain hidden. To find these truly significant features, we looked for features selected for every dataset independently, analyzing their discriminant capability throughout the other dataset (*cross study*). This allowed us to assess the capability of the features used in a general recognition context, which, together with the data preprocessing made and the machine learning technique used, will ultimately define the recognition system.

Several feature vector lengths were evaluated (1, 2, 3, 4, 5 and 10 best ranked respectively) and a 10-cross validation process was applied for each classifier, i.e., data is randomly partitioned into ten pieces of equal size using each piece as the validation set and the remaining 90% as the training set, the performance indices being averaged over the ten cases. To ensure statistical robustness in our study, each experiment was repeated 100 times.

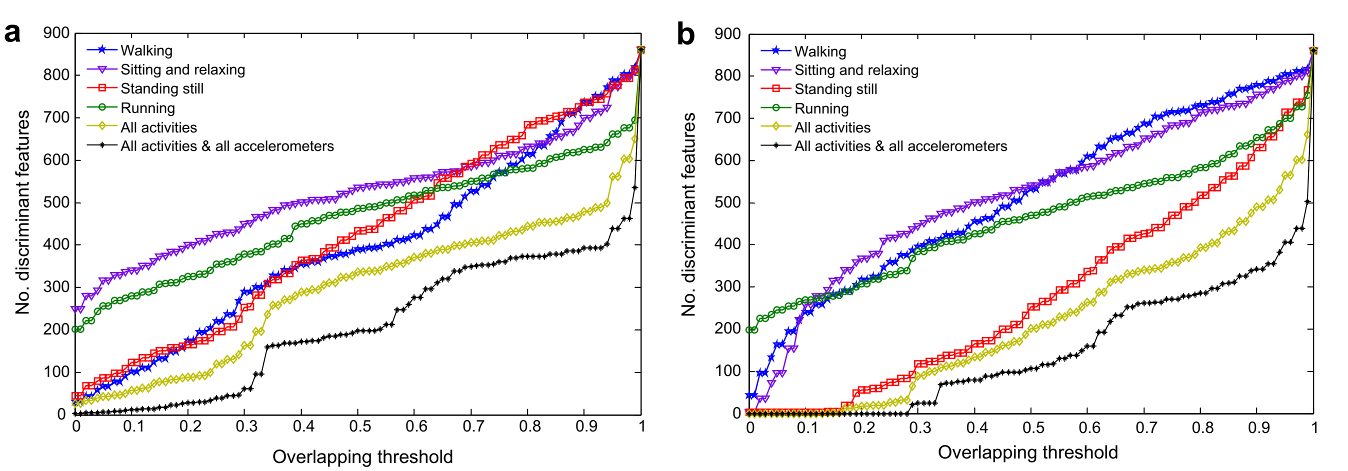

Fig. 3. Evolution of the number of discriminant features for several overlapping probabilities. Features extracted from (a) laboratory and (b) semi-naturalistic data, corresponding to the ankle accelerometer. The "All activities" label refers to common discriminant features for all the activities respectively. The "All activities and all accelerometers" label is for features that retain their "all activities" discriminant capability with all the accelerometers.

3. Results

Having described the method we go on in this section to present the results corresponding to the particular activity-recognition problem analyzed.

Features ranked for laboratory data by applying RQG-DR for $o_{th} = 0$ are shown in Table 2 (an equivalent table was obtained for semi-naturalistic data). For example, the fifth central statistical moment, the maximum and the range of the energy spectral density and the spectrum amplitude were ranked in the quality group #11. These features were able to discriminate two activities using all five accelerometers. The computational simplicity of the proposed feature selection algorithm can be clearly seen in comparison with other selection methods taken from the literature (Fig. 4).

With regard to the best case from the computational complexity point of view, in which only one feature is used for each binary

Table 2
Quality group feature ranking for laboratory data.

DC	R	Features ranked in each quality group	DC	R	Features ranked in each quality group
4	5	#**1**: WAV{a3} – μ_g	2	5	#**11**: ESD/SA – γ_5/max/rang
4	4	#**2**: AMP/AC/WAV{a5,a4,a2,a1,d5,d4,d3} – μ_g	2	4	#**12**: ESD/SA – μ_a
4	3	#**3**: AC – e WAV{a3,a2,d4} – σ^2	2	3	#**13**: ESD – e WAV{a3} – max WAV{d4} – rang WAV{d3} – e WAV{d2} – μ_g
4	2	#**4**: AMP/AC/ XCORR/MPR/WAV{a4,a1,d5,d3,d2,d1} – σ^2 AMP/AC/ XCORR/WAV{a4,a2,a1,d3,d2,d1} – γ_4 AC – γ_5/e/μ_a/max/rang MPR – μ_g	2	2	#**14**: AMP – μ_t/max/rang/zcc ESD/SA – μ_t HLAGS – μ_g/max/rang/σ^2 MPR – max/min/m/rang WAV{a2} – max WAV{a4,a2} – γ_2 WAV{a4,a3,a2} – rang
4	1	#**5**: AMP/AC/ XCORR/HLAGS/WAV{a1,d5} – min/m AMP/MPR/ XCORR/WAV{a4,a3,a2,a1,d5,d4,d1} – e AMP/ MPR/ WAV{a1} – γ_5 AC – μ_a/max/zcc XCORR – μ_g/max XCORR/WAV{d5,d3,d2} – rang ESD/SA/ WAV{a5} – σ^2 ESD/SA/WAV{a3,d5,d4} – γ_4 WAV{a1,d1} – μ_g/zcc	2	1	#**15**: AMP/SA/ WAV {a4,a3,d5} – γ_2 HIST/ WAV {d4,d2} – γ_5 HIST – γ_4/σ^2 SA/MPR – γ_1 HLAGS – μ_a HLAGS/ WAV{a2,a1} – μ_t HLAGS/ WAV{d3,d2} – $\mu_{1/2}$ WAV{a5} – e WAV{a4,a2,a1,d5,d4,d3,d2} – max WAV{a4,a3,a2,d4,d3,d2} – min WAV{a4,a3,a2,d4,d3,d2} – m WAV{a1} – rang WAV{a2,a1,d4} – zcc
3	5	#**6**: *NO FEATURES RANKED IN THIS GROUP*	1	5	#**16**: XCORR – γ_5
3	4	#**7**: *NO FEATURES RANKED IN THIS GROUP*	1	4	#**17**: SA – e MPR – γ_2 WAV{d1} – μ_t WAV{a5} – γ_5
3	3	#**8**: *NO FEATURES RANKED IN THIS GROUP*	1	3	#**18**: CEPS – max ESD/SA – μ_g HLAGS/WAV{a5,d1} – m WAV{a5,d1} – min WAV{a5} – rang WAV{d2} – μ_t
3	2	#**9**: *NO FEATURES RANKED IN THIS GROUP*	1	2	#**19**: AC/ HIST – γ_1 AC/WAV{d3} – μ_t CEPS/HIST – rang CEPS/HLAGS – e ESD/SA – $\mu_{1/2}$ HIST – γ_2 HIST/WAV{a5} – max HIST/WAV{d3} – zcc WAV{d1} – γ_5
3	1	#**10**: WAV{a3,a2,d3} – γ_5	1	1	#**20**: AMP/XCORR/WAV{a3,a1,d1} – $\mu_{1/2}$ AC/HIST/XCORR/WAV{a5,d2} – γ_2 AC/MPR – pmin CEPS – min CEPS/HLAGS/WAV{a4,d5} – γ_5 ESD/SA – pmax HIST/WAV{a4,a3} – μ_t HIST – e XCORR/WAV{a4,a2} – zcc XCORR – Ω
			0	–	#**21**: REST OF FEATURES (***461***)

Features ranked in each quality group for the activity recognition problem (laboratory data). Parameter $o_{th} = 0$ (most restrictive case) is set for selection process.

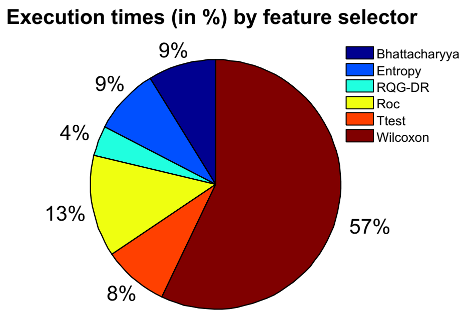

Fig. 4. Normalized time requirements for each feature selection method tested.

classifier (each classifier is specialized in just one class), the accuracy results obtained for all the classification techniques in both studies described in Section 2.6 are set out in Fig. 5. Subplots (a), (b) and (c) correspond to training and validation using the features extracted for the same monitoring context (lab or semi-naturalistic), whereas (d), (e) and (f) are evaluated on the basis of the best features found using the data of the *opposite* monitoring context.

The importance of the performance results obtained should be assessed bearing in mind the number of different features employed for each sensor, which is extremely important in real-time applications. Thus, each bar in Fig. 5 contains a value that identifies the number of different features used by the classifiers for every sensor. Since 4 classes are analyzed in this work, this number ranges from 1 (the best case, using the same feature for every classifier) to 4 (the less efficient case, using a different feature for each classifier). This is in accordance with the features extracted in every selection process for each class and sensor.

Furthermore, the overall number of different features (arrived at by comparison of a single feature chosen for each class for all the sensors) is annotated in brackets together with each feature selection method (see labels on the X-axis). Since four activities and five sensors are considered in our study, the maximum value will be 20, when all the features selected are different for each class and sensor, and the minimum will be just 1, when the same feature is selected for every class and for all the sensors.

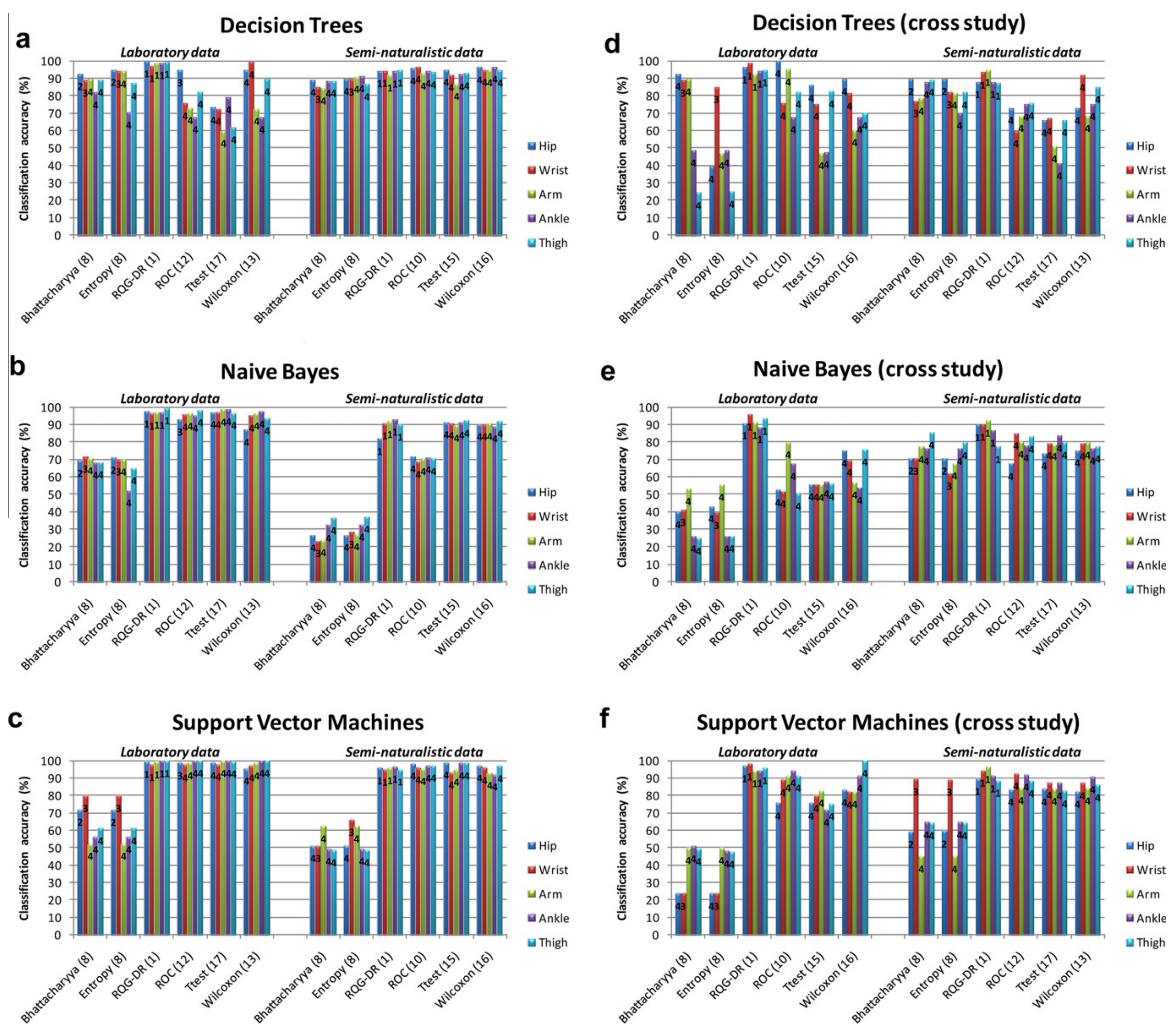

Fig. 5. Accuracy rates using one feature for each classifier. The features used in (a), (b) and (c) correspond to the original monitored data (laboratory/semi-naturalistic), whereas those employed in (d), (e) and (f) are extracted from the *opposite* monitoring context.

4. Discussion

The aim of this study has been to assess the feasibility of activity recognition systems based on a minimum defining feature set, taking into account a huge set of possible discriminant features, many of them disregarded in previous works. An extensive initial feature set is analyzed using several feature selection methods.

Knowledge inference systems designed for both laboratory and semi-naturalistic monitoring contexts achieve quite good results using several methods, as shown in Fig. 5a–c. Particularly remarkable are the results obtained using features selected from RQG-DR, ROC, *T*-test or Wilcoxon in combination with SVM, with accuracy rates close to 100% for laboratory data, and above 95% (more than 99% for some cases such as the hip sensor in *T*-test) for the semi-naturalistic study for each accelerometer. NB classifiers are less accurate, although they do offer good results for the laboratory approach. The results for DT are quite similar to SVM when features selected with our proposed method are used, with good results in general for the semi-naturalistic approach.

Of particular interest is the cross-study considered. The assessment of semi-naturalistic-context recognition systems (defined via the corresponding features selected) applied to laboratory data could permit us to demonstrate their performance beyond the context in which they had originally been defined. This can be seen as a measure of the context-generalization capacity of semi-naturalistic systems versus supervised ones (original laboratory data-based recognition systems applied to semi-naturalistic data).

The results presented in Fig. 5d–f show that the features originally selected by RQG-DR from semi-naturalistic information offer the best overall accuracy rates in DT, NB and especially in SVM when they are applied to laboratory data. The wrist location, considered to be the least intrusive for individuals, stands out with accuracy rates of more than 95% in NB and SVM and close to 100% in DT. Other possibilities are the hip sensor in combination with DT and features chosen by ROC, or the thigh accelerometer with SVM using Wilcoxon's selected features, which lead to complete recognition effectiveness.

On the other hand, the application of features extracted from the supervised context and applied to semi-naturalistic data allows us to identify intrinsic characteristics of the activities analyzed and their particular style of execution. Again, the feature extracted by using RQG-DR is the most accurate (Fig. 5d–f), although the results are slightly poorer compared to the laboratory data tested according to the semi-naturalistic features analyzed before. This is completely reasonable because, as might be expected, the nature of the exercises monitored in the semi-naturalistic approach contains an extensive set of different ways of acting and many different *a priori* unrelated events. Consequently, features extracted and selected for unsupervised data have to deal with a widespread range of situations. Therefore, if a powerful discriminant feature set is found (as occurs in our case), it should necessary discriminate between major executions of every activity analyzed, taking laboratory conduct to be just one of these. Otherwise, since laboratory features are primarily focused on activities developed following the supervised scheme, thus tending to suppress "anomalous" or "non-desirable" movements or gestures, it is extremely complicated to achieve outstanding results when applied to semi-naturalistic information.

We have measured the performance of our systems only in terms of their classification accuracy, but one of the most important objectives of our work has been to achieve computationally efficient systems, a very important issue for real applications. From this perspective, RQG-DR not only obtains the most accurate results but achieves this using the same feature for each classifier and any sensor (*geometric mean of wavelet coefficients obtained through a 3-level Daubechies decomposition* (*a3*) for laboratory data and *X axis – Y axis cross-correlation geometric mean* for semi-naturalistic data). This is extremely important, because the computational requirements in comparison to other functional methods such as ROC, *T*-test or Wilcoxon are less, if we bear in mind that these three techniques find a different feature for each classifier and every sensor.

Finally, in order to find the best sensor location to design a robust and efficient activity recognition system, all the sensors studied can in general be used, and especially remarkable is the fact that no sensor combinations are needed, as they offer extraordinarily good results separately. Nevertheless, as these wearable systems are ultimately designed to be used by people and should not be a hindrance in their daily living activities, the wrist location is considered to be the most suitable. Actually, the best results for both studies usually derived from the data monitored through this sensor, so it would seem to be the best choice for a possible real end-user system.

5. Conclusions

We have described a direct application of feature selection methods applied to daily physical activity recognition systems. An efficient classification method requires a productive and limited feature set, thus requiring an efficient selection process since the initial set of possible candidates is huge. We have designed a highly accurate feature selector based on statistical discrimination and robustness criteria, with very low computational and resource requirements, which represents a competitive alternative to other selection processes.

Apart from its feature selection capabilities and unlike other feature selectors, the method proposed allows the operator to have an objective idea about the discriminant power of every feature. In other words, the feature selection process can be easily interpreted as a fast pre-classification process, with a demonstrated usefulness for features ranked in the top-quality groups.

Features extracted in combination with the proposed feature selector have proved to be particularly interesting for real time activity recognition systems, since only minimal resources are required to offer remarkable efficiency in several classification approaches, DT and SVM being especially interesting because of their speed and simplicity, and the wrist location for its unobtrusive properties.

Acknowledgements

This work was supported in part by the Spanish CICYT Project TIN2007-60587, Junta de Andalucia Projects P07-TIC-02768 and P07-TIC-02906, the CENIT project AmIVital, of the "Centro para el Desarrollo Tecnológico Industrial" (CDTI- Spain) and the FPU Spanish grant AP2009-2244.

We are grateful to Prof. Stephen S. Intille, Technology Director of the House_n Consortium with the MIT Department of Architecture for the experimental data provided.

References

Amft, O., & Tröster, G. (2008). Recognition of dietary activity events using on-body sensors. *Artificial Intelligence in Medicine, 42*(2), 121–136.

Aminian, K., Robert, P., Buchser, E. E., Rutschmann, B., Hayoz, D., & Depairon, M. (1999). Physical activity monitoring based on accelerometry: Validation and comparison with video observation. *Medical & Biological Engineering & Computing, 37*(3), 304–308.

Arcelus, A., Herry, C. L., Goubran, R. A., Knoefel, F., Sveistrup, H., & Bilodeau, M. (2009). Determination of sit-to-stand transfer duration using bed and floor pressure sequences. *IEEE Transactions on Biomedical Engineering, 56*(10), 2485–2492.

Baek, J., Lee, G., Park, W., & Yun, B.-J. (2004). Accelerometer signal processing for user activity detection. *Knowledge-Based Intelligent Information and Engineering Systems, 3215*, 610–617.

Bao, L., & Intille, S. (2004). Activity recognition from user-annotated acceleration data. *Pervasive Computing, 23*, 1–17.

Begg, R., & Kamruzzaman, J. (2005). A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics, 38*(3), 401–408.

Bouten, C. V. C., Koekkoek, K. T. M., Verduin, M., Kodde, R., & Janssen, J. D. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering, 44*(3), 136–147.

Cristianini, N., & Taylor, J. S. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: Wiley.

Engin, M., Demirag, S., Engin, E. Z., Çelebi, G., Ersan, F., Asena, E., et al. (2007). The classification of human tremor signals using artificial neural network. *Expert Systems with Applications, 33*(3), 754–761.

Ermes, M., Parkka, J., Mantyjarvi, J., & Korhonen, I. (2008). Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine, 12*(1), 20–26.

European Economy Commission Directorate-General for Economic and Financial Affairs (2006). The impact of ageing on public expenditure: Projections for the EU25 Member States on pensions, health care, longterm care, education and unemployment transfers (2004–2050). Special Report. vol. 01.

Fahrenberg, J., Foerster, F., Smeja, M., & Müller, W. (1997). Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. *Psychophysiology, 34*(5), 607–612.

Foerster, F., Smeja, M., & Fahrenberg, J. (1999). Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring. *Computers in Human Behavior, 15*(5), 571–583.

He, Z., Liu, Z., Jin, L., Zhen, L.-X., & Huang, J.-C. (2008). Weightlessness feature – a novel feature for single tri-axial accelerometer based activity recognition. Pattern Recognition, 2008. 19th International Conference on ICPR 2008 (pp. 1–4).

Heinz, E. A., Kunze, K.-Steven, Sulistyo, S., Junker, H., Lukowicz, P., & Tröster, G. (2003). Experimental evaluation of variations in primary features used for accelerometer context recognition. In *Proceedings of the 1st European symposium on ambient intelligence* (pp. 252–263).

Karantonis, D. M., Narayanan, M. R., Mathie, M., Lovell, N. H., & Celler, B. G. (2006). Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine, 10*(1), 156–167.

Kern, N., Schiele, B., & Schmidt, A. (2003). Multi-sensor activity context detection for wearable computing. In *Proceedings of EUSAI, LNCS 2875* (pp. 220–232).

Könönen, V., Mäntyjärvi, J., Similä, H., Pärkkä, J., & Ermes, Miikka (2010). Automatic feature selection for context recognition in mobile devices. *Pervasive and Mobile Computing, 6*(2), 181–197.

Laerhoven, K. V., & Gellersen, H.-W. (2004). Spine versus porcupine: A study in distributed wearable activity recognition. Wearable Computers, 2004. Eighth International Symposium on ISWC 2004 (vol. 1, pp. 142–149).

Lee, S.-W., & Mase, K. (2002). Activity and location recognition using wearable sensors. *IEEE Pervasive Computing, 1*(3), 24–32.

Liu, H. (1998). *Feature selection for knowledge discovery and data mining*. Kluwer Academic.

Mantyjarvi, J., Himberg, J., & Seppanen, T. (2001). Recognizing human motion with multiple acceleration sensors. In *IEEE international conference on systems, man, and cybernetics* (vol. 2, pp. 747–752).

Mathie, M. J., Coster, A. C. F., Lovell, N. H., & Celler, B. G. (2003). Detection of daily physical activities using a triaxial accelerometer. *Medical & Biological Engineering & Computing, 41*(3), 296–301.

Mathie, Merryn J., Coster, Adelle C. F., Lovell, Nigel H., & Celler, Branko G. (2004). Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological Measurement, 25*(2), R1–R20.

Maurer, U., Smailagic, A., Siewiorek, D. P., & Deisher, M. (2006). Activity recognition and monitoring using multiple sensors on different body positions. Wearable and Implantable Body Sensor Networks, 2006. International Workshop on BSN 2006 (vol. 4 pp. 112–116).

Minnen, D., Starner, T., Essa, M., & Isbell, C. (2006). Discovering characteristic actions from on-body sensor data. 2006 *10th IEEE international symposium on wearable computers* (pp. 11–18).

Najafi, B., Aminian, K., Loew, F., Blanc, Y., & Robert, P. A. (2002). Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly. *IEEE Transactions on Biomedical Engineering, 49*(8), 843–851.

Nyan, M. N., Tay, F. E. H., Seah, K. H. W., & Sitoh, Y. Y. (2006). Classification of gait patterns in the time-frequency domain. *Journal of Biomechanics, 39*(14), 2647–2656.

Parera, J., Angulo, C., Rodríguez-Molinero, A., & Cabestany, J. (2009). User daily activity classification from accelerometry using feature selection and SVM. In *Proceedings of the 10th international work-conference on artificial neural networks: Part I: Bio-inspired systems: Computational and ambient intelligence* (pp. 1137–1144).

Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., & Korhonen, I. (2006). Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine, 10*(1), 119–128.

Pirttikangas, S., Fujinami, K., & Nakajima, T. (2006). Feature selection and activity recognition from wearable sensors. *Proceedings of the Ubiquitous Computing Laboratory, 4239*, 516–527.

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering, 56*(3), 871–879.

Preece, Stephen J., Goulermas, John Y., Kenney, Laurence P. J., Howard, Dave, Meijer, K., Crompton, R., et al. (2009). Activity identification using body-mounted sensors – a review of classification techniques. *Physiological Measurement, 30*(4), R1–R33.

Ravi, N., Dandekar, N., Mysore, P., & Littman, M. (2005). Activity recognition from accelerometer data. *American Association for Artificial Intelligence, 3*, 1541–1546.

Sazonov, E.S., Fulk, G., Sazonova, N., & Schuckers, S. (2009). Automatic recognition of postures and activities in stroke patients. Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE (pp. 2200–2203).

Sazonov, E. S., Fulk, G., Hill, J., Schutz, Y., & Browning, R. (2011). Monitoring of posture allocations and activities by a shoe-based wearable sensor. *IEEE Transactions on Biomedical Engineering, 58*(4), 983–990.

Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition*. Fourth ed: Elsevier.

Uiterwaal, M., Glerum, E. B., Busser, H. J., & van Lummel, R. C. (1998). Ambulatory monitoring of physical activity in working situations, a validation study. *Journal of Medical Engineering & Technology, 22*(4), 168–172.

Vapnik, V. (1998). *Statistical learning theory*. John Wiley & Sons.

Wang, S., Yang, J., Chen, N., Chen, X., & Qinfeng Zhang, Q. (2005). Human activity recognition with user-free accelerometers in the sensor networks. Neural Networks and Brain, 2005. ICNN&B '05. International Conference on (vol. 2, pp. 1212–1217).

Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity – a review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European association of cardiovascular prevention and rehabilitation. *European Journal of Cardiovascular Prevention and Rehabilitation, 17*(2), 127–139.

Wu, J., Osuntogun, A., Choudhury, T., Philipose, M., & Rehg, J. M. (2007). A scalable approach to activity recognition based on object use. In *IEEE 11th international conference on computer vision, 2007. ICCV 2007* (pp. 1–8).

Yu, X. (2008). Approaches and principles of fall detection for elderly and patient. e-health Networking, Applications and Services, 2008. In *10th international conference on HealthCom 2008* (pp. 42–47).

Zhang, K., Sun, M., Kevin Lester, D., Xavier Pi-Sunyer, F., Boozer, C. N., & Longman, R. W. (2005). Assessment of human locomotion by using an insole measurement system and artificial neural networks. *Journal of Biomechanics, 38*(11), 2276–2287.

Zwartjes, D. G. M., Heida, T., van Vugt, J. P. P., Geelen, J. A. G., & Veltink, P. H. (2010). Ambulatory monitoring of activities and motor symptoms in Parkinson's disease. *IEEE Transactions on Biomedical Engineering, 57*(11), 2778–2786.