User-Independent Recognition of Sports Activities From a Single Wrist-Worn Accelerometer: A Template-Matching-Based Approach

Jenny Margarito*, Rim Helaoui, Anna M. Bianchi, *Member, IEEE*, Francesco Sartor, and Alberto G. Bonomi

***Abstract*—*Goal*: To investigate the accuracy of template matching for classifying sports activities using the acceleration signal recorded with a wearable sensor. *Methods*: A population of 29 normal weight and 19 overweight subjects was recruited to perform eight common sports activities, while body movement was measured using a triaxial accelerometer placed at the wrist. User- and axis-independent acceleration signal templates were automatically extracted to represent each activity category and recognize activity types. Five different similarity measures between example signals and templates were compared: Euclidean distance, dynamic time warping (DTW), derivative DTW, correlation and an innovative index, and combining distance and correlation metrics (*Rce*). Template-based activity recognition was compared to statistical-learning classifiers, such as Naïve Bayes, decision tree, logistic regression (LR), and artificial neural network (ANN) trained using time- and frequency-domain signal features. Each algorithm was tested on data from a holdout group of 15 normal weight and 19 overweight subjects. *Results*: The *Rce* index outperformed other template-matching metrics by achieving recognition rate above 80% for the majority of the activities. Template matching showed robust classification accuracy when tested on unseen data and in case of limited training examples. LR and ANN achieved the highest overall recognition accuracy ∼85% but showed to be more vulnerable to misclassification error than template matching on overweight subjects' data. *Conclusion*: Template matching can be used to classify sports activities using the wrist acceleration signal. *Significance*: Automatically extracted template prototypes from the acceleration signal may be used to enhance accuracy and generalization properties of statistical-learning classifiers.**

***Index Terms*—Activity classification, dynamic time warping, template prototypes, overweight subjects.**

I. Introduction

RECENT and rapid advances in wearable sensing and computing technologies have opened new avenues to a wide range of applications in the manufacturing, entertainment, and healthcare domain [1]. Special attention has been dedicated to the automatic recognition of users physical activity with the intention to foster context-aware applications for health and lifestyle monitoring [2], [3].

Manuscript received January 6, 2015; revised June 24, 2015; accepted August 10, 2015. Date of publication August 21, 2015; date of current version March 17, 2016. This work was supported by COMMIT-NL and carried out in the context of the SWELL project. *Asterisk indicates corresponding author.*

*J. Margarito is with Philips Research, Eindhoven 5654 EC, The Netherlands (e-mail: jenny.margarito@philips.com).

R. Helaoui, F. Sartor, and A. G. Bonomi are with Philips Research.

A. M. Bianchi is with Politecnico di Milano.

Several researchers have approached the challenge of recognizing human activities such as walking, running, and cycling from wearable sensors [3], [4]. A tradeoff is often observed between complexity of the classification problem and user-friendliness of the measurement system. For example, to successfully achieve activity classification using a single wearable accelerometer, different activity types are often grouped in a few large clusters which are then targeted by the classification algorithm [5]–[7]. On the other hand, identification of several activity types often requires recording discriminative features from multiple accelerometer units. The seminal work of Bao and Intille [8] employed a large set of accelerometers to discriminate between 20 activities. The authors also tried to keep an acceptable accuracy level while reducing the number of sensors used. However, the simplest system still required data from two accelerometers to preserve classification accuracy. Studies focusing on measuring physical activity with a single wearable sensor [9], [10] showed that the acceleration signal recorded at the trunk and lower limbs offered better discriminatory features for classifying types of movement of the body center of mass like walking [10]. Yet, sports and activities that involve movement of upper limbs are difficult to identify by monitoring the acceleration of the trunk and legs only [8], [11].

Statistical-learning methods have been extensively applied to solve activity recognition problems [12]. Gupta and Dallas [13] developed a Naïve Bayes (NB) and a *k*-nearest neighbor classifier to recognize seven activities, such as walking, jumping, and running. Other popular algorithms are: Artificial neural network (ANN), decision tree (DT), support vector machine (SVM), random forest, etc., [14]–[16]. Despite the successful application of these algorithms for activity recognition, their evaluation is often restricted to a limited set of data obtained in supervised and structured settings. Testing robustness of the classification accuracy on data collected from different population and measurement settings is often neglected. Several studies focused on analyzing data collected from a small and homogeneous group of individuals with minimally varying anthropometric characteristics and motor skills. In addition, there is evidence that laboratory-trained algorithms such as ANN and SVM have poor reproducibility in free living [16], suggesting a significant decrease in classification accuracy when applied on data from previously unseen conditions.

In an attempt to bridge the gap between laboratory experiments and real-life applications, this paper proposes a template-matching-based framework to recognize a set of eight sports activities from a single accelerometer worn at the wrist. We

carefully selected the wrist placement of the sensor for its unobtrusiveness. The focus on recognizing sports activities is derived from their high relevance for healthcare and lifestyle applications. Distinguishing between different types of sports allows precise assessment of energy expenditure and cardiofitness level [17]. Moreover, automatically detecting different types of exercise are relevant for several rehabilitation and lifestyle intervention programs providing coaching tools to guide patients and users toward a better health and fitness.

Template matching has been seldom applied for physical activity. Muscillo *et al.* [18] proposed user-dependent templates to target recognition of arm-specific tasks. Likewise, Chen and Shen [19] focused on recognizing activities performed with the right upper limb using a classification framework based on template matching. Stifefmeier *et al.* [20] proposed an innovative approach consisting of encoding motion data into sequence of finite symbols and performing activity recognition by using string-matching algorithms.

Finally, Muscillo *et al.* [21] applied template matching to the shin acceleration signal to distinguish between walking, climbing, and descending stairs but still based on subject-, sensor-, and even measurement-axis-specific templates which cannot generalize to data from unknown subjects.

To summarize, many existing methods to automatically recognize several types of sports and human physical activity often rely on a set of a few accelerometers. Exceptions that reduce this set to one single sensor are either target very broad activity classes or focus on limb-specific movements, for example arm-related activities. Additionally, most studies fail to test the robustness of the method and ability to generalize on different groups of users. The purpose of this paper was to investigate the use of template matching for recognizing sports activities using one single accelerometer worn at the wrist. In particular, we wanted to evaluate the ability of a template-matching classification algorithm to generalize on a population of overweight subjects and compare it to popular statistical-learning-based classifiers.

II. Background

Template matching has been successfully applied in several domains, such as computer vision [22], speech recognition [23], and gait analysis [24]. Its main idea consists of two major steps: 1) generating templates for each target class using entities with particular patterns, 2) comparing each new entity to the set of generated templates in order to find the best-fitting one. Thus, the unknown entities can be classified to the target class represented by the selected template [25]. In this section, we describe these two steps and provide an overview of their principal techniques.

A. Generating Templates from Time Series Signal

A *template* is defined by Brunelli [22] as: *1) Anything fashioned, shaped, or designed to serve as a model from which something is to be made: a model, design, plan, outline. 2) Something formed after a model or prototype, a copy; a likeness, a similitude. 3) An example, an instance; esp. a typical model or a representative instance.* Thus, a useful template has to be *general* enough to match the corresponding patterns despite their potential distortion, but also *specific* enough so to substantially differ from all other types of patterns. Considering time series, a reference template consists of a specific *periodic waveform* which is repeated in entities of the same class. The template formation can be carried out in two different ways.

The first method consists of extracting repetitive waveforms from few signal instances and selecting those that ensure the highest recognition rate. Often, in order to increase the classification performance, multiple templates are generated for each class. However, this comes at higher computational costs. A compromise to reduce the number of references per class and to improve the template robustness would be to generate a template by *averaging* waveforms extracted from different examples.

The second approach consists of extracting and interpolating fiducial points, which synthesize the specific shape, by applying spline function [26]. Such an approach can be seen as a smoothing procedure that may affect the waveform specificity and lead to a high misclassification incidence when classes present some similarities.

B. Matching Time Series Signal with Generated Templates

Brunelli [22] defines *matching* as *comparing with respect to similarity; to examine the likeness or difference of*. Hence, once a template is generated for each target class, unseen entities are compared to available templates and are associated to the class with the highest matching level. The matching level is usually evaluated based on distance or correlation measures. The smaller the distance between two time series, or the higher their similarity, correlation and, hence, likelihood that they belong to the same class. Some of the most popular similarity measures that have been applied in the context of template matching are described below.

1) Euclidean Distance: It is the simplest similarity index employed to quantify the correspondence between two sequences of points [27]. Let X and Y be two time series as follows:

$$X = x_1, x_2, \ldots, x_i, \ldots, x_n, \qquad Y = y_1, y_2, \ldots, y_i, \ldots, y_m.$$

The distance $d(i)$ with $i = 0, \ldots, m-n-1$ and $n < m$, is calculated by comparing X with Y divided into overlapped segments composed by n samples. For each ith sample, the norm of the difference (distance) between X and n samples of the signal Y is computed. The Euclidean metric, $d(i)$ is computed as:

$$d(i) = \sqrt[2]{\sum_{k=1}^{n} (Y(i+k) - X(k))^2}. \tag{1}$$

If the compared entities have the same overall shape but different time scale, this metric would fail to assess the real similarity between them. This is the case of acceleration signals detected from activities characterized by different speed of execution. To overcome this limitation, nonlinear algorithms such as dynamic time warping (DTW) and derivative dynamic time warping (DDTW) [28] are often employed.

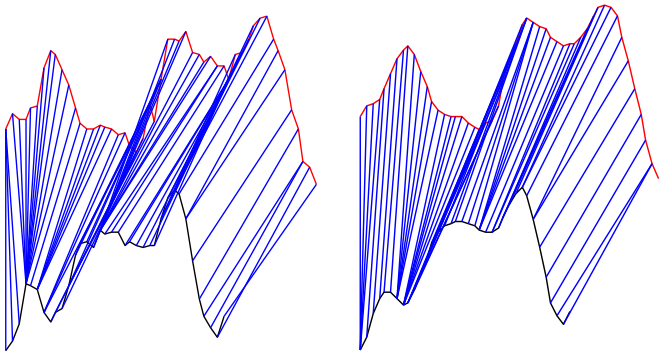

Fig. 1. Example of alignment between two time series, using DTW (to the left) and DDTW (to the right). Both methods present *singularities*, although DDTW shows to provide a more accurate alignment.

2) DTW: It searches for the best alignment of two series by considering all the compressions or stretching of the *x*-axis (time axis for time series), in order to find the mapping which provides the lowest cumulative distance. Thus, it addresses the challenge of similarity evaluation between sequences of points having same shape yet, unaligned on the *x*-axis.

Basically, given the two series X and Y introduced above, the algorithm consists of

1) Building a $m \times n$ matrix, where elements represent the distance $d(y_i, x_j)$ between each couple of points y_i, x_j.
2) Defining all the sequences W of matrix elements, called warping paths, with $w_k = (i, j)_k$, where $w_1 = (1, 1)$ and $w_k = (m, n)$ and constrains of continuity and monotonicity between adjacent w_k elements are respected.
3) Finding the optimal warping path W^* by minimizing the cost function $\text{Dist}_W(Y, X)$, which represents the normalized cumulative distance associated to each W

$$W^* = \arg\min_{W \in \mathbf{W}} \text{Dist}_W(Y, X). \tag{2}$$

Thus, the distance between X and Y is given by adding up the norm difference between those pairs of points belonging to the optimal warping path. The constraints regulating the construction of this warping, however, might lead to nonoptimal alignments. Indeed, as illustrated in Fig. 1, one single point of one sequence may be mapped to multiple points of the other sequence producing the so-called $singularities$. Also, differences in the amplitude scales of the two signals might result in aligning the peak of one sequence to the valley of the second, which in turn, affects the similarity evaluation.

3) DDTW: It is a modified version of DTW which allows us to prevent misalignment resulting from offset translation or different amplitude range. This is achieved by comparing the derivative of the considered series. The derivative of a sequence X is calculated as

$$D_{x[q]} = \frac{(q_i - q_{i-1}) + ((q_{i+1} - q_{i-1})/2)}{2}. \tag{3}$$

Interestingly, DDTW outperforms DTW [21], [28] even though both methods have the same complexity $O(mn)$. Nevertheless, DDTW is not exempt from the presence of $singularities$. Indeed, calculating the derivative of noisy signals might enhance noise and result in unreliable comparisons. Thus, a prefiltering step is usually required, despite the increased computational cost.

4) Cross Correlation: It offers an alternative matching approach to the distance measures presented above. It is based on a cross-correlation index [29] which allows the quantification of the relationship between waveforms both in terms of shape and mutual time delay. Given two time series X of length n points and Y of length m points, the cross correlation $C_{YX}(\tau)$, with $\tau = 0, \ldots, m-1$, is defined as

$$C_{YX}(\tau) = \frac{1}{n-1} \sum_{i=0}^{n-1} [Y(i+\tau)][X(i)]. \tag{4}$$

Often the cross-correlation index is normalized with respect to the standard deviation of the two signals and is called *correlation coefficient* $\gamma_{YX}(\tau)$. It is defined as

$$\gamma_{YX}(\tau) = \frac{C_{YX}(\tau)}{\sigma_{YY}\sigma_{XX}} \tag{5}$$

where σ_{YY} and σ_{XX} refer to the standard deviations of Y and X, respectively. Its value ranges between -1 and $+1$. The $\gamma_{YX}(\tau)$ equals to -1 indicates same shape but opposite phase, 0 indicates the absence of similarity, and 1 refers to total similarity.

When a signal is compared with itself, the cross correlation is called autocorrelation. It is defined as follows:

$$\hat{R}_{YY}(\tau) = \frac{1}{m-1} \sum_{i=0}^{m-\tau-1} [Y(i+\tau)][Y(i)]. \tag{6}$$

Such function is usually exploited to distinguish periodic signals from white noise and for identification of repetitive patterns and signal periodicity [30].

5) Rce: It is an innovative index that we proposed in this paper as additional measure to find shape similarity between signal segments. This was calculated as the ratio of *correlation coefficient* to *Euclidean distance*

$$Rce = \frac{\gamma_{YX}}{\text{dist}(Y, X)}. \tag{7}$$

III. Recognition of Sport Activities Using Template Matching

In this section, we present our data analysis and classification framework and explain the different processing steps carried out to classify sports activities using wrist acceleration data.

A. Data and System Overview

The dataset used in this study was collected in three different trials. In each trial, we used the same set of sensors but included different types of common activities, which could be mapped into eight main classes. Twenty-two primitive activity types were identified from the trials protocol and grouped in eight main categories: cycling, cross trainer, rowing, squatting, stepping, running, walking, and weight lifting, representing the target of our classification systems. The activities were carried out by 48 subjects (24 females and 24 males, age:

29 ± 9 years) wearing a triaxial accelerometer at the wrist (Philips DirectLife, Philips Consumer Lifestyle, Amsterdam, The Netherlands) characterized by a dynamic range of $\pm 2g$ and sampling frequency of 20 Hz. The population included 29 normal weight subjects (16 females and 13 males, body mass index BMI $= 22$ kg/m^2 on average) with a BMI below the threshold of 25 kg/m^2 and 19 overweight subjects (8 females and 11 males, BMI $= 27$ kg/m^2 on average) with a BMI higher than 25 kg/m^2. Data from 50% of the normal weight subjects (randomly selected) were used to generate the signal templates (training group). The same subset was also used for training statistical classifiers for comparison purposes as explained later. Data from the remaining participants as well as the entire overweight group were kept for testing purposes (test group). Because data were gathered in separate experiments, a different number of participants performed the considered activity types. Each activity lasted about 3 min except the squatting and weight lifting which had a duration of 45 and 60 s, respectively. Annotation on start and stop time of each activity were used to select intervals of the time series data belonging to each activity type. Visual inspection of the signal allowed to precisely detect activity onset and termination.

As explained in Section II, our approach consisted of two major steps: 1) generation of templates for each target activity type based on training data, and 2) classifying unseen data by comparing it to the set of generated templates and searching for the best match. Based on the collected data, a detailed description of these two phases are provided in the following sections.

B. Signal analysis

To test the viability of template-based activity classification, we investigated the periodicity of the acceleration signal using frequency-domain analysis. We designed a method independent from sensor orientation, thus we focused the analysis on a combination of the acceleration components x, y, z in the vector magnitude (VM)

$$\text{VM} = \sqrt{x^2 + y^2 + z^2}. \tag{8}$$

Examples of VM signals for different activities are shown in Fig. 2. From the VM signal of each activity type, we calculated the averaged power spectral density, using Welch's algorithm, and the spectral entropy, using the histogram method. The signal entropy was calculated as average of the spectral entropy values calculated for each signal of a specific activity. Low spectral entropy generally indicates high periodicity, while high spectral entropy indicates the absence of a repetitive pattern. The obtained results are reported in Table I. As expected from targeted sport activities, the highest entropy was related to cycling. Unlike the others, this activity does not induce a cyclic wrist movement.

C. Template Generation

Signal templates were generated by an automatic process consisting of two phases: 1) calculation of the template length then 2) extraction of relevant repetitive waveform in the signal. The length $2n$ of a template corresponds to the mean repetition period of each example signal belonging to the same type, which is expressed in number of samples n. It is calculated by averaging the peak-to-peak distance in the autocorrelation function of each signal. This first phase is illustrated in Fig. 3. The second phase of template generation encompasses several steps as illustrated in Fig. 4 and summarized in Algorithm 1. To extract relevant and repetitive waveforms from the example signals, the following steps were undertaken.

1) Normalizing all the instances by the *standard score*. Given y as a signal instance, $\hat{\mu}_y$ as a mean estimator of y, and $\hat{\sigma}_y$ as an estimator of standard deviation, the score normalization is calculated as follows:

$$y_{\text{norm}} = \frac{y - \hat{\mu}_y}{\hat{\sigma}_y}. \tag{9}$$

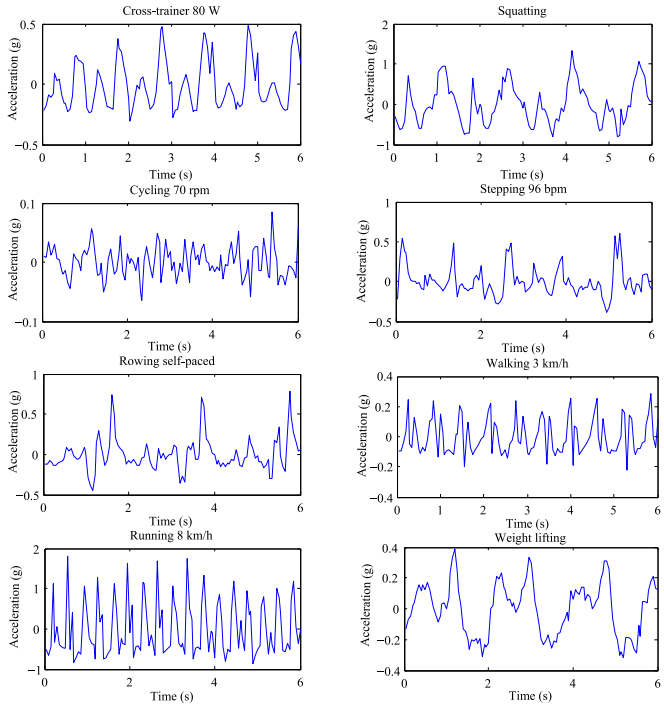

Fig. 2. Plots show segments of centered VM signals extracted from eight sports activities recorded with an accelerometer located at the wrist. The depicted activities are: Cross trainer 80 W, cycling at 70 r/min, rowing self-paced, running at 8 km/h, squatting, stepping 96 b/min, walking at 3 km/h, and weight lifting.

TABLE I
AVERAGE SPECTRAL ENTROPY CALCULATED PER ACTIVITY CLASS FOR TRAINING AND TEST DATA

	Entropy-Training set		Entropy-Test set	
Activity	Mean	SD	Mean	SD
Cycling	4.20	0.35	4.20	0.28
Cross trainer	2.47	0.75	2.55	0.72
Rowing	3.25	0.58	2.96	0.69
Running	1.10	0.24	1.28	0.33
Squatting	1.65	0.15	1.93	0.24
Stepping	3.01	0.26	2.96	0.30
Walking	1.96	0.58	2.02	0.65
Weight lifting	1.47	0.13	1.71	0.41

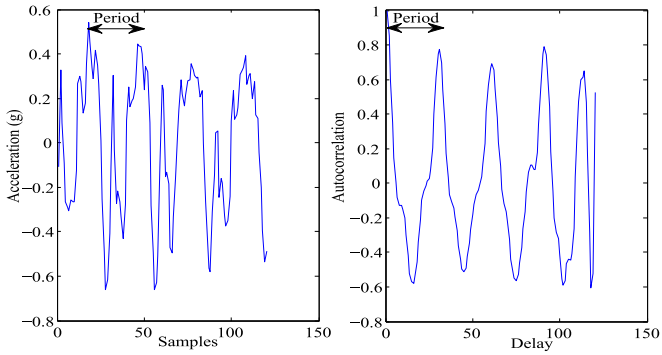

Fig. 3. Example of autocorrelation function (to the right) applied to VM signal recorded during squatting (to the left) for the extraction of the signal period. The autocorrelation function shows peaks reproducing the periodicity of the original signal.

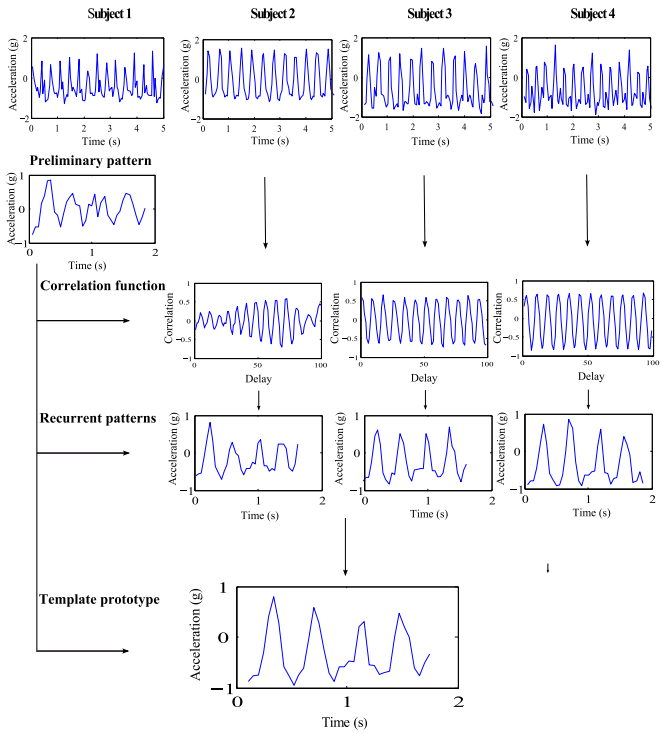

Fig. 4. Illustration of the template generation algorithm. The first row of plots shows segments of running at 8 km/h belonging to the subjects included in the training set. The second row shows the $2n$ segment extracted from the signal of subject 1 and representing the *preliminary pattern*. The third row shows a segment of the cross-correlation function calculated between the *preliminary pattern* and the others three signals. The fourth row shows the new $2n$ segments selected according to the analysis between cross correlation maximum peaks. The final row shows the template obtained by outliers correction and averaging over all the signal segments.

2) Selecting one instance for each activity type and extraction of a segment composed by $2n$ samples from a stationary activity phase (e.g., at the middle of the signal).
3) Cross correlating between the segments extracted in the previous step and all the others instances belonging to the same activity.
4) Extracting signal segments with length $2n$ samples from each instance starting from the sample corresponding to the maximum cross-correlation coefficient.

Algorithm 1 Template generation

Require: Activity signals set, dim
Ensure: Template
$N \leftarrow num(Signals)$
$k \leftarrow 1$
{Build a matrix where each row stores a segment of each signal}
$Smatrix \leftarrow zeros(N, dim)$
{Extraction of the first segment from the middle of the first signal}
$S_1 \leftarrow Signals(1)$
$start \leftarrow length(S1)/2 - dim/2$
$end \leftarrow start + dim/2 - 1$
$Segm_1 \leftarrow S_1(start : end)$
$Smatrix(k, :) \leftarrow Segm_1$
{Find max correlation between the first segment and the other signals}
for k = 2 to N **do**
  $S_k \leftarrow Signals(k)$
  $Ind_{max} \leftarrow corr(S_k, Segm_1)$;
  $start \leftarrow Ind_{max}$
  $end \leftarrow start + dim - 1$
  $Smatrix(k, :) \leftarrow S_k(start : end)$
end for
{Outliers correction based on statistic distribution}
for $j = 1$ to dim **do**
  $q \leftarrow quantile(Smatrix(:, j))$
  for $i = 1$ in N **do**
    if $(Smatrix(i, j) < q_{0.25}$ **or** $SMatrix(i, j) > q_{0.75}$ **then**
      $Smatrix(i, j) \leftarrow q_{0.50}$
    end if
  end for
end for
{Averaging over the Smatrix rows}
$Template \leftarrow sum(SMatrix, 1)/N$
Return $Template$

5) Construction of a statistical distribution of values for each sample in the $2n$ length of the segments extracted at step 4 from each example signal.
6) Calculating the $quantiles = \{q_{0.25}, q_{0.50}, q_{0.75}\}$ for each sample among the segments.
7) Removing outliers by deleting samples from the $2n$ segments and substituting those with $value > q_{0.75}$ or $value < q_{0.25}$ with $q_{0.50}$.
8) Averaging over all the signal segments to obtain the template prototype for the activity.
9) Rescaling the signal amplitude based on activity-type-specific amplitude ranges.

D. Template Matching

Once the activity templates have been generated, any unknown signal could be classified into the target activity classes by comparing it to each of these templates. In order to speed up

the classification, we first applied a discriminating rule based on the amplitude range of the instances. This step also allows to simply distinguish between signals with similar waveforms, yet clearly different amplitude range. The acceleration signal was segmented in windows of 120 samples and was compared to the template set using the five similarity functions explained in the background section: Euclidean distance, DTW, DDTW, correlation, and *Rce*. A dc offset removal was applied to each window of acceleration signal and a moving average filter was applied before computation of the DDTW distance, in order to prevent noise enhancement due to the calculation of signal derivative.

E. Activity Classification Performance

The accuracy of the classification algorithm was evaluated by calculating for each activity class the true positives rate, known as *sensitivity*, and the true negative rate, known as *specificity*. To evaluate the robustness of the presented methods, two different target populations have been tested. The first consisted of 29 normal weight subjects and the second of 19 overweight subjects. One challenging aspect of the collected data is that the number of participants varied from one activity to another. Indeed, while a total of 48 participants performed walking activities, only 8 engaged in squatting exercise. This resulted in an unbalanced training data across activity classes.

We applied a nonoverlapping windowing technique to process the acceleration data. The windows consisted of 120 samples that with a sampling rate of 20 Hz corresponded to a duration of 6 s. This window size was chosen in agreement with previous studies on activity recognition based on statistical features [8], which assessed that windows approximately of 6 s are sufficient to characterize different types of body motion and capture activity periodicity. The total number of windows was 7848 for normal weight subjects data and 4187 for the overweight one. Twenty-two templates have been generated, one for each primitive activity type in the protocol, and grouped in eight classes in accordance with the number of target classes. Therefore, each activity category was represented by multiple templates. The classification algorithms were tested on 15 normal weight subjects (3994 epochs) and on the entire overweight population (4187 epochs). The overweight subjects were excluded from the training set and used for testing in order to investigate whether different biomechanical characteristics due to excess weight had an impact on the performance of the activity classification system built on data extracted from normal subjects.

IV. Experimental Results and Discussion

A. Template-Matching Metrics for Activity Classification

In our framework, we compared the activity classification accuracy of five different template-matching metrics listed in Tables II and III. The true negative rate was above the 90% for all the classification methods, while sensitivity varied substantially. The results suggest that both correlation-based matching techniques and *Rce* generally outperformed the similarity measures (Euclidean, DTW, and DDTW). Sensitivity for the correlation metric and *Rce* was above 80% for the majority of activities on both test datasets, normal weight and overweight subjects. The classification accuracy for certain activity types worsened when using DTW and DDTW metrics as compared to Euclidean distance which could indicate the presence of $singularities$ generated by the DTW and DDTW algorithms (see Section II-B) during waveforms alignment, which leads to a higher cumulative distance. Poor classification accuracy was obtained by the DTW for squatting (sensitivity $= 8.7\%$) on the normal weight population. This may be due to increased distance metric and suboptimal template-signal alignment related to differences in amplitude scales between the template prototype and the waveforms used as test set. Correlation measures and DDTW, which are less sensitive to differences in signal amplitude, showed larger classification accuracy for squatting. Nonetheless, all matching techniques showed a clear variability in the recognition accuracy between activity types. Despite the lack of abundant training data, the squatting activity type could be well classified by the *Rce method* (see Fig. 6). Indeed, the squatting class showed a large recognition accuracy with template matching even if only two subjects were available in the template generation phase. The template extracted from the squatting acceleration signal was well representative and specific for such activity type regardless the amount of example data available.

The poor classification accuracy obtained for some periodic activities, such as cross trainer and rowing, might be explained

TABLE II
Template-Matching Performance Evaluated on a Normal Weight Population

	Sensitivity%				
Activity (*)	Euclidean	DTW	DDTW	Correlation	*Rce*
Cycling (14)	82.2	87.7	83.2	48.2	88.4
Cross trainer (12)	19.8	15.5	19.5	5.0	7.6
Rowing (12)	38.3	43.4	54.0	10.6	52.3
Running (12)	86.0	70.5	62.5	94.0	73.8
Squatting (2)	73.9	8.7	47.8	82.6	91.3
Stepping (12)	10.5	14.8	39.9	86.0	68.1
Walking (14)	80.4	23.4	72.5	81.9	87.4
Weight lifting (4)	62.2	37.8	56.8	83.8	73.0

*Number of subjects per activity included in the training set.

TABLE III
Template-Matching Performance Evaluated on an Overweight Population

	Sensitivity%				
Activity(*)	Euclidean	DTW	DDTW	Correlation	*Rce*
Cycling (14)	85.9	91.6	87.1	54.7	92.4
Cross trainer (12)	39.3	15.2	32.6	10.9	12.9
Rowing (12)	49.4	54.4	60.2	6.5	58.7
Running (12)	85.3	77.2	76.4	95.8	84.8
Squatting (2)	68.2	45.5	36.4	86.4	100.0
Stepping (12)	58.7	10.6	10.2	80.6	79.7
Walking (14)	80.5	24.3	70.4	88.0	91.2
Weight lifting (4)	57.7	75.7	79.3	63.1	64.0

*Number of subjects per activity included in the training set.

TABLE IV
LIST OF SELECTED FEATURES

Features	Domain
Mean	Time
Variance	Time
Root Mean Square of the derivative signal	Time
Range (maximun-minimun)	Time
Total energy	Time
Skewness	Time
Main frequency	Frequency
Entropy	Frequency
Quantile 0.25	Frequency
Quantile 0.50	Frequency
Quantile 0.75	Frequency

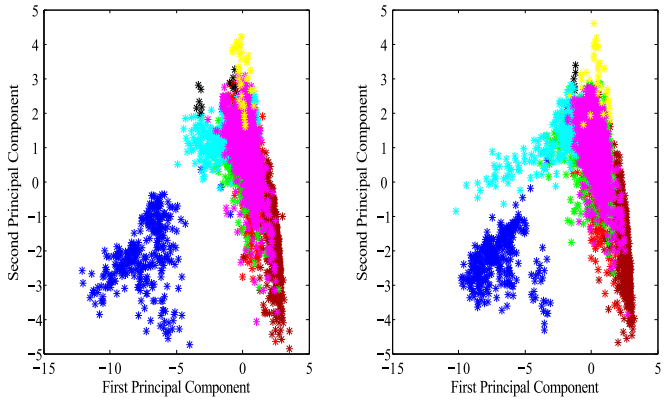

Fig. 5. Distribution of the selected features represented into a two-dimensional space. The first plot (to the left) shows the projection of the training data, while the second plot (to the right) shows the projection of the test data. Each region belongs to a specific activity. Brown: Cycling; Red: Cross trainer; Green: Rowing; Blue: Running; Black: Squatting; Cyan: Stepping; Magenta: Walking; Yellow: Weight lifting.

by the subject-specific interaction with the gym equipment employed in the execution of the activity. Indeed, unforeseen wrist acceleration patterns may have been present in the test datasets due to peculiar preference in the usage of supporting handlebars. This may depend on the participant skill and familiarity with the exercise, as no specific instructions were given to the user on how to perform the activities. Conversely, the cycling activity was recognized with high accuracy despite its nonperiodic nature. This was due to its discriminatory amplitude range that differs it from other types of exercise as explained in the previous section. Combining distance-based and correlation-based techniques (*Rce*) yielded the highest average classification results, while showing reproducible accuracy in both test datasets (see Table V).

B. Template-Matching Metrics Versus Classic Statistical-Learning Algorithms

To gain better insights into the viability of the presented template-matching approach, we compared the classification results to those of common classification algorithms for physical activity recognition. These included linear classifiers (NB and LR), DT and ANN. A set of 11 acceleration features were selected using the Relief method (see Table IV) to obtain a ranking among the 13 most commonly used time- and frequency-domain features. A principal component analysis representation of the data [31] is illustrated in Fig. 5. Such analysis shows a clear distinction of the different activity classes, except for squatting since this activity was poorly represented in the training data, as mentioned earlier. The classification results for all the proposed methods are shown in Table V. Statistical-learning algorithms (DT, NB, LR, ANN) only slightly outperformed the best template-matching method (*Rce*). The ANN and LR achieved the highest recognition accuracy ($\sim$85%) showing to be more accurate than the other proposed techniques. The LR results deserve particular attention since this model offered recognition rate close to the ANN, despite the fact that is a simpler classifier and includes a much lower number of parameters which need to be optimized during training. When comparing statistical-learning algorithms to template-matching methods, the first ones were less robust when tested on unseen data. In fact, we observed higher generalization ability for the template-based methods as

TABLE V
OVERALL CLASSIFICATION PERFORMANCE

	Aggregated Sensitivity%		
Classifier	Normal	Overweight	Δ
Euclidean	66.1	72.6	6.5
DTW	43.4	47.5	4.1
DDTW	64.9	64.7	−0.2
Correlation	62.3	63.9	1.6
Rce	74.7	78.7	4.0
DT	81.9	80.6	−1.3
NB	79.6	79.7	0.1
LR	84.6	83.5	−1.1
ANN	86.7	85.9	−0.8

compared to DT, LR, and ANN whose accuracy decreased when applied to the overweight subjects test dataset. Each template-based classifier and the NB maintained their classification performance when tested on both normal weight and overweight subjects. The activity-specific accuracy of the methods as obtained on data from the overweight population is depicted in Fig. 6.

The need for abundant and representative training data is especially clear for the squatting activity type, which achieved a high recognition rate only with the *Rce* method. The results confirmed the robustness of the template-matching approach against deterioration in classification accuracy due to previously unseen data and subjects. This aspect may increase the likelihood of obtaining large recognition accuracy in free-living conditions, where the performance of statistical-learning algorithm usually decreases. Several advantages of the template-matching method emerged from our analysis. First, template matching showed high robustness with respect to unseen data: Classification accuracy was preserved when the models were applied to the overweight subject data. The template-matching accuracy was unaffected by small training datasets: High classification accuracy could be obtained for the squatting and weight lifting categories with training examples from only two to four subjects.

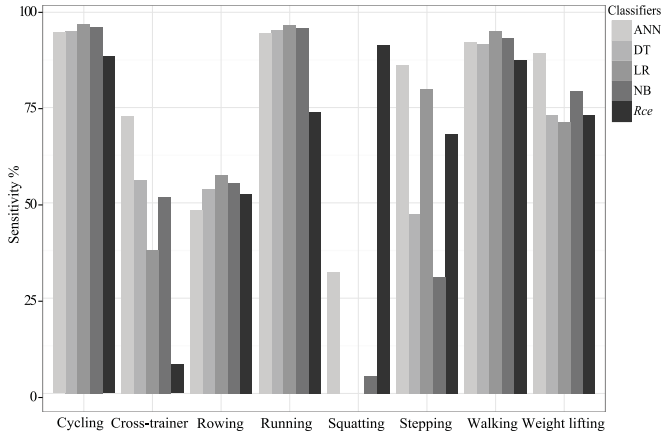

Fig. 6. Classification accuracy for the overweight population.

Thus, template matching successfully discriminated most of the sport activities by using user-independent and automatically extracted template prototypes. On the other hand, template matching was poorly effective for classifying activities characterized by high signal spectral entropy (e.g., wrist acceleration signal during cycling) for which a repetitive template could not be determined. This suggests that activity classification based on template matching alone may be inadequate for recognizing certain activity types given a sensor wearing position. Additionally, the use of template matching for embedded signal processing in wearable sensors may be restricted by its computational complexity. Given these strengths and weaknesses, combining template matching to the output of a statistical-learning algorithm could lead to superior performance in filed applications [32]. Template matching can compensate the problem arising from poor features separation for activities showing similar periodicity, regularity, and intensity but different morphology (e.g., walking and cross trainer). At the same time, template matching may compensate statistical features vulnerability to discriminate between activities carried out at different and previously unseen speed or intensity. Indeed, classic temporal and frequency features may substantially vary within the same activity category with changing speed and execution intensity while the main repetitive pattern would remain unchanged.

V. Conclusion

In this paper, we investigated the use of template matching for classifying sport activities using one single accelerometer worn at the wrist. User-independent signal templates were created and five distance- and correlation-based matching techniques were proposed for activity classification to target eight sports activities, such as cycling, cross trainer, rowing, squatting, stepping, running, walking, and weight lifting. The framework was evaluated using data collected in a gym environment involving two different groups of volunteers: normal weight and overweight subjects. The viability of our classification approach was analyzed by comparing its recognition accuracy with that of four popular classifiers: DT, NB, LR, and ANN. Based on this comparison, we concluded that template matching is well suited for the recognition of sport activities due to their inherent periodic nature. In particular, our template-matching-based method showed to be robust to data generated by previously unseen subjects with different biometric characteristics and possibly motor skills. The overall classification accuracy of the best (*Rce*) template-matching metric was lower than that offered by the statistical classifiers, which can be imputed to the lower accuracy in classifying cross trainer and rowing activities. However, a template-matching framework preserved classification accuracy during testing, whereas several statistical-learning algorithms did not. Future work should focus on exploring the added value of measurement-axis-specific templates for improving the classification accuracy and on investigating the advantages of combining template-matching metrics with statistical classifiers.

References

[1] M. Chen *et al.*, "Body area networks: A survey," *Mobile Netw. Appl.*, vol. 16, no. 2, pp. 171–193, Apr. 2011.

[2] A. G. Bonomi and K. R. Westerterp, "Advances in physical activity monitoring and lifestyle interventions in obesity: A review," *Int. J. Obesity*, vol. 36, no. 2, pp. 167–177, 2011.

[3] A. Avci *et al.*, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *Proc. 23rd Int. Conf. Archit. Comput. Syst.*, 2010, pp. 1–10.

[4] L. Gao *et al.*, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems," *Med. Eng. Phys.*, vol. 36, no. 6, pp. 779–785, 2014.

[5] L. Atallah *et al.*, "Sensor positioning for activity recognition using wearable accelerometers," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 4, pp. 320–329, Aug. 2011.

[6] M. Altini *et al.*, "Estimating energy expenditure using body-worn accelerometers: A comparison of methods, sensors number and positioning," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 219–226, Jan. 2015.

[7] A. G. Bonomi *et al.*, "Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer," *J. Appl. Physiol.*, vol. 107, no. 3, pp. 655–661, 2009.

[8] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*. New York, NY, USA: Springer-Verlag, 2004, pp. 1–17.

[9] M. Mathie *et al.*, "Detection of daily physical activities using a triaxial accelerometer," *Med. Biol. Eng. Comput.*, vol. 41, no. 3, pp. 296–301, 2003.

[10] S. J. Preece *et al.*, "Activity identification using body-mounted sensors: A review of classification techniques," *Physiol. Meas.*, vol. 30, no. 4, pp. R1–R33, 2009.

[11] A. Mannini *et al.*, "Activity recognition using a single accelerometer placed at the wrist or ankle," *Med. Sci. Sports Exercise*, vol. 45, no. 11, pp. 2193–2203, 2013.

[12] L. Barros *et al.*, "Wearable computing: Accelerometers data classification of body postures and movements," in *Advances in Artificial Intelligence—SBIA 2012* (Ser. Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 2012, pp. 52–61.

[13] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single tri-axial accelerometer," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1780–1786, Jun. 2014.

[14] S. I. de Vries *et al.*, "Evaluation of neural networks to identify types of activity using accelerometers," *Med. Sci. Sports Exercise*, vol. 43, no. 1, pp. 101–107, 2011.

[15] M. Ermes *et al.*, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Trans. Inform. Technol. Biomed.*, vol. 12, no. 1, pp. 20–26, Jan. 2008.

[16] I. C. Gyllensten and A. G. Bonomi, "Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2656–2663, Jun. 2011.

[17] F. Sartor *et al.*, "Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings," *Sports Med.*, vol. 43, no. 9, pp. 865–873, 2013.

[18] R. Muscillo *et al.*, "Early recognition of upper limb motor tasks through accelerometers: Real-time implementation of a DTW-based algorithm," *Comput. Biol. Med.*, vol. 41, no. 3, pp. 164–172, 2011.
[19] C. Chen and H. Shen, "A feature evaluation method for template matching in daily activity recognition," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput.*, 2013, pp. 1–4.
[20] T. Stiefmeier *et al.*, "Fusion of string-matched templates for continuous activity recognition," in *Proc. IEEE 11th Int. Symp. Wearable Comput.*, 2007, pp. 41–44.
[21] R. Muscillo *et al.*, "Classification of motor activities through derivative dynamic time warping applied on accelerometer data," in *Proc. IEEE 29th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2007, pp. 4930–4933.
[22] R. Brunelli, *Template Matching Techniques in Computer Vision: Theory and Practice*. Hoboken, NJ, USA: Wiley, 2009.
[23] L. Deng *et al.* "Structure-based and template-based automatic speech recognition—Comparing parametric and non-parametric approaches," in *Proc. INTERSPEECH*, 2007, pp. 898–901.
[24] H. Zhang *et al.*, "Gait modeling and identifying based on dynamic template matching," *J. Comput. Inform. Syst.*, vol. 7, no. 4, pp. 1155–1162, 2011.
[25] J. Tou and R. Gonzales, *Pattern Recognition Principles*. Boston, MA, USA: Addison-Wesley, 1974.
[26] V. Guardabasso *et al.*, "A versatile method for simultaneous analysis of families of curves," *FASEB J.*, vol. 2, no. 3, pp. 209–215, 1988.
[27] C. Vercellis, *Business Intelligence:Data Mining and Optimization for Decision Making*. Hoboken, NJ, USA: Wiley, pp. 296–297, 2009.
[28] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proc. SIAM Int. Conf. Data Mining*, 2001, vol. 1, pp. 5–7.
[29] M. Akay. (2012). Biomedical Signal Processing. [Online]. Available: http://books.google.co.uk/books?id=wr8BjI7tBv4C
[30] W. Deburchgraeve *et al.*, "Automated neonatal seizure detection mimicking a human observer reading EEG," *Clin. Neurophysiol.*, vol. 119, no. 11, pp. 2447–2454, 2008.
[31] I. Jolliffe. (2005). Principal Component Analysis. [Online]. Available: http://tocs.ulb.tu-darmstadt.de/182444961.pdf
[32] P. Siirtola *et al.*, "Improving the classification accuracy of streaming data using sax similarity features," *Pattern Recog. Lett.*, vol. 32, no. 13, pp. 1659–1668, 2011.