

# A multivariate symbolic approach to activity recognition for wearable applications

M. Terzi \* A. Cenedese \*\* G.A. Susto \*\*

\* *Human Inspired Technology Center, University of Padova*

\*\* *Department of Information Engineering, University of Padova*

---

**Abstract:** With the aim of monitoring human activities (in critical tasks as well as in leisure and sport activities), wearable devices provide enhanced usability and seamless human experience with respect to other portable devices (e.g. smartphones). At the same time, though, wearable devices are more resource-constrained in terms of computational capability and memory, which calls for the design of algorithmic solutions that explicitly take into account these issues. In this paper, a symbolic approach for activity recognition with wearable devices is presented: the Symbolic Aggregate approximation technique is here extended to multi-dimensional time series, in order to capture the mutual information of different dimensions. Moreover, a novel approach to identify gestures within activities is here presented. The performance of the proposed methodology is tested on the two heterogeneous datasets related to cross-country skiing and daily activities.

*Keywords:* Activity Recognition, Machine Learning, Time Series Learning, Wearable Devices

---

## 1. INTRODUCTION

Activity Recognition (AR) is a prominent research area with applications to home automation (Belgioioso et al., 2014), gaming (Gowing et al., 2014), sport (Cenedese et al., 2016) and health care (Clifton et al., 2013). In particular, the rapid growth of IMUs (Inertial-Measurement Units) has allowed, in recent years, the development of compact sensor-equipped devices (e.g. smartwatches and smartphones), which lead efficient monitoring of human activities to be feasible and to have a strong impact on the quality of life (Clifton et al., 2013).

On the other hand, wearable devices present some limitations in terms of computational capability and memory, which force the algorithm design to be at the same time efficient and simple. Moreover, decision algorithms need to be portable, i.e., classification tasks have to be taken at a device-level (Cenedese et al., 2015).

It is important in this context to differentiate between *activities* and *gestures*. Gestures (also called in the following *atomic gestures*) are here considered as basis movements that compose an activity that is conversely completely characterized by one or more atomic gestures: for example, in swimming, a stroke (gesture) completely characterizes the style (activity).

Due to the vastness of application scenarios, it is helpful to categorize the AR problems into three main types:

- *continuous-repetitive* - activities that are continuous and composed by repeated gestures with a periodic behaviour within the same activity type;
- *continuous-spot* - continuous activities with non-repetitive gestures;
- *isolated* - activities composed by isolated gestures.

This work is focused on the *continuous-repetitive* type (Morris et al., 2014), that is typical of sports (e.g. rowing and swimming) and health monitoring applications.

AR problems are usually solved by means of Machine Learning (ML) approaches, however, the aforementioned restrictions on computation and memory capacity cause great limitations in choosing the ML algorithms to be employed. For instance, in (Cenedese et al., 2016), Relevance Vector Machines are chosen over popular Support Vector Machines in order to meet the parsimony constraint required by wearables in the ML algorithm complexity.

Within this context the above restrictions can otherwise be handled by adopting symbolic representation techniques (Rajagopalan and Ray, 2006) in the treatment of IMU-generated time series data; with symbolic approaches, time-series are mapped into *strings*, which implies dimensional and numerosity reduction. Moreover, symbolic representations allow avoiding one pre-processing phase, called *Feature Extraction*, which is common to AR solution and often critical in the selection of parameters to be retained in the models.

With these premises, the contributions of this paper are:

- (1) a symbolic approach based on Symbolic Aggregate ApproXimation (SAX) (Lin et al., 2003) for AR. SAX is a popular symbolic approach intended for univariate time-series: since in many AR problems multiple IMU-generated time-series are available, we extend here the approach to multi-dimensional time series in order to exploit mutual information from multiple axes; the results of our experiments confirm that this multi-dimensional extension leads to superior accuracy w.r.t. univariate approaches.
- (2) a procedure to extract atomic gestures and a classification model for Gesture Recognition (GR) is built directly on gestures; more interestingly, a model that is *invariant* to duration and amplitude warpings is depicted. Then, we perform AR starting from GR classification results, through a window-based approach;
- (3) an *Event Identification* (EI) procedure is designed in order to detect time windows where activities to be identified are not performed; these time windows are labelled as *all the other movements* (AOM).

The remainder of the paper is organized as follows: Section 2 is dedicated to discuss related works, while in Section 3 univariate symbolic classification problem is illustrated; Section 4 discusses the multivariate extension of the SAX approach while the gesture extraction phase is depicted in Section 5; in Section 6 the experimental results are shown, while Section 7 is dedicated to final remarks.

## 2. AR PROCEDURES

As stated above, AR/GR problems are usually tackled by means of ML approaches (Morris et al., 2014); more precisely AR/GR problems are generally *classification* ones: the activity or gesture in exam has to be associated with one of the a-priori defined  $K$  possible classes of activities/gestures  $\mathcal{C} = \{c_i\}_{i=1}^K$ .

The main challenge in applying ML algorithms in AR problems is to translate the informative content contained in the IMU-generated time series into a static format that can be handled by ML classifiers (Susto et al., 2016; Cenedese et al., 2016). Typically, this is achieved with a flow chart of operations as in the scheme depicted in Figure 1: the pipeline contains two blocks, window extraction and feature extraction, aiming to translate the informative content in the classical form  $X \in \mathbb{R}^{N \times p}$ , where  $N$  and  $p$  are, respectively, the number of observations (time windows in this context) and the number of features extracted from each window.

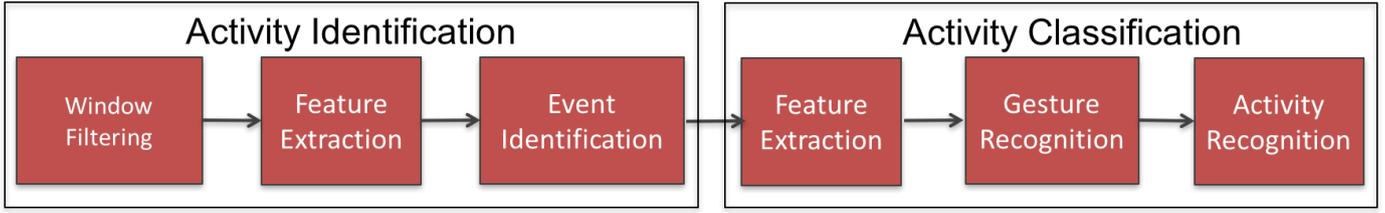


Fig. 1. Scheme of the classical ML approach to AR/GR problems.

The main drawback of using window-based methods is that there is no one-to-one correspondence between gestures and windows; in fact, gestures could have different shapes (at least locally) and warping in time and amplitude domains that dramatically change gestures duration and the window-based statistics (features)  $\phi \in \mathbb{R}^p$ . Moreover, the aforementioned approach has other two major issues: (i) single gestures are not isolated, which of course make the approach not feasible for GR problems; (ii) the feature extraction phase may be computationally expensive, causing this approach to be almost impracticable for wearable applications.

An alternative procedure suggests to directly compare raw signals with specific distance definitions and with the usage of a distance-based classifier (like Nearest Neighbor, NN (Susto et al., 2015)); classifying directly on the time-series allows to bypass the feature extraction phase, but the window extraction procedure is still required. In this sense, one of the most popular approaches for defining a distance between time series of different length is the Dynamic Time Warping (DTW) (Berndt and Clifford, 1994): unfortunately, DTW is a dynamic programming technique that hardly meets computational complexity and memory requirements of wearable devices.

In this paper, instead, we adopt a symbolic approach, which is graphically summarized in Figure 2. The proposed

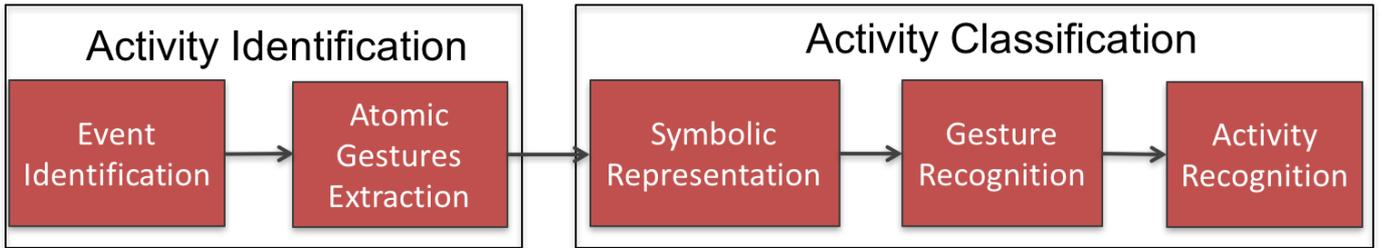


Fig. 2. Scheme of the symbolic approach to AR/GR problems adopted in this work.

method is not relying on window-based extraction procedures, but exploits a preliminary *atomic gesture* extraction phase. After the extraction phase, each gesture is then symbolized through the SAX technique (briefly described in the following); afterwards, a classification model is built over the symbolic representation of a input collection  $\mathcal{X} = \{X_i \in \mathbb{R}^{m \times p}\}_{i=1}^N$ , where  $m$  is the dimensionality,  $p$  the feature cardinality and  $N$  represents the number of gestures. The *Activity Classification* phase will be detailed in Sections 3, 4 then in Section 5 the *Activity Identification* phase is described.

We here recall the SAX technique<sup>1</sup>, which mainly consists on 3 phases:

- signals standardization in order to obtain a zero mean and unit variance signal;
- Piecewise Aggregate Approximation (PAA), described afterwards;
- Symbolic mapping through discretization on amplitude domain.

After normalization, in the PAA phase, a signal  $T = t_1, \dots, t_q$  is discretized on time in  $w$  frames in order to obtain a vector  $\bar{T} = \bar{t}_1, \dots, \bar{t}_w \in \mathbb{R}^w$ . Formally, the resulting  $i$ -th element  $\bar{t}_i$  is defined by:

$$\bar{t}_i = \frac{w}{q} \sum_{j=\frac{q}{w}(i-1)+1}^{\frac{q}{w}i} t_j \quad (1)$$

Then, the SAX representation procedure (i.e. the discretization on amplitude domain) can be summarized as follows. Let  $a_i$  denote the  $i$ -th element of the alphabet  $\mathcal{A}$ , with  $|\mathcal{A}| = \alpha$ . The mapping from the PAA approximation to the correspondent word  $\tilde{T} = \tilde{t}_1, \dots, \tilde{t}_w$  of length  $w$  is obtained as follow:

$$\tilde{t}_i = a_j \quad \text{iff} \quad \beta_{j-1} \leq \bar{t}_i < \beta_j, \quad (2)$$

where  $\{\beta_j\}_{j=1}^{\alpha-1}$  are break-points tuned to have symbols with equiprobable occurrence. One of the advantages of introducing the *SAX representation*, is that a new distance measure - which is a lower bound of euclidean distance -

<sup>1</sup> We refer the interested reader to (Lin et al., 2003) for a more detailed treatment of the SAX technique.

can be immediately defined. Let  $T$  and  $U$  be two time-series of the same length  $q$  and  $\tilde{T} = \tilde{t}_1, \dots, \tilde{t}_q$  and  $\tilde{U} = \tilde{u}_1, \dots, \tilde{u}_q$  be their *SAX symbolic representation*; the SAX distance is defined as:

$$D_{SAX}(\tilde{T}, \tilde{U}) = \sqrt{\frac{q}{w} \sum_{i=1}^w \text{dist}(\tilde{t}_i, \tilde{u}_i)^2}. \quad (3)$$

### 3. ACTIVITY CLASSIFICATION FOR 1-D SIGNALS

To tackle the AR problem, it is convenient to have gestures that are normalized in duration; for this reason we assume in the following that fixed length gestures are available, without loss of generalization since resampling procedures can be in place before the Activity Classification phase.

Let assume of having  $N$  1-dimensional gestures (or more generally signals)  $g_i \in \mathbb{R}^{1 \times q}$  of *fixed* length  $q$ , collected in a row-wise matrix  $X = [g_1; \dots; g_N] \in \mathbb{R}^{N \times q}$ ; a classification problem with  $K$  classes of gestures/activities  $\mathcal{C} = \{c_i\}_{i=1}^K$  is considered. By applying the SAX procedure to the elements of  $X$  we obtain a collection of strings  $\tilde{\mathcal{G}} = \{\tilde{g}_i\}_{i=1}^N$  of length  $\frac{q}{w}$ , with characters belonging to  $\mathcal{A}$ . After the symbolization procedure, following the scheme depicted in Section 2, we employ a ML classifier that exploits the distance defined by SAX. It has been shown (Ding et al., 2008) that a simple 1-NN classification guarantees state-of-the-art performance in terms of classification accuracy with time-series data; hence in this work, 1-NN is adopted in combination with SAX.

For any new observation to be classified, 1-NN (like all  $k$ -NN classifiers) needs to evaluate distances with all the available training data, requiring large space to store the entire training dataset; to make the approach feasible for the problem at hand, we use *templates*, i.e.  $s \in \mathbb{N}^+$  observations for each class that are chosen as representative of all the observations available in the training dataset; the motivation of having more than one template per class, is that, input signals may drastically change with different scenarios (like wearable devices located in different body positions or activities performed in different environmental conditions) which could not be captured without using multiple templates.

The classification model is based on a set of  $\mathcal{M} = \{m_{ij} \in \mathcal{A}^{1 \times w}, i = 1, \dots, K, j = 1, \dots, s\}$ . For each class, the  $s$  most representative templates can be found through a clustering technique based on SAX-distance, or, in a simpler fashion by selecting the most frequent  $s$  observations; in this work we adopt this last criterion.

In the case  $s = 1$ , the prediction task for a new observation  $x$  is easily performed by:

$$\hat{c}_i = \arg \min_i d(x, m_{i1}), \quad i = 1, \dots, K, \quad (4)$$

where  $d := d_{SAX}$ . When  $s > 1$ , given the premises on the choice for  $s > 1$ , a natural extension is given by:

$$\hat{c}_i = \arg \min_i d(x, m_{ij^*}), \quad i = 1, \dots, K, \quad (5)$$

where, for the  $\bar{i}$  class the optimal template is selected

$$j^* = \arg \min_j d(x, m_{\bar{i}j}), \quad j = 1, \dots, s. \quad (6)$$

Once gestures are classified, the focus must be placed on the AR (see Figure 2). Activities are compositions of gestures: in order to move from gestures to activities, a sliding window approach is here employed. A fixed length window of width  $l_w$  is considered: in each window the predicted activity is chosen as the mode of the classified gestures; the window is then shifted forward of one gesture.

### 4. SAX MULTIVARIATE EXTENSION

The framework proposed in the previous Section is limited by the fact that SAX technique only operates with 1-d time-series, which can lead to poor performance in AR problems. In this Section, we discuss extensions of the SAX approach to multidimensional cases.

Let assume of having  $N$   $p$ -dimensional gestures  $\mathbf{g}_i = [g_i^{(1)}; \dots; g_i^{(p)}] \in \mathbb{R}^{p \times q}$  where  $g_i^{(\cdot)} \in \mathbb{R}^{1 \times q}$ . We define  $d^{(i)}$  as the 1-dimensional SAX distance independently computed for the  $i$ -th dimension; we also define the multivariate distance  $\mathbf{d} = \Psi(d^{(1)}, \dots, d^{(p)})$ . In the following, three different versions of  $\mathbf{d}$  are presented:

$$\mathbf{d}_{\min} = \min\{d^{(1)}, \dots, d^{(p)}\} \quad (7)$$

$$\mathbf{d}_{\text{mean}} = \frac{1}{p} \sum_i^p d^{(i)} \quad (8)$$

$$\mathbf{d}_{\text{geom}} = \sqrt[p]{\prod_{i=1}^p d^{(i)}}. \quad (9)$$

With a trivial extension, we employ the just defined multivariate distances to classify accordingly to Figure 2<sup>2</sup>.

The motivation behind distances (7) and (9) is that we want to discard, as much as possible, disturbances derived from changes in orientation; they act similarly with the difference that  $d_{\text{geom}}$  is less biased towards the minimum. These two measures could be unstable, especially where the intra-class and inter-class variance of distance are similar (i.e. when having low discrimination power). Distance (8) avoids this issue but could cause a convergence "phenomenon" where multivariate gestures tend to be equally distant.

## 5. ACTIVITY IDENTIFICATION AND GESTURE EXTRACTION

In this Section we illustrate a procedure to extract gestures from streams of data and to automatically identify non-activities (AOM) regions.

The adopted approach is based on the following assumption: if a non-stationary input signal exhibits periodical (or quasi-periodical) behavior, it can be decomposed into intrinsic mode functions that represent the signal at different time scales. Thus, the *fundamental* mode functions can be extracted in order to capture the activity periodicity. The previous task can be accomplished by the well-known Empirical Mode Decomposition (EMD) (Huang et al., 1998) technique, however, for computational limits, this technique is not suited for the application at hand.

In this work, we employ a simpler filtering method that follows the aforementioned philosophy, a Gaussian filter with kernel  $G_\sigma(t) = \exp[-t^2/(2\sigma^2)]/(\sqrt{2\pi}\sigma)$ . Let  $x(t)$  be the original and the  $y(t)$  its filtered version, defined as

$$y(t) = \int_{-\infty}^{\infty} x(t-\tau)G_\sigma(\tau)d\tau \approx \int_{t-3\sigma}^{t+3\sigma} x(t-\tau)G_\sigma(\tau)d\tau, \quad (10)$$

where  $\sigma$  should be opportunely chosen with *a-priori* knowledge of the minimum fundamental frequency of  $x(t)$ . In fact, the relative cut-off frequency is  $f_c = 1/2\pi\sigma$ , while in the discrete domain (now  $\sigma$  is measured in samples), the cut-off frequency (in physical units) can be calculated from  $f_c = F_s/(2\pi\sigma)$ , where  $F_s$  is the sampling frequency.

---

### Algorithm 1 Gesture extraction algorithm

---

```

1: procedure GESTUREEXTRACTION( $x(t)$ ,  $\sigma$ ,  $q$ ):
2:   Apply Gaussian filter  $y(t) = x(t) * G_\sigma(t)$ 
3:   Event Identification phase. Return  $y'(t)$ , defined as  $y(t)$  without AOM regions.
4:   Find local maxima points  $\{P_i\}_{i=1}^N$  from  $y'(t)$ 
5:   for all  $P_i$  do:
6:     Find borders  $(b_{1i}, b_{2i}), b_{2i} > P_i > b_{1i}$ 
7:     Extract gesture  $g_i$  delimited by borders  $b_{1i}$  and  $b_{2i}$ 
8:     Resample to  $q$  samples
9:   end for
10:  return  $\mathcal{G} = \{g_i \in \mathbb{R}^{p \times q}\}_{i=1}^N$ 
11: end procedure

```

---

### Algorithm 2 Activity recognition algorithm

---

```

1: procedure TRAINING( $\mathcal{G}_{\text{train}}, \theta$ ):
2:   Symbolize gestures in  $\mathcal{G}$  and return  $\tilde{\mathcal{G}}$ 
3:   Compute templates:

$$\mathcal{M} = \{\mathbf{m}_{ij}, i = 1, \dots, K, j = 1, \dots, s, \mathbf{m}_{ij} \in \mathcal{A}^{p \times w}\},$$

   where  $\mathbf{m}_{ij}$  is the  $j$ -th most frequent gesture for class  $i$ 
4:   return  $\mathcal{M}$ 
5: end procedure
6: procedure CLASSIFICATION( $\mathcal{G}_{\text{test}}, \mathcal{M}, l_w$ ):
7:   for all  $g_i \in \mathcal{G}_{\text{test}}$  do
8:     Symbolize  $g_i$  into  $\tilde{g}_i$ 
9:     Predict gesture through 1-NN classifier:

$$\hat{c} = \arg \min_{i, \mathbf{m}_{ij} \in \mathcal{M}} d_{SAX}(\tilde{g}_i, \mathcal{M})$$

10:  end for
11:  Predict activities using a sliding window  $l_w$ 
12: end procedure

```

---

After filtering, gestures are extracted as follows:

- first run a EI procedure in which non-AOM regions are detected; we accomplish this tasks by analyzing the auto-correlation of  $y(t)$  on a sliding window over a reference univariate signal and by monitoring the overcoming of an

<sup>2</sup> Rigorously speaking, one should select multivariate templates by choosing the  $s$  most frequent vectors  $\{(m_{1j}, \dots, m_{pj})\}_{j=1}^s$ . However, this multivariate space is heavily sparse and this choice is affected by the typical "curse of dimensionality" issue Hastie et al. (2009) that lead, in our simulations, to the choice of poor representatives for the classes. To overcome this issue, in this work we select templates independently for each dimension.

opportune threshold  $\eta_{th}$  on the *fundamental* correlation peak. A dedicated set of experiments have been performed to choose a optimal value for  $\eta_{th}$ : as a result, in this work we adopt  $\eta_{th} = 0.5$ .

- a peak detection algorithm discovers maxima, which represent a one-to-one relation with each gesture;
- then, starting from each point of maximum, progress backwards and forward until two points of local minimum are found. These points are the borders of the examined gesture.

This procedure returns a collection of gestures of variable lengths. In order to make gestures comparable, i.e. invariant to *time warping*, we resample to a fixed length  $q$ : in this paper we use the Akima interpolation (Akima, 1970). For the sake of clarity, Algorithm 1 and 2 summarize the fundamental AR steps.

## 6. EXPERIMENTS

The proposed methodology has been tested on two different datasets:

- *HAR dataset* - a reduced version of the public UCI Human Activity Recognition (HAR) Using smart-phones Dataset (Anguita et al. (2013)) where 3 continuous-repetitive normal day activities are considered: walking (WLK), walking upstairs (WUS) and walking downstairs (WDS). We are therefore examining a 3 classes AR problem. The dataset includes experiments that were carried out by 30 people where all the participants were wearing a smartphone (a Samsung Galaxy S II) on the waist during the experiment; 3-axial linear acceleration and 3-axial angular velocity have been captured at a constant rate of 50Hz using the embedded accelerometer and gyroscope of the device.
- Cross-Country Skiing dataset (*XC dataset*)- a private dataset of Cross-Country Skiing where 3 different styles were performed by 8 skiers; the three styles are
  - (1) double poling (DP), where both poles are used in parallel by the skier;
  - (2) diagonal stride (DS), where the poles are used in succession;
  - (3) kick-double-pole (KDP), a variant of DP, where an asymmetrical kick is performed by the skier.
 Athletes were wearing a smart-watch placed on the wrist; 3-axial linear acceleration, angular velocity and magnetic field, have been captured at a constant rate of 100Hz using the embedded accelerometer, gyroscope and magnetometer of the device.

Here we focus on 3D-acceleration signals ( $p = 3$ ), which best capture signal variability. Furthermore, we set the fixed length of each gesture to  $q = 150$  samples for both the datasets. For tuning purposes and in order to assess the algorithm performance, a cross-validation procedure on parameters  $\theta = (w, \alpha, s)$  is run. Moreover, the univariate classification (D1) of single dimension and the multivariate classification with the defined distances are compared: for the sake of simplicity, we will present only the results deriving from the dimension that guarantees the best accuracy (i.e. z-axis).

In the experiments, a L1-distance is also employed for accuracy comparison with the SAX-based classification: in this type of experiments, the classification is done directly on gestures, bypassing the symbolic representation based on a L1-distance. Templates, in this case, are chosen after a  $k$ -means clustering procedure Hastie et al. (2009). Finally, for reasons of space, we shall report only the final AR results, which are our main concern.

*Remark 1.* Reported results was obtained following cross-validation Hastie et al. (2009) paradigms in order to guarantee a fair evaluation of the proposed approaches.

In the following, for each dataset, results are reported following an accuracy improvement order.

### *Results of HAR dataset*

In Figures 3 we report the results for the univariate and multivariate AR problems for the SAX and L1 models, respectively. In particular, Figures 3(a), 3(b), 3(d) and 3(e) refer to the HAR dataset while Figures 3(c) and 3(f) refer to the XC dataset. The results for the 3-classes problem are unsatisfying and this is caused by the fact that WDS and WUS are hardly distinguishable analyzing only z-axis, which is the axis towards the motion direction. Thus we consider a new binary problem with classes WLK vs WS (i.e. WDS and WUS are treated as a single class, WS).

As demonstrated, limiting the analysis to *one* axis could be too restrictive; in the following, it will be shown how the performance can be improved by employing the multi-variate extension presented in Section 4. All the distances (7-9) were tested and it was verified that the mean distance  $d_{mean}$  (8) performs systematically better then the others either in terms of robustness and of accuracy. Therefore, only the results for such distance are presented in the following.

Finally, in Table 1 we report the confusion matrices for SAX and L1 models at the optimal value of the tuning parameters  $\theta$ , which represent the best cross-validation accuracy results for the univariate and multivariate problems, respectively.

### *Cross-Country Skiing dataset*

As already stated, the original data at hand potentially allow the study of 3-classes recognition problem (DP, KDP, DS). However, classes DP and KDP can be considered indistinguishable given the data available: the gestures only differ by a kick that seems to be not 'observable' from the wrist, the location of the wearable collecting the data. Hence, we

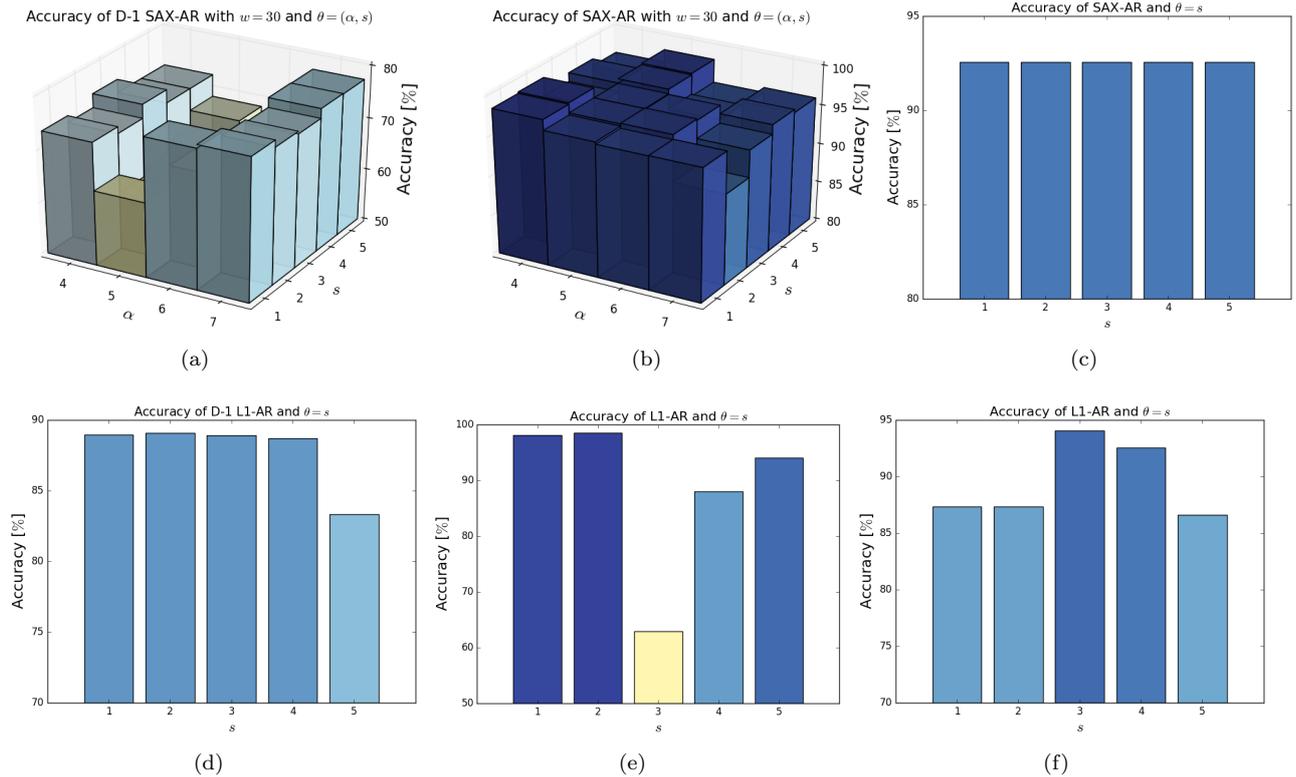


Fig. 3. Results of AR. Comparison between univariate and multivariate approaches. Panels 3(a), 3(b), 3(e), 3(d) refer to the HAR dataset while Panels 3(c), 3(f) refer to the XC dataset.

Table 1. HAR dataset. Confusion matrices (cross-validated results). (a): univariate problem with  $\theta_{SAX} = (\alpha, w, s) = (5, 15, 3)$  and  $\theta_{L1} = s = 3$ . (b): multivariate with  $\theta_{SAX} = (\alpha, w, s) = (6, 30, 2)$  and  $\theta_{L1} = s = 2$ .

Tab1a		Predicted (SAX)		Predicted (L1)	
		WLK	WS	WLK	WS
True	WLK	1293 (85.5%)	218 (14.43%)	1489 (98.54%)	22 (1.46%)
	WS	94 (12.53%)	656 (87.47%)	151 (20.13%)	599 (79.87%)
Tab1b		Predicted (SAX)		Predicted (L1)	
		WLK	WS	WLK	WS
True	WLK	1504 (99.54%)	7 (0.46%)	1497 (99.07%)	14 (0.93%)
	WS	22 (2.93%)	728 (97.07%)	19 (2.53%)	731 (97.47%)

simplify the original classification problem in a 2-classes problem where DP and KDP are considered as the same class (K/DP). For the sake of conciseness, only the multivariate experiments are here reported (Figures 3(c) and 3(f)), while cross-validated results are reported in Table 2.

### Discussions

From the experiments reported in this work, it can be concluded that the proposed multivariate extension of SAX allows good classification accuracy for AR problems; this is achieved at a feasible cost for wearable devices in terms of complexity and memory. While the previous outcome is clearly prove in the case of the HAR dataset, for the XC dataset, results of multivariate and univariate problems are almost identical: in fact, all the informative content is contained in the  $x$ -axis, i.e. the axis which points to the direction of wrist motion (this was verified through a PCA procedures, indicating that the first Principal Component is represented by the  $x$ -axis). On the other hand, we experimentally verified that by increasing the number of templates  $s$  and, therefore, the complexity of the solution, accuracy does not systematically improves. We suppose that one of the reasons of this results is  $s$  should be set differently for each

Table 2. *XC dataset*. Confusion matrices (cross-validated results). (a): univariate problem with  $\theta_{SAX} = (\alpha, w, s) = (7, 30, 1)$  and  $\theta_{L1} = s = 3$ . (b): multivariate problem with  $\theta_{SAX} = (\alpha, w, s) = (7, 30, 2)$  and  $\theta_{L1} = s = 3$ .

Tab2a		Predicted (SAX)		Predicted (L1)	
		K/DP	DS	K/DP	DS
True	K/DP	81 (98.78%)	1 (1.22%)	81 (98.78%)	1 (1.22%)
	DS	9 (17.31%)	43 (82.69%)	7 (13.46%)	45 (86.54%)
Tab2b		Predicted (SAX)		Predicted (L1)	
		K/DP	DS	K/DP	DS
True	K/DP	82 (100.0%)	0 (0.0%)	81 (98.78%)	1 (1.22%)
	DS	10 (19.23%)	42 (80.77%)	7 (13.46%)	45 (86.54%)

element of the set  $R \times K$ , where  $D$  is the dimension set and  $K$  is the templates set. For example, the optimal value  $s = s_0$  for class  $k$  and dimension  $r$ , could be too low for certain classes/dimensions (which causes an under-estimation of complexity) or too high (which could cause the injection of noise).

## 7. CONCLUSIONS

In this paper, an effective warping-invariant symbolic approach to the AR problem, whose parsimony is suitable for wearable devices, is presented. The procedure allows, for repetitive activities, to jointly solve the GR and the AR problem by adopting a Gaussian filtering-based approach to gesture extraction. Furthermore, an extension to symbolic approximation procedure SAX is here proposed: this is achieved with the employment of ad-hoc defined multi-variate distances within the SAX-based classification, allowing a remarkable improvement of the performance. Future studies will be dedicated to investigate the role of the number of templates  $s$ , generalizing to procedure to adopted different  $s_{ij}$  for each class  $i$  and dimension  $j$ .

## REFERENCES

- Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM*, 17(4), 589–602.
- Anguita, D., Ghio, A., Oneto, L., Parra Perez, X., and Reyes Ortiz, J.L. (2013). A public domain dataset for human activity recognition using smartphones. In *21th Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 437–442.
- Belgioioso, G., Cenedese, A., Cirillo, G.I., Fraccaroli, F., and Susto, G.A. (2014). A machine learning based approach for gesture recognition from inertial measurements. In *IEEE Conf. on Decision and Control*, 4899–4904.
- Berndt, D.J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, 359–370. Seattle, WA.
- Cenedese, A., Susto, G.A., Belgioioso, G., Cirillo, G.I., and Fraccaroli, F. (2015). Home automation oriented gesture classification from inertial measurements. *IEEE Trans. on Automation Science and Engineering*, 12(4), 1200–1210.
- Cenedese, A., Susto, G.A., and Terzi, M. (2016). A parsimonious approach for activity recognition with wearable devices: an application to cross-country skiing. In *European Control Conference*, 1243–1252.
- Clifton, L., Clifton, D.A., Pimentel, M.A., Watkinson, P.J., and Tarassenko, L. (2013). Gaussian processes for personalized e-health monitoring with wearable sensors. *IEEE Trans. on Biomedical Eng.*, 60(1), 193–197.
- Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., and Keogh, E. (2008). Querying and mining of time series data: experimental comparison of representations and distance measures. *VLDB Endowment*, 1(2), 1542–1552.
- Gowing, M., Ahmadi, A., Destelle, F., Monaghan, D., O’Connor, N., and Moran, K. (2014). Kinect vs. low-cost inertial sensing for gesture recognition. *Lecture Notes in Computer Science*, 8325, 484–495.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*.
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., and Liu, H.H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 1971, 903–995.
- Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In *ACM SIGMOD Research Issues in Data Mining and Knowledge Discovery*, 2–11.
- Morris, D., Saponas, T.S., Guillory, A., and Kelner, I. (2014). Recofit: Using a wearable sensor to find, recognize, and count repetitive exercises. In *SIGCHI Human Factors in Computing Systems*, 3225–3234.
- Rajagopalan, V. and Ray, A. (2006). Symbolic time series analysis via wavelet-based partitioning. *Signal Processing*, 86(11), 3309–3320.

- Susto, G.A., Schirru, A., Pampuri, S., and McLoone, S. (2016). Supervised aggregative feature extraction for big data time series regression. *IEEE Transactions on Industrial Informatics*, 12, 1243 – 1252.
- Susto, G.A., Schirru, A., Pampuri, S., McLoone, S., and Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 11(3), 812–820.