Assessment of Homomorphic Analysis for Human Activity Recognition from Acceleration Signals

Sebastián R. Vanrell, Diego H. Milone, H. Leonardo Rufiner

Abstract

Unobtrusive activity monitoring can provide valuable information for medical and sports applications. In recent years, human activity recognition has moved to wearable sensors to deal with unconstrained scenarios. Accelerometers are the preferred sensors due to their simplicity and availability. Previous studies have examined several classic techniques for extracting features from acceleration signals, including time-domain, time-frequency, frequency-domain, and other heuristic features. Spectral and temporal features are the preferred ones and they are generally computed from acceleration components, leaving the acceleration magnitude potential unexplored. In this study, a new type of feature extraction stage, based on homomorphic analysis, is proposed in order to exploit discriminative activity information present in acceleration signals. Homomorphic analysis can isolate the information about whole body dynamics and translate it into a compact representation, called cepstral coefficients. Experiments have explored several configurations of the proposed features, including size of representation, signals to be used, and fusion with other features. Cepstral features computed from acceleration magnitude obtained one of the highest recognition rates. In addition, a beneficial contribution was found when time-domain and moving pace information was included in the feature vector. Overall, the proposed system achieved a recognition rate of 91.21% on the publicly available SCUT-NAA dataset. To the best of our knowledge, this is the highest recognition rate on this dataset.

Index Terms

human activity recognition, accelerometer, signal processing, cepstrum.

S. R. Vanrell, D. H. Milone, and H. L. Rufiner are with Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL/CONICET, Argentina. Tel.: +54 342 4575233 ext 117; Fax: +54 342 4575224; Email: {srvanrell, dmilone, lrufiner}@sinc.unl.edu.ar

H. L. Rufiner is with Laboratorio de Cibernética, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina

I. Introduction

Research interest in human activity recognition (HAR) has increased in the last two decades. Nowadays, HAR applications spread over physical monitoring [1]–[4], health [1], [5]–[7], entertainment [5], [8], [9], sports [10], security [5], [9], [11], and industry [9], [12]. For instance, activities of patients can be reliably tracked, which can help a physician to counsel behaviors for physical activity and healthy lifestyle [3], [6]. Historically, research in computer vision has been in the vanguard of HAR [5], [9]. However, efforts to recognize activities in unconstrained daily life settings caused a shift toward the use of wearable sensors. Accelerometers, gyroscopes, and compasses are the sensors typically used, either individually or combined. In this work, we will focus on HAR systems that consider accelerations as input signals.

The first studies on activity recognition using accelerometers were published in the middle and late 1990s, with applications in medical assessment [13]. They used unidirectional accelerometers on two or three locations of the body, and applied a set of simple rules to distinguish between dynamic and static activities. At present, HAR systems are based on triaxial accelerometers and they aim to recognize a broader set of activities. These systems rely on pattern recognition, in which the fundamental stages include signal pre-processing, feature extraction, and activity classification. Several proposals for each of these stages have been made to improve said systems [1], [2], [8], [14]–[16].

Ideally, a feature extraction stage should be able to extract all the discriminative information in a compact representation. Discriminative information helps to distinguish activities from one another. For example, the periodicity combined with a spectral description of movements are good candidates. A compact representation helps to keep the recognizer as simpler as possible and avoids high computational costs. Previous studies have explored a wide range of techniques for extracting features from acceleration signals. However, features related to activities are not discriminative enough and others are not compact. In this study a compact representation is proposed and its potential is fully assessed.

Feature extraction techniques explored in previous studies can be grouped by the type of their outputs as heuristic, time-domain, time-frequency, or frequency-domain features [14]. Heuristic features are derived from and characterized by an intuitive understanding of how an activity produces changes in the signal. For example, a static posture such as standing or lying can be recognized by the direction of the recorded gravity [17]. Other examples are signal magnitude area, mean rectified value, peak-to-peak acceleration, and root mean square, which have been related to the intensity of an activity [18].

In contrast to heuristic features, time-domain features are not directly related to specific aspects of individual movements or postures. Instead, they are computed from a windowed signal and they are typically statistical measures. Common examples include mean, median, variance, skewness, kurtosis,

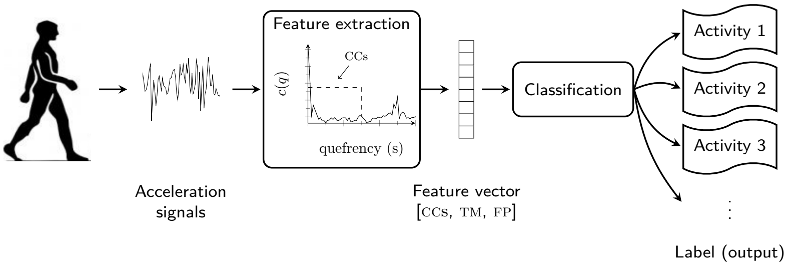

Fig. 1. General diagram of the proposed recognizer with its two main stages: feature extraction and classification. CCs: cepstral coefficients; TM: time-domain measures; FP: fundamental period.

interquartile range, and correlation between accelerometer axes [14], [19], [20]. Recently, a two-directional feature for incremental learning showed promising results [21]. Both time-domain and heuristic features are the simplest to compute, and they were useful for distinguishing between static and dynamic activities. However, they were barely successful in discriminating between dynamic activities [14], [22].

Time-frequency features, such as wavelets, were effective for detecting transitions between activities, but they were outperformed by other features in the activity classification task [23]. By contrast, frequency-domain features are usually the preferred choice in HAR [19], [22], [24]–[27]. These features are obtained using the fast Fourier transform (FFT) or the discrete cosine transform. Basis coefficients may be directly adopted as features or an additional method may be used to characterize the spectral distribution, such as subsets of coefficients, or filter banks. Several studies have reported their best results using FFT coefficients as features, either individually [19], [22], [25], [26] or fused with time-domain features [20], [24], [28]. In these studies, features are usually extracted from each of the acceleration-component signals, leaving the capabilities of acceleration-magnitude signal unexplored, which have processing and storage advantages.

The feature extraction stage proposed in this work uses a well-known technique: the cepstral analysis. It is based on the theory of homomorphic analysis and has been successfully used for characterizing seismic echoes, deblurring images, and recognizing speech [29], but it has not been considered on HAR, except for two studies [30], [31]. However, these studies have neither examined the compliance of hypothesis to apply such analysis nor exploited all the information that cepstral representation can capture. Li *et al.* [31] claimed that signals have quasi-periodic characteristics although the analysis was applied on narrow windows, which cannot encompass that periodicity. In [30], a mel scale was used to map frequency bands in a nonlinear fashion. This mel scale was designed for speech analysis and recognition based on the perceptual scale in the human cochlea [29]. Although this makes no sense for HAR, a high resolution at low frequencies could be useful. In a related study [32], the features proposed in [30] were also used for user identification. For this task, adapted perceptual linear prediction coefficients were the best alternative.

In the present study, the proposed representation explores the capabilities of cepstral coefficients as is (i.e., without using intermediate filter banks or dimension reduction techniques). In addition, it is the first time that body dynamics and activity periodicity are explained in terms of cepstral features. The advantage of the proposed cepstral representation is that it compresses the spectral information of body dynamics in few coefficients while retaining its discriminative power. In the experiments, features extracted from magnitude and components of acceleration were contrasted based on their discriminative power and an exhaustive analysis on the number of required coefficients was performed. Furthermore, we assessed the performance improvements when including time-domain measures and the fundamental period of the signal. To complete the system, a support vector machine was chosen for the classification stage, because it showed better performance compared to other well-known classifiers, such as multilayer perceptron, random forest, and naive Bayes. In addition, the selection of support vector machine allows a fair comparison with previous studies.

The remainder of this paper is organized as follows. Section II describes the proposed HAR system. Section III presents the experimental setup. Results and discussion are given in Section IV. Finally, conclusions are drawn in Section V.

II. Proposed recognition system

As most automatic recognition systems, the proposed recognizer can be seen as a pipeline with two main stages: feature extraction and classification. An schematic diagram of the system based on homomorphic analysis is shown in Figure 1. The first stage will be explained in detail from Section II-A to II-C. The second stage will be covered in Section II-D.

A. Acceleration signals

In this study, a three-dimensional accelerometer is used for capturing information about movements. The captured signals are the acceleration components $a_x(t)$, $a_y(t)$, and $a_z(t)$. In addition, the acceleration magnitude $|\mathbf{a}(t)|$ is computed from components as another input signal. Components are recorded in directions relative to the orientation of the device, thus being altered by gravity when the device rotates. On the contrary, magnitude is unaltered by the orientation of the device because gravity is a constant offset. These signals are processed as described below for a generic signal called $a(t)$.

In what follows the acceleration signal $a(t)$ is considered as the output of a linear convolutive system. The excitation $m(t)$ is associated to the moving pace, which is originated by muscle activity and external body interactions. Body dynamics is modeled in the impulse response $h(t)$, which will depend on the activity that is being performed. Since the goal is to recover information about said activity, the analysis should be able to isolate $h(t)$. However, excitation and impulse response are unknown; therefore recovering $h(t)$ is not straightforward and a blind deconvolution is required. We propose using a homomorphic analysis to carry out the recovering task.

B. Cepstral coefficients

Homomorphic analysis was developed as a general method for separating signals that have been non-additively combined [29]. Homomorphic processing involves converting this mixture into a linear combination, in which the analysis techniques are well understood. For instance, if two signals are convolved in time-domain as

$$a(t) = m(t) * h(t),$$

their Fourier transforms will be multiplied in frequency-domain as

$$A(f) = M(f) \cdot H(f),$$

and an appropriately defined logarithm will produce the sum

$$\log |A(f)| = \log |M(f)| + \log |H(f)|.$$

At this point, $\log |M(f)|$ and $\log |H(f)|$ are additively combined. Taking the inverse Fourier transform of $\log |A(f)|$, a new time-representation,

$$c(q) = \mathcal{F}^{-1} \left\{ \log |A(f)| \right\},$$

can be obtained. This is known as the cepstrum and its domain is the quefrency. At the beginning of $c(q)$, the low-rate variations of $\log |A(f)|$ are codified. The remaining information of $c(q)$ codifies the high-rate variations of $\log |A(f)|$. Generally, low- and high-rate variation components are separated, and they can be linked to the impulse response and the excitation, respectively. Thus, by taking the first coefficients of $c(q)$, the homomorphic analysis can recover the desired information about the underlying system. In the quefrency domain, this operation is known as liftering.

In the time domain, acceleration signals exhibit a quasi-periodic behavior, as exemplified in Figure 2.a. For this example, the period is $\sim$0.5 s, which is the expected moving pace for a daily human activity. This signal gives some insight into the excitation signal $m(t)$, but the impulse response $h(t)$ remains hidden. In the frequency domain, this period can be seen as a peak of fundamental frequency at $\sim$2 Hz and its corresponding harmonics (Figure 2.b). These are the major contributions, but they mask the remaining

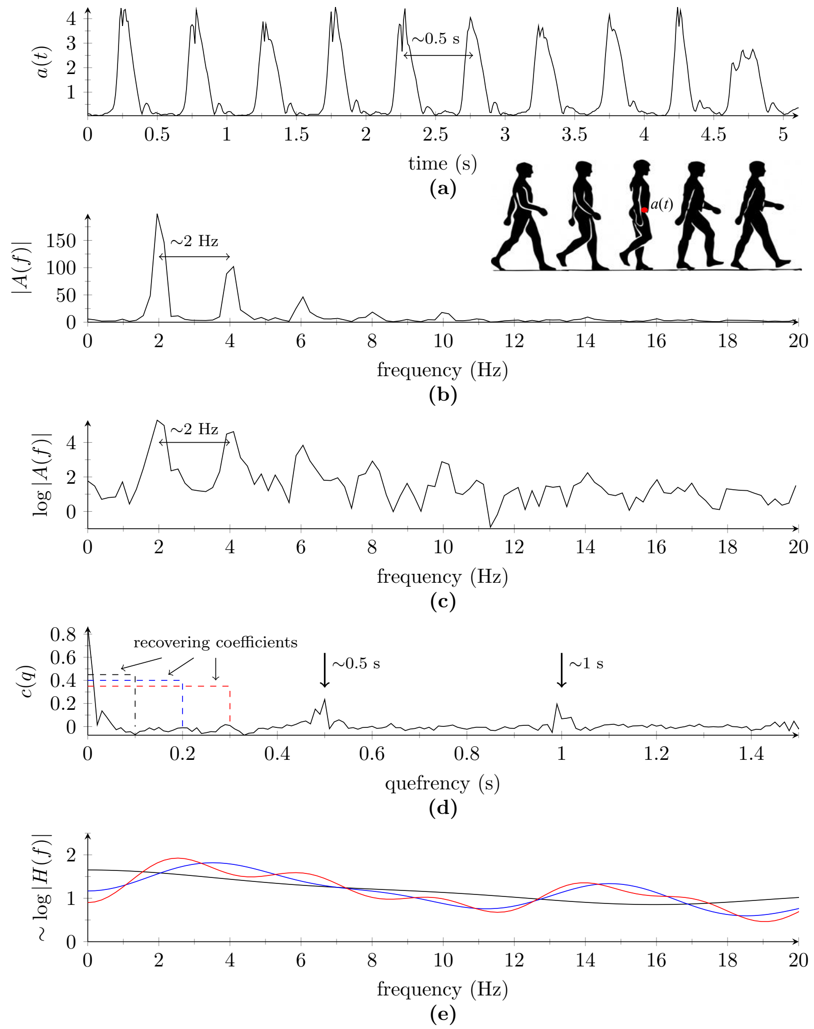

Fig. 2. Computation of cepstral coefficients. (a) Quasi-periodic acceleration signal $a(t)$; (b) spectrum of $a(t)$; (c) logarithm of the spectrum; (d) cepstral coefficients; (e) recovered approximations of $\log|H(f)|$.

information about the impulse response of the system. Figure 2.c shows the signal obtained by applying the logarithm to $|A(f)|$. Afterwards, the inverse Fourier transform is applied and the cepstrum, $c(q)$, is obtained (Figure 2.d). Peaks related to excitation are still present at ~0.5 s and ~1.0 s but no longer overlapped with the cepstrum of impulse response.

For example, if $a(t)$ is captured on the hip of a subject performing an activity, the first coefficients of $c(q)$ codify the dynamics of hip movements. The hip was reported as the best position to represent the whole body, thus the coefficients at the left of the first peak codify the global dynamics of the body [2], regardless of the activity pace. Besides, the separation between peaks in $c(q)$ represents the pace at which the activity is performed.

As explained above, the beginning of $c(q)$ contains the low-rate variations of $\log|A(f)|$, which is the isolated information about $h(t)$. To verify this, a smooth representation of $\log|A(f)|$ can be recovered by applying the Fourier transform only to the first coefficients of $c(q)$, i.e., the transform is applied to a liftered $c(q)$. The lifters used in the recovering are indicated by dashed rectangles in Figure 2.d and the smoothed versions of $\log|A(f)|$ are plotted in Figure 2.e using the respective colors. These curves are estimations of $\log|H(f)|$. It is clear that the number of coefficients defines the degree of details retained in the recovered information about the impulse response of the body dynamics.

Computation of cepstral coefficients requires the application of a Fourier transform, a logarithm and an inverse Fourier transform. These transformations require $O(n\log(n))$, $O(n)$, and $O(n\log(n))$, respectively, where n is the length of the signal analyzed. Thus, cepstral features computation is of the same order as any other FFT-based feature, which requires $O(n\log(n))$.

C. Feature vector composition

In this study, the analysis of a signal $a(t)$ may lead to three kinds of features: cepstral coefficients, time-domain measures, and fundamental period. Cepstral coefficients (CCs) refer to the first coefficients of $c(q)$, up to a predefined quefrency. Time-domain measures (TM) are five values computed directly from $a(t)$: standard deviation, energy, maximum amplitude, minimum amplitude, and peak-to-peak amplitude. The fundamental period (FP) contains a value that represents the moving pace.

The example in the previous section shows that FP is revealed as one or more peaks in $c(q)$ (Figure 2.d). However, there are activities that present an unclear or absent peak. Therefore, we chose to extract FP with the unbiased autocorrelation of $a(t)$ [33]. To be considered as a periodical activity, the amplitude of autocorrelation must exceed a given threshold. In such a case, the lag corresponding to the first local maximum is assigned to FP. On the contrary, FP will be set to zero, which means that no periodicity was detected.

As $a(t)$ can be $a_x(t)$, $a_y(t)$, $a_z(t)$, or $|\mathbf{a}(t)|$, each of these signals can generate its own set of features. Thus, different combinations could be selected to define the final feature vector. Tables I and II summarize the combinations evaluated in this study. Vectors composed by cepstral coefficients of $a_x(t)$, $a_y(t)$, and $a_z(t)$ are given in Table I. Vectors composed by cepstral coefficients of $|\mathbf{a}(t)|$ are presented in Table II. These tables present the names assigned to each composition, which will be used in Section IV. Combinations were selected to perform a comparative study which can assess the partial contribution of each feature. First, the study will evaluate the influence of the number of CCs, without additional features (F_1 in Table I and F_{17} in Table II). Then, the contributions of temporal and pace information will be considered in F_2, F_3 and F_4 (Table I), and F_{18}, F_{19} and F_{20} (Table II). The study will also consider the

TABLE I

COMPOSITION OF FEATURE VECTORS WHEN CEPSTRAL COEFFICIENTS (CC) ARE OBTAINED FROM $a_x(t)$, $a_y(t)$, AND $a_z(t)$. FP: FUNDAMENTAL PERIOD. TM: TIME-DOMAIN MEASURES.

Feature	Signal	Composition of feature vectors															
		F$_1$	F$_2$	F$_3$	F$_4$	F$_5$	F$_6$	F$_7$	F$_8$	F$_9$	F$_{10}$	F$_{11}$	F$_{12}$	F$_{13}$	F$_{14}$	F$_{15}$	F$_{16}$
CC	a_*	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
TM	$\|\mathbf{a}\|$		●		●		●		●		●		●		●		●
FP	$\|\mathbf{a}\|$			●	●			●	●			●	●			●	●
TM	a_*					●	●	●	●					●	●	●	●
FP	a_*									●	●	●	●	●	●	●	●

TABLE II

COMPOSITION OF FEATURE VECTORS WHEN CEPSTRAL COEFFICIENTS (CC) ARE OBTAINED FROM $|\mathbf{a}(t)|$. FP: FUNDAMENTAL PERIOD. TM: TIME-DOMAIN MEASURES.

Feature	Signal	Composition of feature vectors															
		F$_{17}$	F$_{18}$	F$_{19}$	F$_{20}$	F$_{21}$	F$_{22}$	F$_{23}$	F$_{24}$	F$_{25}$	F$_{26}$	F$_{27}$	F$_{28}$	F$_{29}$	F$_{30}$	F$_{31}$	F$_{32}$
CC	$\|\mathbf{a}\|$	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
TM	$\|\mathbf{a}\|$		●		●		●		●		●		●		●		●
FP	$\|\mathbf{a}\|$			●	●			●	●			●	●			●	●
TM	a_*					●	●	●	●					●	●	●	●
FP	a_*									●	●	●	●	●	●	●	●

influence of the signal which gave origin to a feature. For instance, F$_2$ and F$_5$ are comprised by the same CCs but differ on the source signals to compute TM ($|\mathbf{a}|$ or a_*). Finally, combined effects will be studied by considering all the combinations of these variants.

The above description assumes that there is only one three-dimensional acceleration signal of short duration. However, each of these signals may be of long and variable duration. Thus, a collection of acceleration signals $\mathbf{a}^1(t), \mathbf{a}^2(t), \cdots$ is obtained from a raw acceleration data by a fixed-size sliding window, with overlap between consecutive windows (top of Figure 3). Therefore, the feature extraction process generates a sequence of feature vectors $X = \left\{\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \cdots, \mathbf{x}^L\right\}$, where L is the number of windows (bottom of Figure 3).

D. Classification

In the classification stage, a well-known static classifier is used: the support vector machine (SVM) [34]. This classifier has proven to be one of the best machine-learning techniques for binary classification

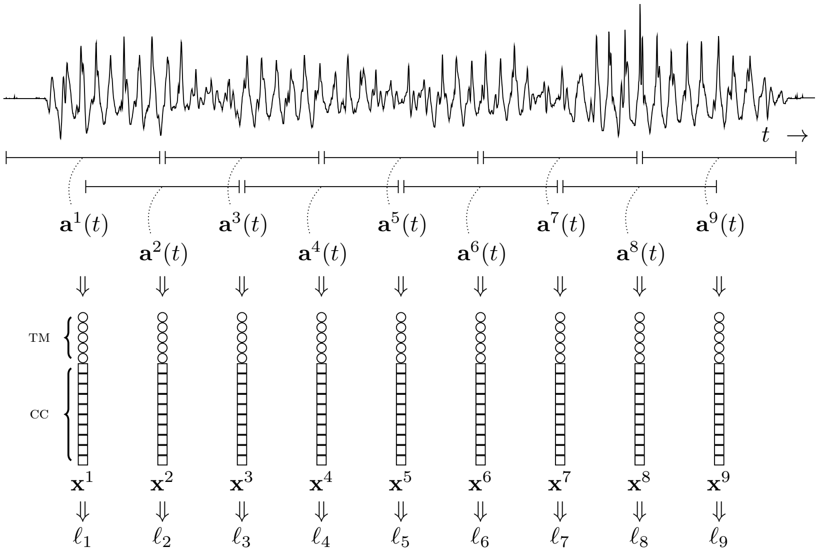

Fig. 3. Sequence of feature vectors generated from a raw acceleration signal. In this example, feature vectors are composed by time-domain measures (TM) and ten cepstral coefficients (CC) extracted from the acceleration magnitude.

problems [35]. First, input features are non-linearly mapped to a very high-dimension feature space. The mapping is done using a kernel function. Then, a linear decision surface is built in this new feature space. That surface is a hyperplane and it is located so that the maximum margin of separation is achieved between the two classes (see Hastie *et al.* [36] for a detailed explanation). In this study, a radial basis function $K(\mathbf{x}, \mathbf{y}) = e^{-\gamma|\mathbf{x}-\mathbf{y}|^2}$ was chosen as a kernel, where $\mathbf{x}$ and $\mathbf{y}$ are feature vectors. Other kernel choices were discarded in preliminary experiments. Also, a soft margin penalty for misclassifications was considered. Penalty coefficient C and parameter γ were optimized with a logarithmic grid search on the training dataset. In the proposed system, the number of classes is equal to the number of activities to recognize. For this multiclass classification task, the one-against-one approach was followed [37], [38].

An SVM requires a number of operations during training that is different from classification time. During training, complexity of an SVM depends greatly on the number of vectors in the dataset, rather than the dimension of those vectors. In addition, SVM can perform very well with relatively large feature vectors. During classification, the number of required operations is linear with the number of support vectors and the dimension of feature vectors.

Feature selection may be included prior to classification and there is a vast diversity of techniques proposed in the literature for this task. For instance, there are techniques that can be applied to any kind of classifiers [39] and others that are embedded and specifically designed for SVM [40]. However, there is no clear definition on which technique is suitable in advance. In general, it depends on chosen classifier and dataset characteristics, such as number of examples and inter-relationship among features. A thorough evaluation of selection methods with cepstral representation falls outside the scope and extension of the present study.

As was previously described, a raw acceleration data corresponding to a single activity is translated into a sequence of feature vectors $X = \left\{\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^L\right\}$. SVM classifies each vector $\mathbf{x}^i$ independently from one another and assigns an activity label ℓ_i to each one. Then, the label count is computed by activity, and the label that corresponds to the maximum count is assigned to the input acceleration data.

III. Experimental setup

The recognition experiments were carried out on the publicly available SCUT-NAA dataset [22], which contains 1278 records from 44 subjects collected in naturalistic settings. During the recording sessions, subjects were asked to perform ten activities, one time each, thereby being a dataset with balanced classes. Each recorded signal represented a single activity (i.e., the provided signals were already segmented by activity), with sample frequency fixed at 100 Hz. Subjects wore a triaxial accelerometer on their waist belts, trouser pockets, and shirt pockets, alternately. Previous studies have demonstrated that the waist is the best location for a single sensor because it can better represent the major human motion [2]. Thus, only records captured on waist belt position will be used in the following experiments. Features extracted considering this position will contain information of the whole body dynamics.

Features were extracted from raw acceleration signals using a window size of 5.12 s. The overlapping between consecutive windows was 50%. Signals were sampled at 100 Hz, hence, windows size was of 512 samples, with 256 samples overlapping between consecutive windows. Each segment was smoothed with a Hamming window. Several cepstral lengths were tested: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 s (in quefrency), which correspond to $n_{cc} = \{10, 20, 30, \cdots, 90\}$ coefficients, respectively. Regardless of final cepstrum length, time-domain signals were transformed with a 512-points FFT. A previous study [31] performed cepstral analysis with narrow windows (<1 s) that cannot capture a single period of an activity. Also, the resolution of cepstral representation was not explored. Authors in [30] tested different length of their speech-adapted representation but the performance of isolated acceleration features was not studied.

In this study, leave-one-subject-out cross-validation was used to conduct the experiments, which allows a fair comparison with previous results on the same dataset. In the first fold of this scheme, signals from subject 1 were taken for testing, while the remaining signals (from the other subjects) were used for selecting the best parameters of the classifier and for training the models. The following folds switch the test signals to another subject until all subjects are considered in the test. As in previous studies [19],

[22], activity recognition rate will be used as a performance measure, which is simply defined as the number of correctly classified signals over the total number of classified signals. Also, the standard recall ($t_p/[t_p + f_n]$) and precision ($t_p/[t_p + f_p]$) are reported on confusion matrices, where t_p are true positives, f_p false positives, and f_n false negatives counts for class c. In addition to standard measures, the relative error reduction will be used to compare the performance of different systems. This measure is useful when the recognition rates are close to 100%. For two recognizers with absolute errors ϵ_A and ϵ_B, where A is the reference system, the relative error reduction is $\delta = (\epsilon_B - \epsilon_A)/\epsilon_A$.

IV. Results and discussion

The following subsections present a detailed analysis of the results obtained. First, feature vector compositions are compared based on their discriminative power. Second, these compositions are distinguished from other approaches based on cepstral analysis. Third, the proposed system is compared with state-of-the-art systems. Fourth, a related recognition task that joins similar activities is formulated. For this modified task, the proposed system and some adaptations are evaluated and discussed.

A. *Comparative analysis of feature vector compositions*

Feature vectors can comprise features extracted from $|\mathbf{a}(t)|$, $a_x(t)$, $a_y(t)$, and $a_z(t)$. A comparative study was conducted to select the most suitable composition. This included the influence of the number of CCs, and the contributions of temporal and pace information. Both points will be addressed in the following paragraphs.

The influence of the number of CCs on the recognition rate per activity is shown in the violin plots of Figure 4 for F_1 and F_{17}. The composition F_1 corresponds to the CCs obtained from the acceleration components and F_{17} corresponds to the CCs obtained from the acceleration magnitude (Tables I and II). As more coefficients are selected (from 10 to 90), the mean of recognition rates improves and the distributions get narrower. Particularly, improvements are obtained with up to 50 coefficients. A possible explanation is that this resolution of the cepstral representation can capture most of the discriminative information in the impulse response. Also, discriminative cues can be associated with the first periodicity peak that typically appears between the 35th and the 50th cepstral coefficient (i.e., 0.35 and 0.50 s at 100 Hz sampling frequency). Using more coefficients showed no significant improvements. Hence, up to 70 CCs will be considered in the following experiments.

The possible compositions were expanded when other features are appended to CCs. The results for all of the proposed compositions when CCs are extracted from $a_x(t)$, $a_y(t)$, and $a_z(t)$ are compiled in Table III. The number of CCs is specified in the header of the table. The results in each row correspond

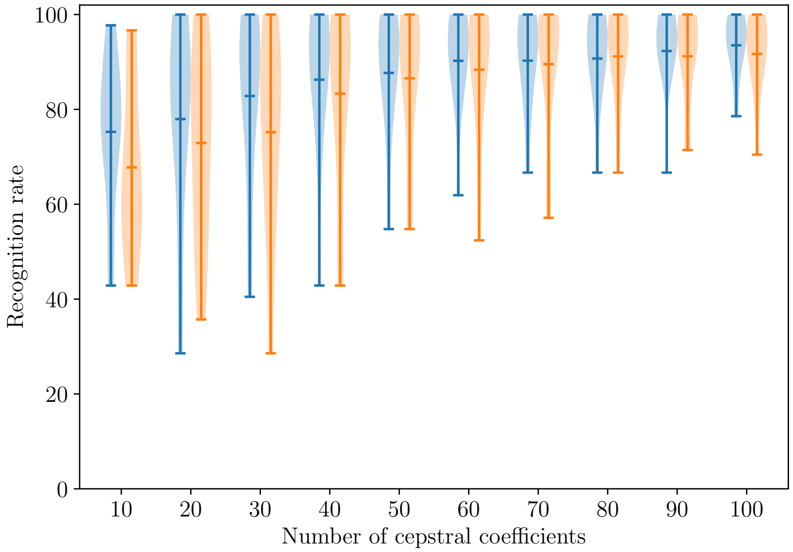

Fig. 4. Violin plots of the recognition rate per activity versus the number of CCs. Feature vector composition: (blue) CCs of the acceleration components (F_1), (orange) CCs of the acceleration magnitude (F_{17}).

to an exclusive composition that only differs in the number of CCs. The improvements in recognition rates are clearly higher when TM or FP are appended to a small number of CCs, because they compensate the lack of discriminative power of low-resolution cepstral representations. However, when more CCs are used, the cepstral representation by itself achieves high recognition rates and the contribution of other features is reduced. For example, if 10 CCs are used (F_1), the recognition rate is 74.82%, and the rate improves to 83.85% when TM and FP (F_4) are added. However, if 70 CCs are selected (F_1), the recognition rate is 90.02%, and no improvement is observed when TM and FP (F_4) are added. In the absence of TM, FP helps to distinguish between activities, achieving a high rate of 90.97% for 70 CCs (F_{19}).

Table IV presents the results for CCs extracted from $|\mathbf{a}(t)|$ (F_{17}-F_{32}). Clear improvements are achieved by appending TM or FP obtained from $|\mathbf{a}(t)|$ to CCs (compare the results for F_{18}-F_{20} with F_{17}). Time-domain measures help to avoid confusion between activities that are clearly different in their waveform. For example, jumping involves a high acceleration that must oppose to gravity. Thus, the recorded signal $|\mathbf{a}(t)|$ will exhibit a great amplitude and energy that can be easily distinguished from the other activities. However, adding TM slightly improves the discriminability of similar activities. For example, activities such as walking and walking quickly are expected to be very similar in their temporal and cepstral representations. Fundamental period can be of help in this situation, when repetitive actions are performed at different paces. As a result, it is observed that, regardless of the number of CCs, the recognition rates are higher when TM and FP (F_{20}) are combined than when they are used individually.

TABLE III

ACTIVITY RECOGNITION PERFORMANCE WHEN CCS ARE OBTAINED FROM $a_x(t)$, $a_y(t)$, AND $a_z(t)$.

Features (size)	n_{cc}: Number of cepstral coefficients						
	10	20	30	40	50	60	70
F$_1$ (0+3n_{cc})	74.82	77.43	82.42	85.99	87.41	**90.02**	90.02
F$_2$ (5+3n_{cc})	78.86	80.76	87.17	86.94	87.65	89.07	**89.31**
F$_3$ (1+3n_{cc})	79.57	82.66	85.51	88.12	88.60	90.74	**90.97**
F$_4$ (6+3n_{cc})	83.85	84.09	88.36	87.89	88.84	**90.02**	90.02
F$_5$ (15+3n_{cc})	80.05	84.32	85.75	87.41	87.89	**89.31**	89.31
F$_6$ (20+3n_{cc})	80.05	83.61	85.75	87.41	87.89	87.41	**88.60**
F$_7$ (16+3n_{cc})	83.37	86.94	88.12	89.31	88.60	89.07	**89.79**
F$_8$ (21+3n_{cc})	83.85	86.94	88.36	88.60	88.60	88.12	**90.12**
F$_9$ (3+3n_{cc})	79.81	84.56	84.32	86.22	88.84	89.79	**90.26**
F$_{10}$ (8+3n_{cc})	81.47	84.80	86.46	87.41	87.65	**89.07**	89.07
F$_{11}$ (4+3n_{cc})	80.29	84.32	86.46	87.41	89.79	**90.26**	90.26
F$_{12}$ (9+3n_{cc})	81.00	85.75	87.89	88.84	88.12	89.31	**89.55**
F$_{13}$ (18+3n_{cc})	83.61	85.99	86.70	86.70	**88.84**	88.60	88.84
F$_{14}$ (23+3n_{cc})	83.14	85.99	87.17	87.65	88.36	88.60	**88.84**
F$_{15}$ (19+3n_{cc})	84.32	86.94	88.36	87.41	88.84	88.84	**89.79**
F$_{16}$ (24+3n_{cc})	82.90	87.41	88.36	88.12	88.60	89.07	**89.79**

As presented in Table III, the improvements in recognition rates are greater if TM and FP are appended to low-resolution cepstral representations rather than high-resolution representations. It is evident that discriminative information of both features is overlapped, since the improvement is lower when they are combined (F$_{20}$) than when they are used individually (F$_{18}$ or F$_{19}$). In addition, a recognition rate of 90.02$\%$ is an interesting result for a composition (F$_{20}$) that uses features exclusively obtained from $|\mathbf{a}(t)|$, because this signal is independent of the orientation of the recording device. Moreover, this result is comparable to the ones presented in Table III, although it was obtained with a smaller feature vector (corresponding sizes are specified in Supplementary Material).

Additional information can be included in feature vectors by appending TM and FP from acceleration components (F$_{21}$-F$_{32}$). For example, activities such as walking and walking upstairs involve only one direction, thereby affecting only one particular component of the acceleration. Walking primarily modifies the acceleration in the axis related to forward movement, while walking upstairs affects axes related to vertical and forward movements. Thus, temporal information of acceleration components could reveal unseen differences between activities. This may explain why higher recognition rates are achieved with F$_{22}$, F$_{24}$, and F$_{28}$ compared to F$_{20}$. Although compositions that use TM and FP obtained from components

TABLE IV

Activity recognition performance when ccs are obtained from $|\mathbf{a}(t)|$.

Features (size)	n_{cc}: Number of cepstral coefficients						
	10	20	30	40	50	60	70
F_{17} (0+n_{cc})	66.98	72.21	74.58	82.90	86.22	88.12	**89.31**
F_{18} (5+n_{cc})	75.53	83.37	85.99	85.51	88.12	87.89	**88.36**
F_{19} (1+n_{cc})	76.72	80.52	83.85	84.32	87.65	88.60	**88.84**
F_{20} (6+n_{cc})	82.66	85.04	87.17	87.65	88.84	89.07	**90.02**
F_{21} (15+n_{cc})	80.76	84.56	84.56	85.51	86.70	88.84	**89.31**
F_{22} (20+n_{cc})	82.66	85.75	84.80	87.65	88.84	89.31	**90.97**
F_{23} (16+n_{cc})	83.61	88.84	86.70	87.65	88.12	**90.02**	88.84
F_{24} (21+n_{cc})	84.09	87.41	87.41	88.36	90.02	90.50	**91.21**
F_{25} (3+n_{cc})	76.01	82.19	83.61	85.04	88.36	**89.79**	88.84
F_{26} (8+n_{cc})	80.05	85.51	86.94	87.65	**89.07**	89.07	89.07
F_{27} (4+n_{cc})	75.77	82.66	84.56	85.99	88.84	**89.31**	89.31
F_{28} (9+n_{cc})	81.95	86.46	88.84	88.60	**90.50**	90.50	90.02
F_{29} (18+n_{cc})	85.27	86.70	88.36	85.99	87.17	**89.31**	88.84
F_{30} (23+n_{cc})	84.56	87.89	88.36	89.07	**89.79**	88.84	88.12
F_{31} (19+n_{cc})	85.27	87.89	87.65	87.89	87.17	**89.07**	88.36
F_{32} (24+n_{cc})	83.61	87.65	88.36	88.60	**90.50**	89.31	89.31

(F_{29}-F_{32} in Table IV) achieve good recognition rates, they are not as good as the ones mentioned above. Specifically, said compositions show their best results for 50 and 60 ccs. This may be explained by the combined contributions of tm and fp from components, which compensate the loss from the use of a slightly coarse cepstral representation.

The best recognition rate (91.21%) was achieved with the feature vector F_{24}. Other well-known classifiers such as Naive Bayes, Random Forest, and Multilayer Perceptron were tested but none of them showed better results than SVM. Table V compare activity recall and overall recognition rate for F_{24} composition and these classifiers. This composition combines the best of all the proposed features. It has the best cepstral representation (70 ccs), a complete compendium of temporal information (tm from the acceleration magnitude and components), and the pace at which the activity was performed (fp from $|\mathbf{a}(t)|$). The second best result (90.97%) corresponds to a similar composition (F_{22}), which only lacks the latter feature. Finally, it is remarkable the good result (90.02%) obtained with F_{20} composition, which comprises features extracted from a single signal ($|\mathbf{a}(t)|$). The computational cost of processing a single signal is a clear advantage.

TABLE V

PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS FOR FEATURE VECTOR F_{24} WITH 70 CCS.

Activity	SVM	NaiveBayes	Random Forest	MLP
Step walking (s)	**97.73**	13.64	90.91	95.45
Jumping (j)	**100.00**	**100.00**	**100.00**	**100.00**
Bicycling (b)	**100.00**	36.67	**100.00**	**100.00**
Walking (w)	**75.00**	0.00	65.91	**75.00**
Walking backward (wb)	**90.70**	83.72	69.77	**90.70**
Walking quickly (wq)	**76.19**	2.38	66.67	73.81
Running (r)	**100.00**	93.18	**100.00**	**100.00**
Relaxing (re)	97.73	**100.00**	97.73	97.73
Downstairs (d)	**86.36**	72.73	81.82	81.82
Upstairs (u)	**90.70**	81.40	81.40	79.07
Overall	**91.21**	59.14	85.04	89.07

B. Discussion about other approaches based on cepstral analysis

In the present study, the cepstral coefficients were exclusively extracted from acceleration signals and an exhaustive analysis on the number of required coefficients was performed. In addition, the inclusion of time-domain and moving pace information was also evaluated. As already mentioned, two earlier studies used cepstral analysis for feature extraction. However, those approaches presented some important differences with our study such as they used extra input signals in addition to the acceleration, their data was captured in a laboratory environment, and actual feature vectors are different from ours.

In [30], body and gravitational acceleration, and gyroscope signals were the inputs to the recognizer and they were fused at feature level. This study considered only 3 dynamic activities. Also, it assumed that body and gravitational accelerations can be separated by low-pass filtering from recorded acceleration components. However, components were recorded in directions relative to the orientation of the device, thus being altered by gravity when the device rotates. In addition, proposed features compute frequency band energies using an stretched mel scale. This scale was designed for speech analysis based on the perceptual scale in the human ear, which makes no sense for HAR. In contrast, in the present study cepstral coefficients were computed without any intermediate filter bank. Moreover, body dynamics was explained in terms of cepstral features and a dataset with a vast diversity of dynamic activities was considered.

In a related study [32], the same features proposed in [30] were evaluated for a completely different task: user identification. The authors have found that adapted perceptual linear prediction (PLP) coefficients yielded the lowest error rate for user recognition. Computation of PLP coefficients consider frequency bands defined by a filter bank with a Bark distribution [33], similar to the mel scale discussed before. Since mel and Bark scales were designed for audio signals, it would be interesting to explore the design of a filter bank for acceleration signals. Future works could evaluate evolutionary approaches [41] to obtain an optimized filter bank for activity recognition.

In [31], acceleration and electrocardiogram signals were separately analyzed and fused at decision level. Feature vectors were obtained applying heteroscedastic linear discriminant analysis on normalized cepstral features and their first derivative. Thus, actual features were derived from cepstral analysis but they are not cepstral coefficients. Proposed systems were trained in a subject-dependent fashion, considering 5 subjects in the experiments. By contrast, the present study evaluated subject-independent recognizers, considering 44 subjects in the experiments. Finally, the feature extraction proposed in [31] was applied on too narrow windows (such as 0.48 s) that cannot reveal the quasi-periodic characteristics of human activities in cepstral domain. Our system used a longer window (approx. 5 s) that can exhibit several periods of an activity, which makes sense to capture that periodicity.

C. Comparison with previous results on the same dataset

The comparative analysis was performed using the feature vector F_{24} with 70 CCs, a composition that was selected as the best proposal. This vector comprises 91 features, 76 of which are computed from the acceleration magnitude and the remaining 15 from the acceleration components. The features from acceleration components boost the magnitude features by considering extra information related to axis signals. The comparison will be made with the best system proposed by Xue *et al.* [22], which obtains the best recognition rate reported on the SCUT-NAA dataset. The system uses an SVM as the classifier and a feature vector based on the spectrum of acceleration components. The feature vector has a total of 189 attributes, comprising 63 FFT coefficients obtained from each signal, $a_x(t)$, $a_y(t)$, and $a_z(t)$. SVM with a radial basis function was used for a fair comparison with previous studies. As in the present proposal, features are extracted from raw acceleration data using a fixed-size sliding window with overlap between consecutive windows.

In Table VI, the proposed system and the reported system are compared by recognition error rate. A comparison with other features is not shown for the sake of space. Features like discrete cosine transform, time-domain features, and autoregressive coefficients were included in [22], with the same experimental setup. The recognition error per activity was highly reduced by the system based on CCs, between 25% and 100%, except for walking and walking quickly. Notably, three activities (jumping, bicycling, and running) were perfectly recognized. The overall recognition error was reduced by 33%. The statistical

TABLE VI

GLOBAL PERFORMANCE COMPARED BY RECOGNITION ERROR.

Activity	Xue *et al.* [22]	F_{24} with 70 CCs	Relative error reduction (%)
Step walking (s)	13.64	**2.27**	83.36
Jumping (j)	2.33	**0.00**	100.00
Bicycling (b)	16.67	**0.00**	100.00
Walking (w)	**22.73**	25.00	–
Walking backward (wb)	18.18	**9.30**	48.84
Walking quickly (wq)	**18.18**	23.81	–
Running (r)	2.27	**0.00**	100.00
Relaxing (re)	4.55	**2.27**	50.1
Downstairs (d)	18.18	**13.64**	24.97
Upstairs (u)	15.91	**9.30**	41.55
Overall	13.18	**8.79**	33.31

significance of this reduction was verified with a binomial test performed for the overall recognition rate [42]. Specifically, the binomial test rejected the null hypothesis of both overall rates being equal with a p-value of 0.016.

The confusion matrix (Table VII) shows that walking and walking quickly were misclassified because the system usually confused between them (low precision and recall). Nevertheless, this confusion is not serious since both activities are expected to be very similar, only differing in their pace. Indeed, an activity that may be considered as walking quickly for one subject may be just normal walking for others. This may explain why the system does not distinguish between them correctly. The remaining activities were classified without important mistakes, as shown by the almost diagonal confusion matrix.

The comparative analysis indicates that the proposed recognition system based on CCs outperformed other reported systems. The proposed feature vector combines a variety of information (cepstral, temporal, and pace), therefore it is more robust than a single type of spectral information. Cepstral features give a compact representation of the impulse response (body dynamics) but use a smaller number of coefficients than spectral representations. For instance, 70 CCs are used, in comparison to the 189 FFT coefficients used in [22]. Also, higher recognition rates were obtained by extracting cepstral features from a single signal ($|\mathbf{a}(t)|$), rather than computing frequency-domain features for each of the acceleration components. Adding temporal and pace information prevents confusions based on cepstral information, because they show other aspects of the signals that are not necessarily present in the cepstral or spectral domain. Finally, temporal features from acceleration components add the final piece of information for a minimum extra

TABLE VII

Confusion matrix of feature vector F_{24} with 70 CCs.

Activity	s	j	b	w	wb	wq	r	re	d	u	Recall
Step walking (s)	43		1								97.7
Jumping (j)		43									100
Bicycling (b)			30								100
Walking (w)				33	2	5			1	3	75.0
W. backward (wb)				2	39	1				1	90.7
W. quickly (wq)				6	1	32			2	1	76.2
Running (r)							44				100
Relaxing (re)			1					43			97.7
Downstairs (d)					1		3		38	2	86.4
Upstairs (u)					1				3	39	90.7
Precision	100	100	93.7	80.5	88.6	84.2	93.6	100	86.4	84.8	

TABLE VIII

Confusion matrix of feature vector F_{24} with 70 CCs on the walking activities joined.

Activity	s	j	b	w+wq	wb	r	re	d	u	Recall
Step walking (s)	43		1							97.73
Jumping (j)		43								100.0
Bicycling (b)			30							100.0
W. forward (w+wq)				80	1			2	3	93.02
W. backward (wb)				3	39				1	90.70
Running (r)						44				100.0
Relaxing (re)			1				43			97.73
Downstairs (d)					1	3		38	2	86.36
Upstairs (u)				1	1			3	38	88.37
Precision	100	100	93.7	95.2	92.9	93.6	100	88.4	86.4	

cost.

D. Recognition with walking activities joined

In these experiments, the recognition task was slightly modified. On the assumption that walking and walking quickly are very similar activities, they were joined in a single walking-forward activity. The outputs of previously discussed systems can be post-processed to deal with the new problem, with no need for retraining. If the system has classified a signal as walking or walking quickly, then this signal will be labeled as walking forward. Hence, only 9 activities can be recognized in the new recognition task.

An overall error rate of 10.35% was obtained using the system based on FFT (21.5% of relative error reduction compared to the 10-activity task). The recognition error of walking forward was 6.82% and the recognition errors of the remaining activities were preserved as shown in Table VI. Regarding the system based on CCs and considering the confusion matrix presented in Table VII, the overall error was reduced to 6.18% (29.7% of relative error reduction compared to the 10-activity task). In this case, the recognition error of walking forward was 11.63%. Therefore, when walking and walking quickly are joined, the overall error rate of both systems is significantly reduced.

A different approach for solving the 9-activity task involves building a new classification model, i.e., training a new SVM. Under this scheme, the SVM is trained to recognize 9 activities, including walking forward. This approach assumes that a single model can be a good representation of the walking-forward activity defined above. The results are summarized in the confusion matrix of this approach presented in Table VIII. In the 10-activity task (Table VII), signals that correspond exclusively to walking or walking quickly were misclassified 10 times. By contrast, in the 9-activity task, the same signals that correspond to walking forward were misclassified only 6 times (Table VIII). This comparison demonstrates that joining the activities is a good strategy. As a global result, the newly trained recognizer for 9 activities and the feature vector F_{24} with 70 CCs identified walking forward with a 6.98% error, and it increased precision and recall above 93%. Moreover, the new recognizer reduced the overall error from 6.18% to 5.46%, i.e., a relative error reduction of 11.7% compared to the best 10-activity task recognizer.

V. Conclusions

In this study, a human activity recognition system was developed based on the homomorphic analysis of acceleration signals. Body dynamics was captured by a single accelerometer and it was then translated into a compact representation known as cepstrum. The advantage of this representation is that it compresses spectral information while retaining its discriminative power. High recognition rates were achieved with this exclusive representation, which were later improved by the fusion of cepstral, temporal, and pace information at feature level. Error recognition rates were reduced from 100.00% for some activities to 24.97% for others. Therefore, the overall recognition error rate was reduced by 33.31%. The highest recognition levels were achieved with cepstral, temporal, and pace features extracted from acceleration magnitude, combined with temporal features from acceleration components. This is an advantage since acceleration magnitude is known for its independence of the orientation of the recording device (relative to the human body), and the temporal features from acceleration components are not expensive in computational terms.

Future research includes extending the system to work with continuously recorded signals, which requires modeling the long-term dynamics of the system and the automatic segmentation of activities.

ACKNOWLEDGMENTS

This work has been funded by Universidad Nacional del Litoral, PACT CAID 2011-525.

REFERENCES

[1] Y. Meng and H.-C. Kim, "A review of accelerometer-based physical activity measurement," in *Proceedings of the International Conference on IT Convergence and Security 2011*, Springer, 2012, pp. 223–237.

[2] C.-C. Yang and Y.-L. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.

[3] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 156–167, 2006.

[4] A. Dalton and G. ÓLaighin, "Comparing supervised learning techniques on the task of physical activity recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 46–52, 2013.

[5] R. Poppe, "A survey on vision-based human action recognition," *Image and vision computing*, vol. 28, no. 6, 976–990, 2010.

[6] A. Nazábal, P. García-Moreno, A. Artés-Rodríguez, and Z. Ghahramani, "Human activity recognition by combining a small number of classifiers," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1342–1351, 2016.

[7] 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.

[8] Ó. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[9] J. Aggarwal and M. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, 16:1–16:43, 2011.

[10] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 20–26, 2008.

[11] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, and Y. Y. Tang, "Person re-identification by dual-regularized KISS metric learning," *IEEE Transactions on Image Processing*, vol. 25, no. 6, 2726–2738, 2016.

[12] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, 1553–1567, 2006.

[13] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon, "Physical activity monitoring based on accelerometry: Validation and comparison with video observation," *Med. Biol. Eng. Comput.*, vol. 37, no. 3, pp. 304–308, 1999.

[14] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, “Activity identification using body-mounted sensors—a review of classification techniques,” *Physiol. Meas.*, vol. 30, no. 4, R1, 2009.

[15] A. Bulling, U. Blanke, and B. Schiele, “A tutorial on human activity recognition using body-worn inertial sensors,” *ACM Comput. Surv.*, vol. 46, no. 3, p. 33, 2014.

[16] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, “A survey of online activity recognition using mobile phones,” *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015.

[17] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, “Direct measurement of human movement by accelerometry,” *Medical Engineering & Physics*, vol. 30, no. 10, pp. 1364–1386, 2008.

[18] M. J. Mathie, B. G. Celler, D. N. H. Lovell, and A. C. F. Coster, “Classification of basic daily movements using a triaxial accelerometer,” *Med. Biol. Eng. Comput.*, vol. 42, no. 5, pp. 679–687, 2004.

[19] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in *Pervasive Computing*, Springer, 2004, pp. 1–17.

[20] R. San-Segundo, J. Lorenzo-Trueba, B. Martínez-González, and J. M. Pardo, “Segmenting human activities based on HMMs using smartphone inertial sensors,” *Pervasive and Mobile Computing*, 2016.

[21] D. Tao, Y. Wen, and R. Hong, “Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1124–1134, 2016.

[22] Y. Xue and L. Jin, “A naturalistic 3D acceleration-based activity dataset amp; benchmark evaluations,” in *IEEE International Conference on Systems Man and Cybernetics*, 2010, pp. 4081–4085.

[23] S. Preece, J. Goulermas, L. Kenney, and D. Howard, “A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 871–879, 2009.

[24] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes Ortiz, “Training computationally efficient Smartphone–Based human activity recognition models,” in *Artificial Neural Networks and Machine Learning – ICANN 2013*, Springer, 2013, pp. 426–433.

[25] Z. He and L. Jin, “Activity recognition from acceleration data based on discrete cosine transform and SVM,” in *IEEE International Conference on Systems, Man and Cybernetics*, 2009, 5041–5044.

[26] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. 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.

[27] D. Tao, L. Jin, Y. Yuan, and Y. Xue, “Ensemble manifold rank preserving for acceleration-based human activity recognition,” *IEEE transactions on neural networks and learning systems*, vol. 27, no. 6, 1392–1404, 2016.

[28] Ó. D. Lara, A. J. Pérez, M. A. Labrador, and J. D. Posada, “Centinela: A human activity recognition system based on acceleration and vital sign data,” *Pervasive and Mobile Computing*, vol. 8, no. 5, pp. 717–729, 2012.

[29] A. Oppenheim and R. Schafer, “From frequency to quefrency: A history of the cepstrum,” *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.

[30] R. San-Segundo, J. M. Montero, R. Barra-Chicote, F. Fernández, and J. M. Pardo, “Feature extraction from smartphone inertial signals for human activity segmentation,” *Signal Processing*, vol. 120, pp. 359–372, 2016.

[31] M. Li, V. Rozgić, G. Thatte, S. Lee, B. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan, “Multimodal physical activity recognition by fusing temporal and cepstral information,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, pp. 369–380, 2010.

[32] R. San-Segundo, R. Cordoba, J. Ferreiros, and L. F. D'Haro-Enríquez, "Frequency features and GMM-UBM approach for gait-based person identification using smartphone inertial signals," *Pattern Recognition Letters*, vol. 73, pp. 60–67, Apr. 2016.

[33] L. R. Rabiner and R. W. Schafer, *Theory and Applications of Digital Speech Processing*. Upper Saddle River: Prentice Hall, 2011.

[34] V. N. Vapnik, *The nature of statistical learning theory*. Springer, 2000.

[35] I. Steinwart and A. Christmann, *Support vector machines*. Springer, 2008.

[36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd. Springer, 2009.

[37] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 27:1–27:27, 2011.

[38] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, 10–18, 2009.

[39] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005.

[40] Y.-W. Chen and C.-J. Lin, "Combining SVMs with various feature selection strategies," in *Feature extraction*, Springer, 2006, 315–324.

[41] L. D. Vignolo, H. L. Rufiner, D. H. Milone, and J. C. Goddard, "Evolutionary cepstral coefficients," *Applied Soft Computing*, vol. 11, no. 4, pp. 3419–3428, 2011.

[42] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, 3rd. John Wiley & Sons, 2013.