Activity Recognition on 3D Accelerometer Data (Technical Report)

Thomas Bernecker, Franz Graf, Hans-Peter Kriegel, Christian Moennig, Dieter Dill and Christoph Tuermer

Abstract

Physical activity is an important factor that is considered for the prevention of diseases like diabetes or hypertension and for rehabilitation. For this purpose, patients are required to fulfill a certain, regular pensum of activity which follow a training schedule that is integrated into daily life which cannot be supervised. In order to obtain reliable statements about achieved physical activity within a certain time period, accelerometers can act as tools that provide results which are superior to questionnaires filled by the patients. State-of-the-art activity recognition systems already provide good results, but the accuracy of classification algorithms including data preprocessing often depend on the position of the sensors and the data. This paper proposes an activity recognition system designed for accelerometers positioned at the ankle, which, in general, yields accurate and representative acceleration time-series data. An extensive evaluation on real-world datasets shows that, performing an additional reclassification step, excellent classification results outperforming recent work can be achieved.

I. Introduction

Physical activity becomes more and more important in the modern society. Nowadays, cardiovascular diseases cover a significant part of annually occurring affections, which is due to the reduced amount of activity in the daily life [7]. The automation of working processes as well as the availability of comfortable travel options may cause overweight [54], which may result in lifestyle diseases, such as diabetes mellitus [31]. Warburton et al. [50] showed that prevention and therapy of such diseases as well as the rehabilitation after affections or injuries can be supported by continuous and balanced physical activity. For this purpose, patients are required to fulfill a regular pensum of activity which follows a training schedule that is integrated into the daily life, but which cannot be supervised. In order to obtain reliable statements about achieved physical activity within a particular time period, accelerometers can act as tools that provide accurate results, as filled questionnaires tend to be strongly subjective [2], [52]. This statement is obvious, as, according to [14], the patients tend to overestimate their own abilities, which leads to results that are likely to be biased. Furthermore, the evaluation of the results is very complex and time-consuming. In order to improve the quality, i.e., the accuracy and the objectivity of these results, accelerometers serve as suitable devices for medical monitoring. The recordings of sensor observations allow the detection of any type of human motions

T. Bernecker, F. Graf, H.-P. Kriegel and C. Moennig are with the Ludwig-Maximilians-Universität, München, Germany. E-mail: {bernecker,graf,kriegel,moennig}@dbs.ifi.lmu.de. D. Dill and C. Tuermer are with the Sendsor GmbH, München, Germany. E-mail: {dill,tuermer}@sendsor.de.

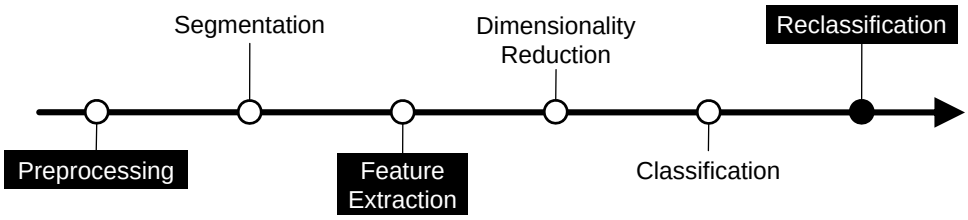


Figure 1. A visualization of the classification process.

that are composed of cyclic patterns. Cycling is, for example, a typical activity where cyclic movements repeatedly occur via pedaling; but periodic patterns can also be detected from other activities, such as walking, running, swimming and even working.

This paper will present an algorithm for activity recognition, where time series data derived from three-dimensional accelerometers is analyzed in order to classify the recorded patterns w.r.t. different types of activities. Here, the three dimensions reflect the direction of physical movement in each axis, i.e., forward/backward, left/right and up/down. The application scenario was published in a medical context [45], [46] within the collaboration with the *Sendsor GmbH*[1], who also provided the accelerometers used to record the datasets that will be used for the experimental evaluation of the approach that will be presented in this paper. These sensors measure acceleration amplitudes of $\pm 2g$.

Diverse pieces of work have shown that the position of an accelerometer has a significant influence on the results obtained while measuring physical activity [6], [38], [44]. The results of this research leads to varying interpretations, as the set of activities that have to be classified strongly depends on the problem definition. There is still no dedicated position where the measurements of an accelerometer are able to provide globally best results that are independent of the set of activities; however, it has been shown that accelerometers positioned at the ankle achieve superior recordings compared to other body positions [44].

Basically, activity recognition requires several steps of preprocessing before the classifier can separate different activities properly (cf. Section II). Avci et al. [4] provide a detailed survey of these steps. An overview of the general processing chain is shown in Figure 1. The main contributions of this paper, highlighted in Figure 1, consist of

- a reconstruction of the data peaks for the case that the measured acceleration values exceed the amplitude range,
- the usage of additional features to represent the recorded physical activities,
- and a reclassification step that corrects classification errors.

The rest of this paper is organized as follows. First, an overview of related work in the context of activity recognition is provided in Section II. The preprocessing steps performed on the time series will be summarized in Section III. Section V will give details about the used features for the classification. Applied techniques for dimensionality reduction of feature vectors will be presented in Section VI. A postprocessing step that corrects classification errors will be presented in Section VII. Section VIII will provide a broad experimental part containing the

[1]http://www.sendsor.de

Notation	Description
X	a one-dimensional time series
n	the length (number of observations) of X
t_i	the ith time stamp of X
x_i	the observation of x occurring at time t_i
X_{t_i}	the subsequence of X starting at time t_i
$\widehat{X}$	a reference sample of X
$\widehat{m}$	the length of $\widehat{X}$
Δ_{opt}	the optimal pattern length
Δ_{max}	the maximum pattern length
ρ_{opt}	the correlation obtained with Δ_{opt}
τ_ρ	the correlation threshold
S	a segment (subsequence) of X
m	the length of S
S_{max}	the maximum amplitude value of S
τ_{min}	the required minimum amplitude for peaks
v	a feature vector
d_v	the dimensionality of v

Table I
TABLE OF NOTATIONS FREQUENTLY USED IN THIS PAPER.

evaluation of the process chain against the recent approach. The experimental evaluation was performed using the data mining framework *Knowing* (*Know*ledge Engineer*ing*) [8]. Finally, Section IX will conclude this paper.

Table I provides an overview of the most frequent notations used in this paper.

II. RELATED WORK

In the context of activity recognition, a dedicated processing chain is performed, including preprocessing steps and the classification. In the following, related work about this process chain is summarized. A more detailed survey of these steps is given in [4].

A. Data Preprocessing

Recorded time series data from accelerometers often contains noise of high frequency, which in many cases distorts the actual signal. Thus, sliding-window-based average [24] or median filters [18] are applied in order to remove outliers. Furthermore, removing the effect of the gravitational force is supposed to distinguish activity from non-activity phases. This is in general obtained by applying a low-pass filter, as shown in [3], [18].

B. Segmentation

In order to separate periodic parts from nonperiodic parts, time series are divided into subsequent segments. In the literature, there exist different techniques for segmenting time series. Sliding-window-based methods [25], [39], [47], [48] are suitable for online processing and provide pattern matching algorithms starting with a reference sample that is extended until

the matching distance exceeds a particular threshold. Top-down approaches in the context of time series processing [29], [30], [42] recursively split a time series into two subsequences w.r.t. an approximation error threshold. Complementary approaches [20], [21], [22] work in a bottom-up manner, starting with $\frac{n}{2}$ segments of size 2 (where n denotes the number of observations) and combining adjacent subsequences until an upper cost bound is reached. A combination of the advantages of sliding-window, which is most efficient, and bottom-up, which provides superior segmentation quality, is provided within a survey on time series segmentation by Keogh et al. [19]

Sliding-window-based methods are still most frequently used in the area of activity recognition [4]. Thus, the segmentation method used this paper is based on a sliding-window algorithm.

C. Feature Extraction

Periodic and nonperiodic segments of time series are commonly described by a combination of features of different types.

- Time-domain features, such as mean, variance and standard deviation [28], [36], [51] or the *Root Mean Square* (*RMS*) [15], [32] are directly derived from the time series. Further prominent examples of this feature type are the average time between the peaks [28] and the number and average value of the peaks [51].
- Many publications apply well-known dimensionality reduction techniques by transforming the time series into the frequency domain. Frequency-domain features can be derived by the *Discrete Fourier Transform* (*DFT*) [1] or the enhanced *Fast Fourier Transform* (*FFT*) and are used in [6], [43]. Features like aggregated FFT coefficients or the entropy of the frequency domain [6], that distinguish activities where similar energy values are detected (e.g., running and cycling), or single FFT coefficients [27] are also used in existing literature.
- A combination of domains w.r.t. time and frequency is given by wavelet features, derived from the *Discrete Fourier Transform* (*DFT*) [1] and is used in the context of gait classification [34]. The extraction of frequency-domain (or combined) features requires additional computational effort for the time series transformation and will thus be omitted in this paper.
- Heuristic features cannot be directly derived from the time series, but require mathematical and statistical methods to be extracted from the three dimensions of accelerometrical time series data simultaneously. A prominent example here is the *Signal Magnitude Area* (*SMA*), which is defined by the sum of the absolute values of all axes within the current time window and that is used in several works [3], [10], [18], [24], [55]. A further feature of this class is given by the *Inter-Axis Correlation*, which is a suitable measure to distinguish between movements measured at different body parts [6]. However, the authors of [36] could prove that this feature performs inferior to the simple features like mean and standard deviation, as also shown in this paper.

The adequate combination of features is an important task, since the classification of the time series highly depends on a good representation.

D. Feature Vector Dimensionality Reduction

In order to reduce the computational effort of the classification process, dimensionality reduction is typically applied in order to remove redundant information; this decreases the size of the feature vectors. In the context of accelerometer data, the literature distinguishes between methods of *feature selection* and *feature transformation*, which can also be used in combination.

- Feature selection methods include, for example, methods based on *Support Vector Machines* (*SVMs* [26], e.g., applied in [49]), or the *forward-backward search* technique [57] (e.g., used in [36] and also in this paper.
- Feature transformation methods further support the separation of different classes. Commonly applied techniques here are the *Principal Component Analysis* (*PCA* [35], e.g., used in [55], [56]), the *Independent Component Analysis* (*ICA* [12], e.g., used in [33]) or the *Linear Discriminant Analysis* (*LDA* [13], e.g., used in [15], [24]).

E. Classification

Regarding the classification step, many different approaches have been used in the literature in the context of activity recognition. Several publications apply supervised classification methods based on pattern recognition and training phases, e.g., decision trees [6], [17], [18], *Hidden Markov Models* [51], *Gaussian Mixture Models* [3], k-NN classifiers [17], [36], *Naïve Bayes* classifiers [17], *Support Vector Machines* (*SVMs*) [26], or *Neural Networks* [24], [28]. Section VII will propose an additional step that improves the classification result. First, the processing chain is started with data preprocessing in the following section.

III. Preprocessing Steps

A. Outlier Removal

In order to remove outliers in the data that emerged from measurement or transmission errors, an average filter is commonly applied [24]. Thus, further processing techniques that are applied on the data are not influenced by noise, such as the time series segmentation (cf. Section IV) and the feature extraction, in particular the computation of the *Average Peak Amplitude* feature (cf. Section V).

B. Peak Reconstruction

In some cases, the sensor recordings may be incomplete. The approach that will be presented in this paper uses a sensor that measures acceleration amplitudes in the range of $\pm 2g$. However, very intense or fast movements with a higher acceleration value create amplitudes that exceed this range. For consecutive values that are beyond this range, these intervals are cut, yielding significant gaps. One solution to overcome this problem is, of course, to use a sensor that supports measurements of a higher amplitude range up to $\pm 12g$ [10]. However, in order to be independent of technical constraints, the following method provides a reconstruction of the original signal.

Algorithm 1 Peak Reconstruction: reconstructPeaks(X, o, $maxAmp$)

Require: X, o, $maxAmp$

$n \leftarrow |X|$ $i \leftarrow 1$, $j \leftarrow i + 1$
while $j \leq n$ **do**
  while $x_i = maxAmp$ **AND** $x_j = maxAmp$ **do**
    $j \leftarrow j + 1$
  end while
  $a \leftarrow i$, $b \leftarrow j - 1$
  compute Δ_{before}, Δ_{after} and Δ_{total} according to Eqs. (1), (2) and (3)
  $c \leftarrow b - a$
  $h \leftarrow \lfloor \frac{c}{2} \rfloor$
  for $i = 1 \rightarrow h - 1$ **do**
    $x_{a+i} \leftarrow x_{a+i} + (\sqrt{i} \cdot \Delta_{total})$
    $x_{b-i} \leftarrow x_{b-i} + (\sqrt{i} \cdot \Delta_{total})$
  end for
  if $c \mod 2 = 1$ **then**
    $x_{a+h} \leftarrow x_{a+h} + (\sqrt{h} \cdot \Delta_{total})$
  else if $|\Delta_{before}| > |\Delta_{after}|$ **then**
    $x_{b-(h-1)} \leftarrow x_{b-(h-1)} + \Delta_{total}$
  else
    $x_{a+(h-1)} \leftarrow x_{a+(h-1)} + \Delta_{total}$
  end if
  $i \leftarrow j + 1$, $j \leftarrow j + 2$
end while

The first step is to identify time intervals where the measured acceleration has exceeded the amplitude range. In these parts, at least two observations must exist with a maximum (minimum) amplitude of exactly $+2g$ ($-2g$). As this scenario is an improbable case, such a sequence is likely to be the result of truncated data. The missing data after truncations can be reconstructed using the preceding and following observations. Based on these values, the average gradients before and after a peak (Δ_{before} and Δ_{after}) are derived (cf. Equations (1) and (2)) and the average total gradient Δ_{total} can be computed (cf. Equation (3)).

$$\Delta_{before} = \frac{1}{o} \sum_{i=a-o}^{a} x_{i+1} - x_i \tag{1}$$

$$\Delta_{after} = \frac{1}{o} \sum_{i=b}^{b+o} x_i - x_{i+1} \tag{2}$$

$$\Delta_{total} = \frac{\Delta_{before} + \Delta_{after}}{2} \tag{3}$$

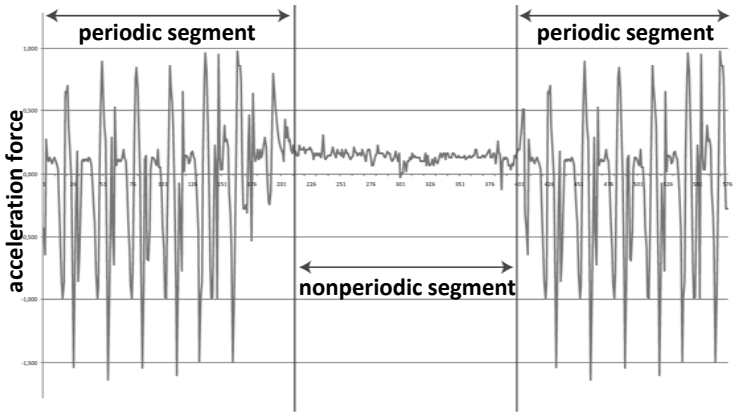


Figure 2. An example time series with periodic and nonperiodic segments.

Here, x_i corresponds to the observation measured at time t_i ($1 \leq i \leq n$) and n corresponds to the length of the time series. A truncated data peak is defined to start at time t_a and to end at time t_b. The variable o denotes the number of observations before and after the peak (respectively) that are considered for the computation of the missing data.

Algorithm 1 presents the steps of the peak reconstruction for one axis. The truncated peaks are determined while scanning the time series once (condition in line 2). If two successive maximum amplitudes are detected (the maximum depends on the sensor characteristics and can be set as a parameter $maxAmp$), the algorithm finds the whole truncated time range (line 4) and computes the average gradients (line 7). The variable c denotes the number of subsequent occurring maximum (minimum) values, which is 2 in the most simple case. Using the variable h, the first (line 11) and last $\lfloor \frac{c}{2} \rfloor$ values (line 12) are interpolated based on a weighted total gradient. If c is odd, then the central value in the peak, now represented by x_{a+h}, has to be recomputed in addition (line 15). Otherwise, both amplitudes $x_{a+(h-1)}$ and $x_{b-(h-1)}$ are equal. Then, a global extremal value within this segment is ensured by increasing one of the middle values (lines 16ff).

IV. Segmentation

Detecting periodic parts in time series is performed via segmentation, where a time series is divided into periodic and nonperiodic sequences (cf. Figure 2), in the following called *segments*. Both periodic and nonperiodic sequences are then processed separately based on extracted features (cf. Section V). For a periodic segment, it can be assumed that the detected activity holds for the entire time period, which is due to the periodicity in the signal. A nonperiodic segment may contain activity changes or indefinable acceleration recordings. A feature vector describing a segment is independent of the segment's length.

The first step of the segmentation is the detection of present periodicity. For this purpose, a commonly used method in the community of activity recognition is to apply a sliding window algorithm [25], [39], [47], [48], where autocorrelation [11] is used in order to measure the self-similarity of time series segments. In general, the autocorrelation $\rho(X, t_1, t_2)$ of a time

series X at two time stamps t_1 and t_2 is defined as

$$\rho(X, t_1, t_2) = \frac{E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})]}{\sigma_{t_1}\sigma_{t_2}}, \tag{4}$$

where X_{t_i} denotes the subsequence of X starting at time t_i and μ_{t_i} (σ_{t_i}) denotes the mean (variance) of X_{t_i} ($i = 1, 2$). Hereby, the length of the subsequence is limited to the current reference sample length $\widehat{m}$, which is defined below. The pseudocode of the segmentation algorithm is illustrated in Algorithm 2.

For each starting point of a potential period (denoted by the time stamp t_i, which is initially set to the first time stamp 1), the algorithm consists of two phases: Phase 1 determines the periodicity of the time series X of length n (lines 15-21), and Phase 2 extracts the periodic segments (lines 29-32). Therefore, X has to be scanned once completely (condition in line 2). In each iteration, a potential period is represented by a reference sample $\widehat{X}$ of length $|\widehat{X}| = \widehat{m}$, which is extracted from X (line 3). Before starting the process in the actual iteration, a maximum shift range has to be set in order to limit the length of a pattern period and, thus, to increase the matching chances. This range will be denoted by Δ_{max} in the following. Generally, Δ_{max} is set as an input parameter. However, its value has to be decreased if it exceeds the time series length (line 4f). Also, if $\Delta_{max} < 2 \cdot \widehat{m}$ holds, i.e., a shift of the reference sample length $\widehat{m}$ cannot be performed, all remaining observations are inserted in a temporary list called L_{np}, which collects all observations that are regarded to be nonperiodic (line 9). The elements will later be materialized as a nonperiodic segment (line 12). Otherwise, Phase 1 is started.

Phase 1, starting from line 15, detects the optimal window shift between periodic occurrences of patterns. The reference sample $\widehat{X}$ is now matched with a sliding window that contains the following observations of the time series within the maximum shift range Δ_{max}. The window shift Δ_{opt} yielding the highest correlation between $\widehat{X}$ and the pattern occurring in the window is considered as optimal. Thus, the optimal distance between patterns of a periodic cycle is given by Δ_{opt} observations. A sufficiently high correlation value is defined by a threshold τ_ρ, which is also an input parameter of the algorithm. Thus, if no correlation higher than τ_ρ is found within the maximum shift range Δ_{max}, the current optimal shift Δ_{opt} does not contain any significant periodic pattern. Then, the algorithm considers the current start time t_i to be nonperiodic (line 23) and continues with the next iteration using the respective sample sequence shifted by one time stamp, now starting at time stamp t_{i+1} (line 24).

If a periodic pattern (with $\rho_{opt} \geq \tau_\rho$) is found, the remaining observations in the L_{np} list, i.e., the observations that have before been marked as nonperiodic, are combined to a nonperiodic segment (line 26) and removed from the L_{np} list (line 27). Then, Phase 2 extracts periodic segments from the time series X, starting from line 29. The pattern is shifted by Δ_{opt}, yielding the new starting time t_j, and the correlation between the current reference sample and the following subsequence is computed. If a minimum correlation of τ_ρ between X_{t_i} and X_{t_j} is obtained (which is obvious with the first shift, as this was the "highest" correlation value that yielded the value for Δ_{opt}), the periodic segment is extended performing further shifts of length Δ_{opt}. This shifting procedure is continued until a correlation value less than τ_ρ is obtained or the periodic pattern reoccurs until the end of the time series is reached (condition in line 29). A periodic segment is required to have a minimum length m_p. If the extracted

Algorithm 2 Time Series Segmentation: segmentation(X, τ_ρ, Δ_{max}, m_p)

Require: X, τ_ρ, Δ_{max}, m_p
$i \leftarrow 1$, $L_{np} \leftarrow []$, $n \leftarrow |X|$
while $i \leq n$ **do**
  $\widehat{X} \leftarrow$ extractReferenceSample(X, t_i), $\widehat{m} \leftarrow |\widehat{X}|$
  if $i + \Delta_{max} = n$ **then**
    $\Delta_{max} \leftarrow n - i - \widehat{m} + 1$ {cut Δ_{max} if too long}
  end if
  if $\Delta_{max} < 2 \cdot \widehat{m}$ **then**
    while $i \leq n$ **do**
      L_{np}.add(x_i) {assign remaining observations to a nonperiodic segment}
      $i \leftarrow i + 1$
    end while
    createNonPeriodicSegment(L_{np})
  else
    $\Delta_t \leftarrow l$, $\rho \leftarrow 0$, $\rho_{opt} \leftarrow 0$, $\Delta_{opt} \leftarrow i$
    while $\Delta_t \leq \Delta_{max}$ **do**
      $\rho \leftarrow \rho(\widehat{X}, t_i, t_i + \Delta_t)$ {search for optimal shift Δ_{opt}}
      if $\rho > \rho_{opt}$ **then**
        $\rho_{opt} \leftarrow \rho$, $\Delta_{opt} \leftarrow \Delta_t$
      end if
      $\Delta_t \leftarrow \Delta_t + 1$
    end while
    if $\rho_{opt} < \tau_\rho$ **then**
      L_{np}.add(x_i) {τ_ρ was never exceeded}
      $i \leftarrow i + 1$
    else
      createNonPeriodicSegment(L_{np}) {materialize nonperiodic segment}
      L_{np}.clear()
      $\rho \leftarrow 0$, $j \leftarrow i$
      while $\rho \geq \tau_\rho$ **AND** $j \leq n - \widehat{m} + 1$ **do**
        $\rho \leftarrow \rho(\widehat{X}, t_i, t_j)$
        $j \leftarrow j + \Delta_{opt}$ {extend periodic segment}
      end while
      if $j - i + \widehat{m} \geq m_p$ **then**
        createPeriodicSegment(X, t_i, t_{j-1}) {materialize periodic segment}
      else
        createNonPeriodicSegment(X, t_i, t_{j-1}) {materialize nonperiodic segment}
      end if
      $i \leftarrow j$
    end if
  end if
end while

segment, thus, consists of m_p or more observations, a periodic segment is created (line 34). Otherwise, a nonperiodic segment is created (line 36). Afterwards, the algorithm restarts from the observation at the first time stamp after the extracted segment (line 38). The segmentation algorithm is finished if all observations have been assigned to a (periodic or nonperiodic) segment. This condition is satisfied after all time stamps of the time series have been tested (line 2).

The presented segmentation algorithm is generally restricted to a single time series. An extension for three-dimensional accelerometer data requires a slight modification. Here, the optimal pattern distance is chosen from the time series where the highest correlation value is obtained, and it is then applied to all time series. In order to mark a periodic segment, the correlation values ρ_{opt} of all three time series have to exceed τ_ρ (line 22).

In general, the autocorrelation method is sensitive w.r.t. slight deviations of the periodicity. Thus, an activity might be divided into several consecutive periodic segments showing different behavior w.r.t. frequency or intensity of the acceleration.

The runtime complexity of the segmentation algorithm is $O(n \cdot \ (\Delta_{max} - \widehat{m}))$. In the worst case, the segmenting process has to find the optimal shift within a range of $\Delta_{max} - \widehat{m}$ for each of the n observations.

V. Feature Extraction

As already stated in Section IV, the segments are of different length, so that classification cannot be directly performed based on the raw segments, since subsequence matching would require a high computational cost. Also, this step will not perform efficiently having segments that consist of a high number of observations. In order to overcome these problems, each segment is represented by a feature vector of the same dimensionality, where the choice of features can vary for periodic and nonperiodic segments. An overview of the final choice of features will be given in Subsection VIII-B.

The feature vectors used for the approach presented in this paper contain time-domain features and heuristic features, which will be summarized in the following. Frequency-domain features will not be considered in order to the save computational cost w.r.t. the transformation of the time series to the frequency domain.

For each periodic segment S, it is sufficient to derive only one feature vector v, as common characteristics for a segment follow directly from the periodicity of the observations contained in the segment. If a minimum length of m_p observations for S is assumed for each axis recorded by the accelerometer, the usage of feature vectors now reduces the dimensionality from at least $3 \cdot m_p$ values (regarding all three axes) to d_v, where d_v denotes the dimensionality of v and in general $d_v \ll m_p$ holds. In an exemplary case of $m_p = 100$ and $d_v = 15$ (which will be the final size for the experimental part), this corresponds to a reduction of at least 95%. For nonperiodic segments, it cannot be assumed that all observations were captured with the same physical activity, as no periodicity has been detected. Thus, a single feature vector would not represent the entire segment that well. In order to minimize this error, a nonperiodic segment is again split up into subsegments having a number of m_{np} observations, each represented by its own feature vector. In most cases, the last subsegment contains less than m_{np} observations,

since a nonperiodic segment, in general, varies in its length. If the last subsegment contains more than $\frac{m_{np}}{2}$ observations, another feature vector is computed. If this is not the case, this subsegment will be neglected in the classification step, as it hardly contains sufficiently enough information for the creation of some features, such as the *ARC* feature, which will be explained in the following. However, this subsegment will be considered again in the reclassification step, which is performed after the classification (cf. Section VII).

The following features will be examined in the context of the proposed approach.

1) Auto Regression Coefficients (ARC): The autoregressive model is commonly used in the context of modeling time series [11]. Let x_i be the observation of a time series X at time t_i. x_i is modeled as the sum of p linearly weighted recent values. The autoregressive model of order p is given by

$$x_i = \sum_{j=1}^{p} a_j \cdot x_{(i-j)} + \varepsilon_i, \tag{5}$$

where a_j is the jth autoregressive coefficient and ε_i is a noise term at time t_i. Given a time series X, the coefficients a_j can be estimated in various ways, such as the method of least squares [9].

Autoregressive coefficients predict the prospective course of a signal based on recent values and have been used in [23], [24] in the context of activity recognition. For the solution of this work, the first three coefficients for each axis of the accelerometer data are used, yielding nine feature values.

2) Signal Magnitude Area (SMA): The *Signal Magnitude Area* (*SMA*) is a well-known heuristic energy measure in the context of activity classification [3], [10], [18], [24], [55]. It is computed by summing up the absolute observation amplitudes of the three accelerometer axes and by normalizing the result w.r.t. the length of the corresponding segment S, i.e.,

$$SMA = \frac{1}{m} \sum_{i=1}^{m} (|x_i| + |y_i| + |z_i|). \tag{6}$$

Hereby, m corresponds to the length of S, and x_i, y_i and z_i are the values of the respective axis at time t_i. As the maximum amplitudes might vary for each axis, activities showing intense acceleration values occurring with high frequency contribute to a high SMA value, whereas low-acceleration activities result in a low SMA.

3) Tilt Angle (TA): The *Tilt Angle* (*TA*) feature is described by the average tilt angle of the lower leg over time. The accelerometer is supposed to be worn in the same position at the ankle. Hence, physical activity can be described by the angle ϑ between the gravitational vector and the positive z-axis of the sensor, i.e., $\vartheta = \arccos(z)$. Recognition of activities like swimming, which enforces a different tilt angle of the lower leg, takes considerable advantage of the TA feature. The TA feature has been used in [24] and will, thus, be examined in the experimental evaluation.

4) Average Peak Amplitude (APA): The *Average Peak Amplitude* (*APA*) will be introduced, which is an energy measure and is, in contrast to the SMA, restricted to the (positive and negative) peak amplitudes within a temporal window of the current segment S.

Algorithm 3 Peak Detection: detectPeaks(S, p_{min}, τ_{min}, Δ_τ, $minDist$)

Require: S, p_{min}, τ_{min}, Δ_τ, $minDist$

1: $S_{max} \leftarrow arg\,max_{x \in S}(S)$
2: $\tau \leftarrow S_{max}$
3: $P_{Peak} \leftarrow \{x \in S : x \geq \tau\}$
4: **while** $|P_{Peak}| < p_{min}$ **AND** $\tau > \tau_{min}$ **do**
5: $\quad \tau_{old} \leftarrow \tau$
6: $\quad \tau \leftarrow \tau - \Delta_\tau \cdot S_{max}$
7: $\quad P_{Peak}.add(\{x \in S : \tau_{old} > x \geq \tau\})$
8: $\quad$ **for all** $x_i, x_j \in P_{Peak}$ **do**
9: $\quad\quad$ **if** $t_i - t_j \leq minDist$ **then**
10: $\quad\quad\quad P_{Peak} \leftarrow P_{Peak} \setminus \{\min(x_i, x_j)\}$
11: $\quad\quad$ **end if**
12: $\quad$ **end for**
13: **end while**
14: $avg \leftarrow \frac{1}{|P_{Peak}|} \sum_{x \in P_{Peak}} x$
15: **if** $|P_{Peak}| \geq p_{min}$ **then**
16: $\quad$ **RETURN** avg
17: **else**
18: $\quad$ **RETURN** $sgn(avg)$
19: **end if**

The process of identifying the peaks is outlined in Algorithm 3. The signal has been cleaned w.r.t. outliers beforehand (cf. Subsection III-A), so that the set of detected peaks is not influenced by erroneous observations. The actual identification step first determines the global absolute extremum of S, denoted by S_{max} (line 1).

Next, a threshold τ is introduced, which defines a minimum amplitude for all potential peaks contained in S. τ is initialized with the value of S_{max} (line 2). The set of observations w.r.t. τ, denoted by P_{Peak}, is initially created (line 3) . As long as the algorithm has not yet detected a mandatory minimum number of peaks p_{min}, τ is decreased by $\Delta_\tau \cdot S_{max}$ (line 6), yielding new elements for P_{Peak} (line 7).

P_{Peak} may contain neighboring observations that actually belong to the same peak. In order to reduce the result to a single observation per peak, a minimum distance $minDist$ between peaks is introduced that has to hold. If two amplitudes identified as peaks show a distance less than $minDist$, the observation with the lower amplitude value will be removed (line 10; here, a numerical ordering w.r.t. the amplitudes is assumed for the observations).

The described procedure is repeated until p_{min} is reached or τ has reached a minimum value of τ_{min} (condition in line 4). In the latter case, S contains only few significant amplidudes.

Finally, the feature value represents the average of all peak amplitudes in the current segment, where the APA is only considered as significant if the number of peaks is at least p_{min} (lines 15ff). Otherwise, a default value depending on the amplitude sign is returned.

5) Surrounding Segmentation Rate (SSR): Physical activity classes can also differ in their fraction of periodic segments. For example, in the context of this work, the sensor recordings for the activity *Cycling* showed to contain more nonperiodic segments than *Running*, as periodic movements often have been interrupted by external influencing factors. This observation leads to the derivation of a simple, but suitable heuristic feature describing the *Surrounding Segmentation Rate* (*SSR*). The computation of the SSR is performed for a temporal window of w_{SSR} seconds surrounding the current segment, which is in particular suitable for long-term activities. Thus, for a window containing overall s segments, s_p and s_{np} denote the numbers of periodic and nonperiodic segments, respectively; it holds that $s = s_p + s_{np}$. Then, the SSR is computed by $SSR = \frac{s_p}{s}$.

VI. Dimensionality Reduction

A. Feature Selection

In order to reduce the computational effort of the actual classification process, a dimensionality reduction of feature vectors is typically applied in order to remove redundant information; this decreases the size of the feature vectors. In the context of accelerometer data, the literature distinguishes between methods of *feature selection* and *feature transformation*, which can also be used in combination. The selection of most relevant features for the feature vector was performed using the *forward-backward search* [57]. This method can be applied to reduce the effort of testing all feature combinations, which would be exponential in the number of features. This is achieved by alternately adding features to the feature vector with the currently best quality (which yields a new feature combination) and removing features (to examine subsets of combinations that have not been considered before).

B. Feature Transformation

The separation of different activity classes can further be supported applying the *Linear Discriminant Analysis* (*LDA*) [13] on the feature vector after the feature selection step. This is called *feature transformation*. The LDA minimizes the variance of features within a class and maximizes the variance of features between different classes, having the side effect of slight performance gain. Applying the benefits of the LDA to the current application scenario, this step neglects person-specific differences with the same physical activity, which is caused by different body heights or variations in the execution of movements. Finally, the LDA leads to a robustness of the classification w.r.t. the exact position of the sensor. Despite the fact that the sensor is fixed on the ankle, continuous movements can lead to a rotation or a shift of the sensor, which influences the quality of the data and, thus, the quality of the classification results. Applying the LDA, these errors can be corrected.

VII. Reclassification

In order to classify short subsegments where no features were extracted (cf. Section V), a postprocessing step is applied which assigns the most likely class label to these subsegments. This likelihood depends on the classification results obtained for the surrounding segments.

Algorithm 4 Reclassification of Observation x_i: reclassObservation(x_i, W, cl)

Require: x_i, W, cl

1: $w \leftarrow |W|$
2: $\mathcal{I} \leftarrow \{0, \ldots, cl-1\}$
3: $\mathcal{W} \leftarrow [0, \ldots, 0]$
4: $L_{old} \leftarrow x_i.label, L_{new} \leftarrow -1$
5: **for all** $x_j \in W \setminus \{x_i\}$ **do**
6: $\quad x_j.weight \leftarrow \frac{w-1}{2} - |j-i| + 1$
7: $\quad \mathcal{W}[x_j.label] \leftarrow \mathcal{W}[x_j.label] + x_j.weight$
8: **end for**
9: $L_{new} \leftarrow arg\,max_{l \in \mathcal{I}}(\mathcal{W})$
10: **if** $L_{old} \neq -1$ **OR** $L_{old} \neq L_{new}$ **AND** $\mathcal{W}[L_{new}] > \sum_{l \in \mathcal{I} \setminus \{L_{new}\}}(\mathcal{W}[l])$ **then**
11: $\quad x_i.label \leftarrow L_{new}$
12: **end if**

Furthermore, this step detects errors that occurred in the actual classification step. These errors may contain highly improbable results. The application of activity recognition provides sufficiently interpretable information for these cases. For example, if two significantly long segments classified as *Cycling* contain a short segment classified as *Elliptical Trainer*, the classification result of the latter segment will be revised. A formal description of the unsupervised reclassification is outlined in Algorithm 4.

Hereby, the variable cl denotes the number of activity classes; also, the activity class labels are represented as indices. Temporally used data structures are the list of class label indices $\mathcal{I}$ and the list of weights $\mathcal{W}$ of the class labels; the weight values are referenced by the respective class label index $i \in \{0, \ldots, cl-1\}$. The reclassification step takes, for each observation, the available information of the neighboring observations into account by considering a temporal window W of size $w = |W|$, containing the current observation x_i as well as $\frac{w-1}{2}$ preceding and $\frac{w-1}{2}$ successive observations. For observations that are close to the border of the time series, W is cut off accordingly. Based on the class labels of each observation x_j contained in W ($j \neq i$), a weighted linear distribution of the occurring labels is computed, which considers more recent observations to have more impact. Thus, the distance-based weight $x_j.weight$ of a neighboring observation x_j corresponds to $\frac{w-1}{2}$, whereas the weight of the most distant observation is 1 (line 6). The distribution of the weights of the observations x_j corresponds to a linear time-fading function. A quadratic or general distribution-based fading function would also be applicable here. If x_i has not been classified before (the label obtained in the classification is denoted by L_{old}, where a value of -1 implies no assignment to a class label) or the class label L_{new} that shows the highest weighted occurrence has a significant influence on x_i (i.e., its relative weighted occurrence is higher than the sum of all other classes), the reclassification was successful and L_{new} is assigned to x_i (line 11).

For the reclassification of each of the n observations, the surrounding $w-1$ observations within the window W have to be regarded; thus, this algorithm requires a runtime complexity of $O(n \cdot\ w)$.

Activity	# Datasets	Duration **(hh:mm:ss)**
Ell. Trainer	3	00:55:07
Walking	6	02:42:09
IL Skating	3	01:55:50
Running	6	02:37:13
Cycling	4	02:32:46
Total	22	10:43:05

Table II
DATASETS USED FOR THE EXPERIMENTAL EVALUATION.

VIII. EXPERIMENTAL EVALUATION

A. Datasets

The application scenario for this paper was given by a collaboration with the *Sendsor GmbH*, who also provided the accelerometers used to record the datasets that will be used in this section. The accelerometers are able to record amplitudes in the range of $\pm 2g$ with a rate of 25 Hz. In order to obtain accurate and representative acceleration measurements, the accelerometer is worn by the patients at the ankle [44].

In the context of this paper, five different activity classes were examined: *Walking*, *Running*, *Cycling*, *In-line Skating* (*IL Skating*) and *Elliptical Trainer* (*Ell. Trainer*). The datasets used for the following experiments are summarized in Table II. The evaluation was performed using the data mining framework *Knowing* [8].

B. Experimental Setup

1) Choice of the Classifier: The evaluation of the presented activity classification approach was performed using the *Naïve Bayes* classifier [17]. In the context of implementing the *Knowing* framework, an evaluation of overall 32 classifiers that are available in *WEKA* [16] has been performed, where Naïve Bayes turned out to provide most effective solutions on periodic and nonperiodic segments. Hereby, the effectiveness of the classifier was measured by the *classification accuracy*, which is a common measure in the field of information retrieval [5].

The first experiment was performed without applying reclassification. Naïve Bayes yielded a classification accuracy of 97.18% (more details will be provided in Subsection VIII-C). Results of slightly minor quality could were obtained using Sequential Minimum Optimization [37] (accuracy of 96.67%) and a normalized Gaussian radial basis function network (accuracy of 94.88%).

In addition, two methods based on *Artificial Neural Networks* (*ANNs*) were tested in order to provide the comparability to the approach of [24]. The latter uses a multilevel perceptron based on *backpropagation learning* [41], [53], which is available in *WEKA*. The second evaluated ANN, which is based on *resilient propagation learning* [40], is available in the *Encog Machine Learning Framework*[2]. In the evaluated settings, each of the neural networks consisted of a hidden layer of ten neurons and an output layer of five neurons, which corresponds to

[2]http://www.heatonresearch.com/encog

Description	Abbreviation	# Values	[24]	This Approach
Auto Regression Coefficients	ARC	9	✓	✓
Signal Magnitude Area	SMA	1	✓	✓
Average Peak Amplitude	APA	3	-	✓
Tilt Angle	TA	1	✓	✓
Surrounding Segmentation Rate	SSR	1	-	✓
Inter-Axis Correlation	IAC	3	-	-
Arithmetic Mean	Mean	3	-	-
Variance	VAR	3	-	-

Table III
SET OF EXAMINED FEATURES.

the number of evaluated activities. While the resilient propagation classifier yielded a total classification accuracy of 93.43%, the backpropagation classifier achieved a very low accuracy value of 27.7%. In addition, the backpropagation classifier required extensive computational cost on the used dataset. Thus, for the comparison experiments in the context of this paper (cf. Subsection VIII-C), the resilient propagation classifier was applied to be used with the approach of [24] instead.

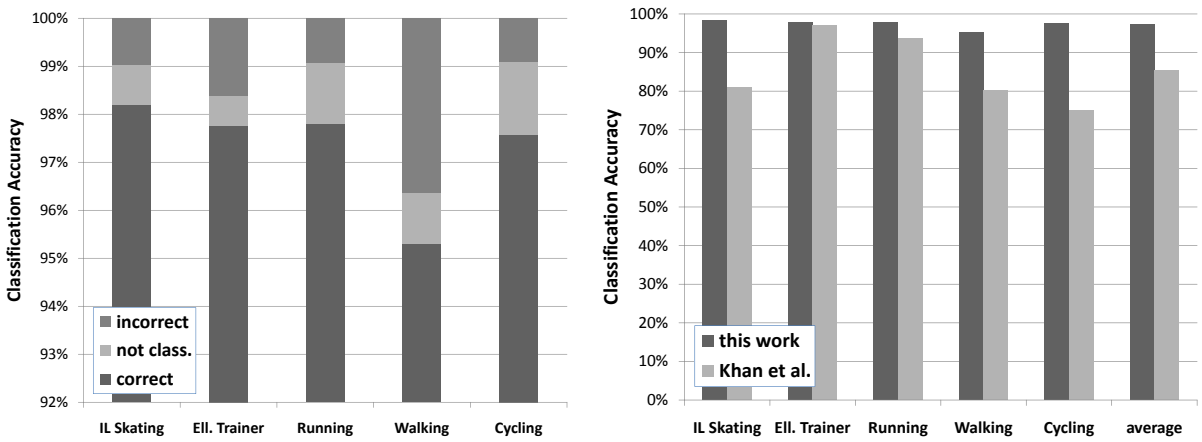
2) Feature Selection: The selection of most relevant features for the feature vector was performed using the forward-backward search technique [57] (cf. Subsection VI-A). For periodic segments, the feature selection step yielded a feature vector dimensionality of $d_v = 15$. For the nonperiodic segments, the selection process yielded the same feature vector as for periodic segments. An overview of the used features for the evaluation of the current approach and the competing approach of [24] is given in Table III. In addition to the features presented in Section V, the simple features *Arithmetic Mean*, *Variance* and *Inter-Axis Correlation*, used in [6], were included into the selection process, but proved to contain no significant information.

For creating the experimental setup as well as in the following experiments, some feature parameters were set to default values (cf. Section V). For the APA feature, values for the minimum peak threshold ($\tau_{min} = 0.7$), the peak threshold step ($\Delta_\tau = 0.02$), the minimum number of peaks ($p_{min} = 3$) and the minimum distance between peaks ($minDist = 10$) were set. For the SSR feature, the window size w_{SSR} was set to 1500, which corresponds to a temporal window of 60 seconds.

3) Further Parameters: The window size for the average filter that is applied to remove outliers is set to 3. For the segmentation (cf. Section IV), the following default values were used:

- The reference sample $\widehat{X}$ consisted of 25 observations, which corresponds to one second.
- The required minimum correlation ρ_{min} was set to 75%.
- The minimum length m_p of periodic segments was set to 100 observations, which corresponds to four seconds.
- The length m_{np} of nonperiodic subsegments was set to 80 (3.2 seconds).

In the reclassification step (cf. Section VII), the size w for the temporal window, which is used to capture the weighted occurrences of the class information of the surrounding observations, consists of 751 observations (750 plus the observation that is to reclassify), which corresponds



(a) Overview.

(b) Comparison of the results of this paper and [24].

Figure 3. Classification result.

to 15 seconds before and after the current observation, respectively. Furthermore, as five activity classes are evaluated in this paper, $cl = 5$.

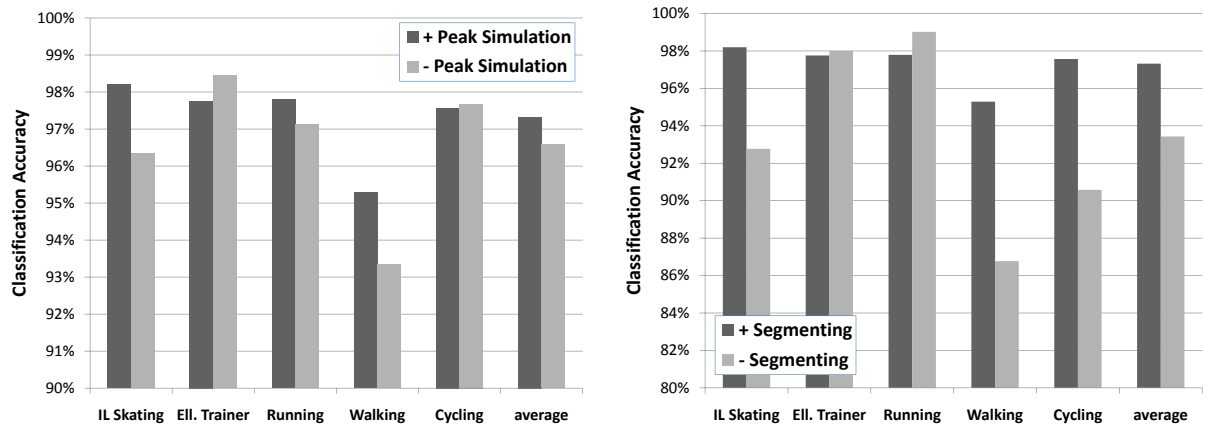
C. Classification Results

The classification algorithm was evaluated by performing a cross validation on the activity classes. For each time series segment containing one of the five activities (cf. Subsection VIII-A), the obtained classification accuracy using the default settings of Subsection VIII-B are depicted in Figure 3(a). The classification yields accuracies of more than 95% for each activity. The highest classification error was obtained with the activity *Walking*, which was classified as *Cycling* with 3.56%, which can simply be explained by the observation that these activities are likely to create similar acceleration measurements. In order to visualize the percentage of segments that were incorrectly classified or could not be classified at all, the reclassification step was omitted in the first experiment. The effect of the reclassification will be examined in Subsection VIII-D.

In [24], the classification of 15 different activities yielded an accuracy of 97.9%. For the evaluation in the context of this paper, a slight adaption of this approach was implemented: the resilient propagation algorithm was used instead of the usually applied backpropagation algorithm due to performance reasons (cf. Subsection VIII-B). Figure 3(b) illustrates the classification results of the approach introduced in this paper compared to the results of [24]. It can be observed that, for each class, the approach of [24] achieves less accuracy compared to the approach presented in this paper.

D. Effect of the Processing Steps

The next experiment will examine the effect of the preprocessing steps. Evaluating the peak reconstruction (cf. Subsection III-B), the classification results could be improved for three out



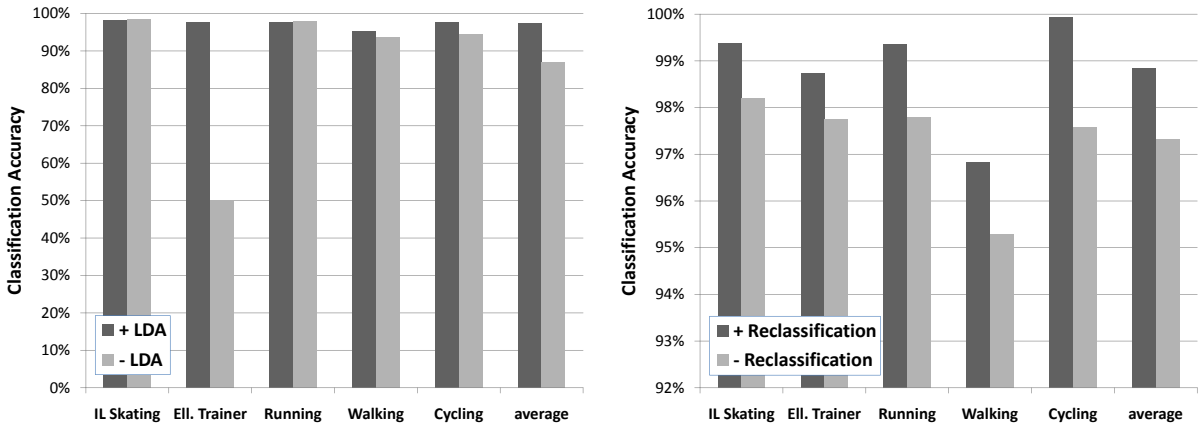
(a) Effect of the peak reconstruction.

(b) Effect of the segmentation.

Figure 4. Effect of data preprocessing.

of five activities (*Walking*, *Running* and *In-line Skating*). This can be explained by the fact that these activities take advantage from their significant peaks because of significant movements, whereas the movements of *Cycling* and *Elliptical Trainer* are indirectly supported by the sports equipment, which may lead to rather smooth movements. Overall, this step yielded an overall precision gain of almost 1% (cf. Figure 4(a)).

Next, the effect of the segmentation (cf. Section IV) was evaluated. Instead of choosing a naïve solution without a segmentation that extracts only one feature vector for the time series, the time series was divided into non-overlapping subsequences of equal length, each containing 80 observations, analogously to the size of nonperiodic segments in the original approach (cf. the experimental setup in Subsection VIII-B), and a feature vector was derived for each segment. The SSR feature could not be applied here, as, for this segmentation variant, no information about the amount of surrounding periodic segments is available. Hence, the used feature vector consisted of 14 features for this variant of the classification approach. The results are shown in Figure 4(b). It can be observed that, for the equal-length segmentation, almost no segment remains unclassified. This is due to the fact that short segments that do not contain enough information to be represented by a feature vector (cf. Section V) only occur at the end of the time series. For long-term activities that are very constant over time, such as *Running* and *Elliptical Trainer*, the equal-length segmentation yields comparable results, as there are only few gaps in the data. For activities consisting of short-term periods interrupted by several breaks due to external influence factors, e.g., in the case of *Cycling*, where pedaling is often noncontinuous, a classification supported by a segmentation into periodic and nonperiodic parts achieves a significant improvement of 4% in average. Similar observations explain the significant improvement with the activities *Walking* and *In-line Skating*, as the step length is not homogeneous.



(a) Effect of the LDA.

(b) Effect of the reclassification.

Figure 5. Effect of the LDA and the reclassification.

E. Effect of the Feature Transformation

In the next experiment, the effect of the LDA (cf. Subsection VI-B) was examined. With the most activity classes, the LDA improves the results only slightly. This shows that the combination of features done by the forward-backward search already yielded representative feature vectors with almost no redundant information. In the processing chain, the forward-backward search step is performed before the application of the LDA.

The only activity that obtains a performance gain with applying the LDA is the activity *Elliptical Trainer*. As this activity is, intuitively, very similar to both activities *Walking* and *Cycling*, an accurate separation among these classes is not always possible. Moreover, the training datasets for this activity class seem to be very inhomogeneous due to significantly different velocities. Here, the LDA maximizes the differences to the other activity classes successfully. Thus, these errors can be corrected. The results of this comparison are depicted in Figure 5(a).

F. Effect of the Reclassification

The reclassification step was omitted with the evaluations of Subsections VIII-D and VIII-E in order to get the amount of unclassified data returned. Finally, the observed results with an additional application of the reclassification step are illustrated in Figure 5(b)). Here, a slight improvement of 1.6% was achieved. Most nonperiodic segments that could not be classified in the actual classification step seem to contain many activity changes within a short time period, which leads to errors in the reclassification step.

G. Conclusions

Concluding, it can be stated that the proposed approach achieve results of high quality, since a state-of-the-art activity recognition method could be outperformed. The evaluation of

the processing step showed that each of them is able to support the actual classification step in order to provide reliable statements about the performed activity. For the case of limited resources, i.e., if the classification algorithm has to run on the firmware of the sensor in order to provide a real-time recognition and to return results quickly, a trade-off solution between storage and processing time requirements and classification accuracy has to be found.

IX. Summary

This paper provided an effective solution for the application scenario of activity recognition on periodic time series that are collected from accelerometers, which emerged from the application scenario of [45], [46]. The proposed solution, which can be categorized as matching-based approach, extends existing methods by integrating additional processing steps, such as a reconstruction of the data peaks, a utilization of suitable periodicity features and a reclassification of the data to improve the classification results. The experimental part showed an improved recognition quality in comparison with existing work.

Acknowledgements

This research has been supported in part by the THESEUS program in the MEDICO and CTC projects. They are funded by the German Federal Ministry of Economics and Technology under the grant number 01MQ07020. The responsibility for this publication lies with the authors.

References

[1] R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. pages 69–84, 1993.

[2] B. E. Ainsworth, D. R. J. Jr., and A. S. Leon. Validity and reliability of self-reported physical activity status: the lipid research clinics questionnaire. *Medicine and Science in Sports and Exercise*, 25(1):92–98, 1993.

[3] F. R. Allen, E. Ambikairajah, N. H. Lovell, and B. G. Celler. An adapted gaussian mixture model approach to accelerometry-based movement classification using time-domain features. In *Proceedings of the 28th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 3600–3603, 2006.

[4] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga. Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. *23rd International Conference on Architecture of Computing Systems (ARCS)*, pages 1–10, 2010.

[5] R. Baeza-Yates and B. B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, 1999.

[6] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. In *2nd International Conference on Pervasive Computing (PERVASIVE), Vienna, Austria*, pages 1–17, 2004.

[7] N. C. Barengo, G. Hu, T. A. Lakka, H. Pekkarinen, A. Nissinen, and J. Tuomilehto. Low physical activity as a predictor for total and cardiovascular disease mortality in middle-aged men and women in finland. *European Heart Journal*, 25(24):2204–2211, 2004.

[8] T. Bernecker, F. Graf, H.-P. Kriegel, N. Seiler, C. Türmer, and D. Dill. Knowing: A generic data analysis application. pages 630–633, 2012.

[9] A. Björck. Least squares problems. In *Encyclopedia of Optimization*, pages 1856–1866. 2009.

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

[11] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[12] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[13] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[14] I. C. for Health and S. Care. Health survey for england - 2008. *National Health Service*, 1, 2008.

[15] H. Ghasemzadeh, V. Loseu, E. Guenterberg, and R. Jafari. Sport training using body sensor networks: a statistical approach to measure wrist rotation for golf swing. In *Proceedings of the 4th International Conference on Body Area Networks (BodyNets)*, pages 2:1–2:8, 2009.

[16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. 11(1):10–18, 2009.

[17] E. A. Heinz, K. Kunze, M. Gruber, D. Bannach, and P. Lukowicz. Using wearable sensors for real-time recognition tasks in games of martial arts - an initial experiment. In *Proceedings of the 2nd IEEE Symposium on Computational Intelligence and Games (CIG), Reno/Lake Tahoe, NV*, pages 98–102, 2006.

[18] 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*, 10(1):156–167, 2006.

[19] E. J. Keogh, S. Chu, D. Hart, and M. J. Pazzani. An online algorithm for segmenting time series. In *Proc. ICDM*, pages 289–296, 2001.

[20] E. J. Keogh and M. J. Pazzani. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In *Proc. KDD*, pages 239–243, 1998.

[21] E. J. Keogh and M. J. Pazzani. Relevance feedback retrieval of time series data. In *Proc. SIGIR*, pages 183–190, 1999.

[22] E. J. Keogh and P. Smyth. A probabilistic approach to fast pattern matching in time series databases. In *Proc. KDD*, pages 24–30, 1997.

[23] A. M. Khan, Y. K. Lee, and T.-S. Kim. Accelerometer signal-based human activity recognition using augmented autoregressive model coefficients and artificial neural nets. In *Proceedings of the 30th International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, pages 5172–5175, 2008.

[24] A. M. Khan, Y.-K. Lee, S.-Y. Lee, and T.-S. Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Biomedical Engineering*, 14(5):1166–1172, 2010.

[25] A. Koski, M. Juhola, and M. Meriste. Syntactic recognition of ecg signals by attributed finite automata. *Pattern Recognition*, 28(12):1927–1940, 1995.

[26] A. Krause, M. Ihmig, E. Rankin, D. Leong, S. Gupta, D. Siewiorek, A. Smailagic, M. Deisher, and U. Sengupta. Trading off prediction accuracy and power consumption for context-aware wearable computing. In *Proceedings of the 9th International Symposium on Wearable Computers (ISWC), Osaka, Japan*, pages 20–26, 2005.

[27] A. Krause, D. Siewiorek, A. Smailagic, and J. Farringdon. Unsupervised, dynamic identification of phsiological and activity context in wearable computing. In *Proceedings of the 7th International Symposium on Wearable Computers (ISWC), White Plains, NY*, pages 88–97, 2003.

[28] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.*, 12:74–82, 2011.

[29] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan. Mining of concurrent text and time series. In *KDD-2000 Workshop on Text Mining*, pages 37–44, 2000.

[30] C.-S. Li, P. S. Yu, and V. Castelli. MALM: a framework for mining sequence database at multiple abstraction levels. In *Proc. CIKM*, pages 267–272, 1998.

[31] J. E. Manson, M. J. Stampfer, G. A. Colditz, W. C. Willett, B. Rosner, C. H. Hennekens, F. E. Speizer, E. B. Rimm, and A. S. Krolewski. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *The Lancet*, 338(8770):774–778, 1991.

[32] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN), Cambridge, MA*, pages 116–121, 2006.

[33] J. Mäntyjärvi, J. Himberg, and T. Seppänen. Recognizing human motion with multiple acceleration sensors. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC), Tucson, AZ*, 2001.

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

[35] K. Pearson. On lines and planes of closest fit to systems of points in space. 2(6):559–572, 1901.

[36] S. Pirttikangas, K. Fujinami, and T. Nakajima. Feature selection and activity recognition from wearable sensors. In *Proceedings of the 3rd International Symposium on Ubiquitous Computing Systems (UCS),Seoul, Korea*, pages 516–527, 2006.

[37] J. C. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods – Support Vector Learning*, pages 185–208. 1998.

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

[39] Y. Qu, C. Wang, and X. S. Wang. Supporting fast search in time series for movement patterns in multiple scales. In *Proc. CIKM*, pages 251–258, 1998.

[40] M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proceedings of the 1st IEEE International Conference on Neural Networks (ICNN), San Francisco, CA*, pages 586–591, 1993.

[41] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Neurocomputing: foundations of research. pages 673–695. 1988.

[42] H. Shatkay and S. B. Zdonik. Approximate queries and representations for large data sequences. In *Proc. ICDE*, pages 536–545, 1996.

[43] A. Sugimoto, Y. Hara, T. W. Findley, and K. Yoncmoto. A useful method for measuring daily physical activity by a three-direction monitor. *Scandinavian Journal of Rehabilitation Medicine*, 29(1):37–42, 1997.

[44] C. Türmer. Konzeptionierung eines Aktivitätsmonitoring-Systems für medizinische Applikationen mit dem 3D-Accelerometer der Sendsor GmbH. Master's thesis, Technische Universität München, Germany, 2009.

[45] C. Türmer, D. Dill, A. Scholz, M. Gül, T. Bernecker, F. Graf, H.-P. Kriegel, and B. Wolf. Concept of a medical activity monitoring system improving the dialog between doctors and patients concerning preventions, diagnostics and therapies. 2010.

[46] C. Türmer, D. Dill, A. Scholz, M. Gül, A. Stautner, T. Bernecker, F. Graf, and B. Wolf. Conceptual design for an activity monitoring system concerning medical applications using triaxial accelerometry. 2010.

[47] H. Vullings, M. Verhaegen, and H. Verbruggen. ECG segmentation using time-warping. 1280:275–285, 1997.

[48] C. Wang and X. S. Wang. Supporting content-based searches on time series via approximation. pages 69–81, 2000.

[49] S. Wang, J. Yang, N. Chen, X. Chen, and Q. Zhang. Human activity recognition with user-free accelerometers in the sensor networks. In *Proceedings of the IEEE International Conference on Neural Networks and Brain (ICNN&B), Beijing, China*, pages 1212 –1217, 2005.

[50] D. E. R. Warburton, C. W. Nicol, and S. S. D. Bredin. Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, 174(6):801–809, 2006.

[51] J. A. Ward, P. Lukowicz, and G. Tröster. Gesture spotting using wrist worn microphone and 3-axis accelerometer. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies (sOc-EUSAI), Grenoble, France*, pages 99–104, 2005.

[52] R. A. Washburn and H. J. Montoye. The assessment of physical activity by questionnaire. *American Journal of Epidemiology*, 123(4):563–576, 1986.

[53] P. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, Cambridge, MA, 1974.

[54] A. K. Yancey, C. M. Wold, W. J. McCarthy, M. D. Weber, B. Lee, P. A. Simon, and J. E. Fielding. Physical inactivity and overweight among los angeles county adults. *American Journal of Preventive Medicine*, 27(2):146–152, 2004.

[55] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern Recognition Letters*, 29:2213–2220, 2008.

[56] H. Yoon, K. Yang, and C. Shahabi. Feature subset selection and feature ranking for multivariate time series. *Knowledge and Data Engineering, IEEE Transactions on*, 17(9):1186–1198, 2005.

[57] T. Zhang. Adaptive forward-backward greedy algorithm for sparse learning with linear models. pages 1921–1928, 2008.