Detecting Leisure Activities with Dense Motif Discovery

Eugen Berlin and Kristof Van Laerhoven
Department of Computer Science
Technische Universität Darmstadt
{berlin,laerhoven}@ess.tu-darmstadt.de

ABSTRACT
This paper proposes an activity inference system that has been designed for deployment in mood disorder research, which aims at accurately and efficiently recognizing selected leisure activities in week-long continuous data. The approach to achieve this relies on an unobtrusive and wrist-worn data logger, in combination with a custom data mining tool that performs early data abstraction and dense motif discovery to collect evidence for activities. After presenting the system design, a feasibility study on weeks of continuous inertial data from 6 participants investigates both accuracy and execution speed of each of the abstraction and detection steps. Results show that our method is able to detect target activities in a large data set with a comparable precision and recall to more conventional approaches, in approximately the time it takes to download and visualize the logs from the sensor.

Author Keywords
activity detection, motif discovery, psychiatric monitoring

ACM Classification Keywords
H.5.m Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Algorithms, Design, Experimentation, Measurement.

INTRODUCTION
The automated recognition of the user's activities has been suggested for over a decade as an attractive system feature in pervasive computing literature. By relying on observations from sensors that are deployed in the environment of the user, or worn on his or her body, knowledge of recognition activities can be extracted. This technology is motivated by establishing a more effective dialogue between user and computer, reducing cognitive load in pervasive computing scenarios, or delivering an improved service by proactively responding to given situations. Numerous applications have been suggested to benefit from activity recognition: The authors of [20] for instance demonstrate how Activities of Daily Living (ADLs) can be detected, which are used in estimating the quality of self-care for elderly users. Further application scenarios for activity recognition include detecting office activities [17], maintenance tasks performed by engineers [21] and specific sports activities [8], finding appropriate advertising based on the user's physical activity [18] and eating and drinking activities [1]. Depending on the application, algorithms can go beyond recognition of activities and detect certain characteristics, such as the number of counts for selected gym workouts [3].

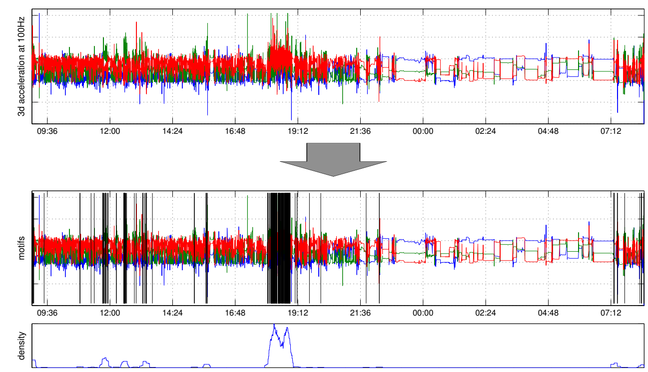

Figure 1. Day-and-night recordings from an unobtrusive sensor (top) are in this paper's approach analyzed for certain leisure activities. Salience detection upon motif discovery (bottom) is used to find typical activity patterns (black marks) as supporting evidence for the activity.

This paper's activity recognition approach is motivated by an application scenario that is relatively new: Psychiatric patient monitoring aims at characterizing both mood and behavioral trends by recording activity data over a period of typically *several months*. Current commercial solutions[1] are able to detect sleep and wake cycles for such long deployments, and come with tools for facilitating the recording of certain physical activities. In this scenario, a few of the general problems in activity recognition become trivial to solve: Patients already detail their activities in diaries so supervised learning methods can be employed, and only a few key leisure activities are of interest among the logged data. Other requirements, however, form new challenges: Sensors need to record for long stretches of time, the large amount of logged data needs to be analyzed fast enough, and detection needs to be robust against a deluge of background data.

[1] e.g., the Actiwatch: `http://www.camntech.com/` [3/2012]

http://doi.acm.org/10.1145/2370216.2370257

Figure 1 illustrates the dense motif discovery method that forms the basis of our detection approach. Occurrences of so-called *motifs* are searched in the data, which are then used to substantiate the presence of an activity within the data. Motifs are discovered at training time by abstracting the raw acceleration samples in function of sequences of peak patterns, and efficiently searching for such similar patterns by means of a data structure called suffix tree. This search can be implemented in linear time, and performs a strong abstraction in comparison to traditional acceleration features (such as mean, variance, or FFT coefficients), at the cost of a processing overhead. Classification is then implemented with a straightforward bag-of-words classifier. An extra advantage of this method is that the illustration of motif occurrences in the time series allows for visual inspection of the activity recognition before the classification step. The main contributions of this work are threefold:

- We use a specific mood monitoring scenario as an activity recognition application with interesting constraints that impact both sensor and the software system design.
- A deployable system has been built consisting of a minimally invasive, wrist-worn sensor that is able to last for two weeks on a single battery charge while recording acceleration samples at 100Hz, and a data analysis tool that can efficiently process the recorded data.
- A novel detection approach is suggested that is suitable for classifying large amounts of long-term activity data, relying on local shape features within the acceleration signal and dense motif discovery.

The remainder of this paper is structured as follows: First, a psychiatry monitoring scenario motivates the need for the proposed fast and accurate detection of when a user performs physical leisure activities. Then a section is dedicated to related work in activity recognition, with particular focus on methods that aim for long-term deployments, and other motif discovery research in particular. The next section will then go into details on the different design choices and steps that constitute our method, such as the linear abstraction of inertial data and the use of suffix trees for finding motifs. An experiment then presents results on several long-term datasets how fast and robustly the chosen activities can be recognized among a large amount of daily activity data. The paper is wrapped up with the conclusions section enumerating the key findings of this paper, as well as the future research potential.

MOTIVATION: PSYCHIATRIC MONITORING SCENARIOS

Research in mood disorders relies frequently on the patients' self-reports, as well as semi-structured interviews with a psychiatrist, both during diagnosis and therapy. Emerging work with actigraphy tools and activity measurement in psychiatry [28, 24] has started to deploy wrist-worn sensors in conjunction with these tools that are recording the activity intensities observed for the patient over intervals of several seconds to minutes at a time. Such studies have found to be valuable in a range of mood disorder studies, such as attention deficit hyperactivity disorder (ADHD) and bipolar disorder [6].

Characterized by severe mood swings between manic or hypomanic, mixed, as well as depressive episodes, it is important in the diagnosis of a bipolar disorder to observe the patient's activity behavior over multiple weeks to months at a time. For mania for instance, energy levels tend to be high and activities tend to be performed in an interleaved fashion or especially vigorously (e.g., performing sports exercises longer without breaks). Similarly, depression tends to show in lower activity levels or the way patients behave during key activities, from not performing them at all or sparsely, to not fully completing them. Apart from daily activities such as sleep and food intake, especially physical and leisure activities are very likely to be impacted: Patients might for instance stop playing tennis when depressed, or vigorously practice for several hours in a manic episode.

In collaboration with psychiatrists specializing on actigraphy and ambulatory assessment of bipolar disorders, several interviews were held to list the basic requirements and expectations that an activity recognition method should adhere to. These were grouped in three categories that are important factors to consider when designing an activity recognition system for this field:

- **Supervised learning.** Patients are normally interviewed at regular intervals of several weeks, and provide in many cases log entries where they report on performed tasks and their mood assessment. Recent actigraphy systems already combine this information with the sensor data, so that in a similarly-developed activity recognition approach this can be used as an approximate annotation to train a patient-specific classifier.
- **Week-long, 24/7 data.** It is crucial that data is continuously captured at all hours of the day, as patients that go through depression or manic episodes are known to perform activities at irregular times, including night time. As a result, the sensor units need to be robust and power-efficient enough to keep recording for such long timespans without breaks, and the amount of data that need to be recorded will be substantial to process.
- **Leisure activities.** The number of activity classes that need to be recognized is relatively small (often 1) and can be determined by the psychiatrist during the first phases of diagnosis. This makes it easier for patients to keep track of what activities were performed, and this also impacts activity recognition, since only few activities need to be detected amongst a large amount of background data that might produce false positives.

This paper focuses first and foremost on a practical capturing and detection method that is able to recognize particular activities, and this within large datasets that tend to include a massive amount of background data, generally holding weeks of activity data at a time. The next section will review some of the literature on activity recognition methods that tackle similar problems, as well as technically related approaches in wearable sensing, data mining, and classification, and situate the proposed approach among peer research.

RELATED WORK

Activity recognition has been suggested as a promising tool before for bipolar studies. Both [23] and [27] have pointed out that the use of automatically monitoring activities would be a useful tool to support the diagnosis of bipolar disorder and detect onsets of depression and mania. In particular the so called Hamilton Depression Scale (HAMD) and Bech-Rafaelsen Mania scale (BRMS) [2] tools contain elements where physical activities are of considerable interest. However, to our knowledge no research has thus far focused on implementing an activity recording method that can practically be deployed on a patient's wrist for a week and allows almost-instant analysis at the psychiatrist's office.

A significant amount of work in the context of activity recognition has focused on automatic feature selection for inertial data and using strong classifiers upon these features to detect activities. Common candidates that have proven worthwhile in previous studies (e.g., [9], [13]) have found basic statistics, in particular mean and variance, and frequency-based features (FFT and Cepstral Coefficients, spectral entropy and energy) over a sliding window to be distinctive features to characterize. Lester et al. [13] use in a combined discriminative-generative classification approach the AdaBoost algorithm to automatically select the best of these features and to learn an ensemble of static classifiers to recognize different activities. Strong classifiers that have proved valuable in activity recognition include Naïve Bayes, Bayesian Networks, Hidden Markov Models (HMMs) or Support Vector Machines (SVMs) [1, 3, 8, 13, 15, 17, 18, 19, 20, 21].

The use of motif discovery has been suggested as an alternative approach in activity recognition that is especially useful when a fully supervised method is not feasible, or when short characteristic gestures need to be spotted that are hard to annotate individually by the system's users. Minnen et al. [16] use motifs to automatically discover gym work-out gestures in inertial data recorded form body-worn sensors, by mapping the sensor data to symbols and using a suffix tree to search efficiently through the resulting large symbolic strings. Similarly, Hamid et al. [7] analyze activities in an instrumented kitchen, and [26] uses motif discovery to detect activities such as walking and falling without supervision. We use motif discovery primarily because (1) the annotations that describe which activity was done when are provided by the system's wearer using self-recall and are thus only approximate, (2) we assume that physical leisure activities can typically be characterized by occurrences of certain short gestures, and (3) because it is an especially fast method that allows parsing of large data sets at once.

Motif discovery techniques generally rely on symbolic abstraction of the original raw sensor data to obtain an especially fast detection method using the suffix tree representation. In [22], the symbolic representation of inertial data is used to facilitate efficient matching of motion patterns. The inertial trajectory in space is approximated after which it is mapped to a character based on the minimum angular distance to the 3 axes that are represented by a small alphabet of 6 symbols, thus resulting in a motion string. The

Figure 2. Our custom-made sensor platform, designed to be worn and used as a wrist watch, is light-weight enough to be worn for recordings of up to 2 weeks. Left: board with controller, accelerometer, micro-SD storage, and USB connector for data access and charging the battery.

approach of [16] uses discrete mapping based on a Gaussian distribution fit on the data, whereas others use probabilistic approaches such as [4], or approximate the sensor signal's time-series first by piecewise constant segments of fixed length [14], which are then mapped to a set of discrete symbols. Our approach uses similarly an approximation of the inertial time-series, but uses for the mapping the subsequent segments' slopes to capture the essence of these short gestures' patterns in accelerometer data from the wrist.

The importance of long-term recording of inertial data, in an unobtrusive manner, has been stressed in several key publications on activity recognition (most notably [5, 10]). Although data sets have been recorded over similar time frames as in this paper, none so far have recorded day and night for several days consecutively. Actigraphy on the other hand does log for extended periods of time, but abstracts the inertial data on-board the sensor and does not retain the original time series at the resolution of this paper's (100Hz).

DATA LOGGING PROTOTYPE

Deploying sensors that are lightweight and wearable is one of the hard challenges in the creation of robust recognition systems, as has been identified in previous research such as the Mobile Sensing Platform [5]. Since there are no current off-the-shelf platforms that allow continuous logging of high-resolution accelerometer data, the experiments in this paper have been recorded from custom-built sensor units that measure and store regularly-timestamped 3D acceleration on 2 Gigabytes of local flash storage (Figure 2). This persistent memory is required since the recording is done by sampling 3D acceleration data at 100Hz. In addition to the accelerometer, light and temperature sensors are also on board.

While recording, the sensor regularly acquires 33 equidistant (10 ms) samples at a time from an ADXL345 accelerometer's first-in first-out (FIFO) buffer and logs these using run-length encoding (RLE) on the on-board micro-SD card. Regular time stamps are produced by a precise real-time clock (RTC) unit embedded in the PIC18F46j50 microcontroller, allowing detailed verification of separate annotations taken by the subjects and individual sensor readings. Between FIFO reads, only the accelerometer is active on the board, with the micro controller in sleep mode preserving the charge of its miniature $180mA$ Li-Polymer battery.

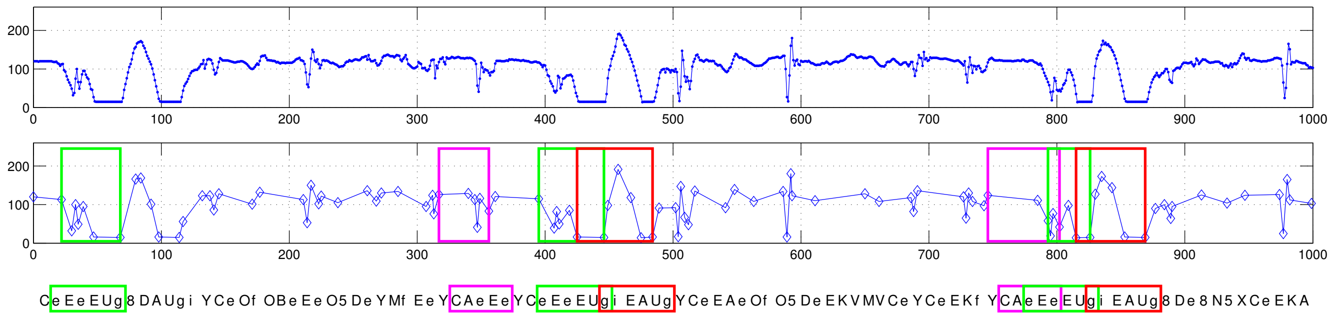

Figure 3. The raw 3D 100 Hz inertial data (top plot) are transformed by a piecewise linear approximation algorithm into segments (bottom plot) that preserve the shape of the signal to facilitate storage and analysis. The segments are subsequently abstracted in discrete symbols (bottom row) to allow fast discovery and matching of motifs. Occurrences for three motifs are highlighted by colored boxes; Note that they can overlap and vary in length.

The entire sensor draws $480uA$ on average while logging. Non-stop logging of 100Hz inertial data should thus last for 15.6 days. From our own experiments, we have observed that the battery is usually drained after 14.5 days, which still allows 2-weeks of uninterrupted logging. To meet the requirements of long-term 24/7 deployment, the unit is packed in a custom shock-proof case and provided with an anti-allergic textile wrist strap. For more inconspicuous deployments, as advised by our collaborators from psychiatry, an OLED display has been added to display the current time from the real-time clock. The display is by default turned off and activated by double-tapping the watch, driven by an on-accelerometer function interrupt. With the OLED attached, deployments generally last one to two days less.

DENSE MOTIF DISCOVERY

This section gives an overview of the search and selection procedure for motifs from raw inertial data, and motivates the use of *dense* motifs. A set of early abstraction steps of the accelerometer data, together with a search-optimized data structure called suffix tree, guarantee that searching through weeks of data becomes feasible on standard computing hardware, and that classification can be done almost simultaneously with the downloading of the prototype's data.

Method Overview

Motif discovery refers to the search for recurring sequences or patterns within a data stream. For this to be applicable in real-world scenarios, previous research has identified several techniques to represent the original data, which often tend to be noisy and hard to match exactly, in a discrete symbolic string. This paper's approach implements a discrete mapping that applies a two-step abstraction process while aiming to characterize patterns in the data (i.e., potential activity-specific gestures) by the shape of the time series.

Figure 3 illustrates how the proposed method transforms inertial data to a string that facilitates the finding of recurring motifs: The original data consists of 3D accelerometer samples taken in equidistant 10 ms steps. Sets of linear segments are created from these, using an online approximation algorithm that minimizes the residual error between original data and segments. The segments are discretized into symbols using the slopes of connected linear segments. Using a suffix tree representation of the target activity's training data, a set of motifs is found using adaptive length thresholds. Motifs that also have occurrences in the training's background data, i.e., the vast set of data that does not belong to the activity, are withheld. As a *dense* set of motifs is trained for, classification is done by searching new data for time windows in which motifs from one particular activity are frequently occurring. This is implemented with a bag-of-words classifier which uses the detected motif occurrences as evidence. To perform the described motif discovery efficiently, the original sensor data needs to be abstracted: The next section will discuss a linear segmentation step.

Approximation from Raw Data

The first abstraction step is crucial from a practical and efficiency point of view: Raw sensor data is sampled at a relatively high frequency (100 Hz) to capture the essence of the gestures and typical motions performed by the sensor's wearer, but this also means that analysis of larger data sets quickly becomes challenging. Even with fast and lossless compression techniques such as run-length encoding, an entire day worth of data typically contains millions of 3D acceleration samples.

We argue that for motif discovery in inertial data, primarily the shape of the acceleration time series is important to retain. The applied technique to reduce the amount of data on the one hand and to preserve the shape of the signal on the other, belongs to the Piecewise Linear Approximation (PLA) family of abstraction algorithms. We used in our algorithm a modification of the original Sliding Window and Bottom-Up (SWAB) algorithm [11] that has been verified to perform well on body-worn accelerometer data [12].

The transformation from raw data to linear segments consists of two steps, the approximation of data on a sliding buffer window and filling the buffer with new sensor samples. The main approximation step is carried out by a Bottom-Up brute-force algorithm that produces the linear segments by merging cheapest adjacent segments until a preset threshold is reached. The leftmost segment is output as a result, and the buffer is filled with new data, whereby the modified version

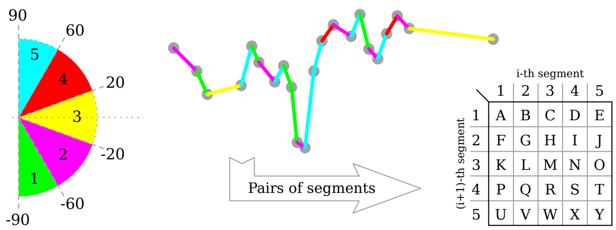

Figure 4. Mapping from linear segments to symbols: The slope range is divided into bins for a given number of separation points which are computed based on the training data segments' slopes histogram. The segments' corresponding bin numbers are then used as indices for the symbols matrix. Sliding through the segmented time series while considering two neighboring segments will thus result in a symbolic string.

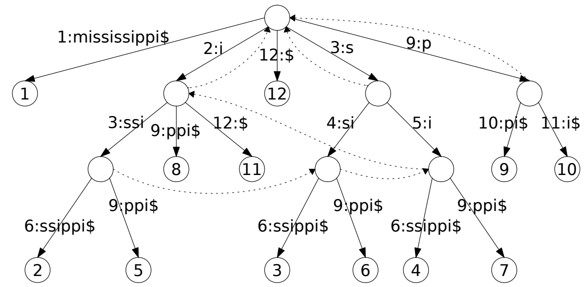

Figure 5. Generalized suffix tree for the string `mississippi` created by adding a unique terminator character $ to the original string. Suffix links are indicated with dotted edges, edge labels give the first occurrence position in the string of subsequent suffixes.

considers slope sign changes, and thus peaks in the signal, in incoming raw data. More details and a study on its efficiency can be found in [11]. The two plots in Figure 3 show a 10-second accelerometer time series and the resulting piecewise linear approximation produced by the modified SWAB algorithm. The approximation was performed per acceleration axis for implementation reasons, although it is also possible to approximate multidimensional time series.

Mapping to Discrete Symbols

After abstracting the raw acceleration data to linear segments, a discretization step is used to obtain a symbolic string representation of the original time series. This abstraction step is first and foremost required to enable fast discovery of motifs, but also for finding matches between motifs.

First, we evaluated two degrees of freedom per segment, considering the length and the slope, and mapping the resulting segments onto symbols in a similar way as was done with the SAX approach by Lin et al. in [14]. Our approach also discretizes the feature value space based on the distribution of the values. The main difference to SAX is that the first abstraction step produces constant segments of fixed length, thus having only one degree of freedom, while SWAB produces linear segments with individual slope and length. With this approach, our initial test showed that very long segments became over-represented in the motif discovery. This is due to inherent properties of accelerometer data, with long segments with a slope close to zero being over-represented, particularly during the night time and sedentary tasks, where little or no changes are present in the signal.

Being interested in mainly short and characteristic gestures, focus went to the slopes of two neighboring segments, whereby we use the angular representation of the slope defined as $\theta = \arctan(m)$. To achieve discretization, the slope range from -90 to 90 degrees was divided into bins, whereby the borders (quantiles) are selected on the basis of representative data in a histogram during the training phase. To avoid over-representation of non-motion motifs, segments with a slope close or equal to zero were not considered. The rest, where we do not assume Gaussian distribution, is used to compute the quantiles for a given number of bins (which was found to produce the best results when set to 5).

Mapping the linear segments to discrete symbols is realized by sliding through the time series, considering the slopes of two neighboring segments at a time, and converting them to one character (cf. Figure 4) using a 2 dimensional matrix. Converting an approximated time series using this approach will result in a long symbolic string, as shown in Figure 3, that can be parsed for motifs with the help of suffix trees. The advantages of this approach are two-fold: First, the length of a linear segment is not constrained to a fixed value, as it is the case with the SAX approach, and common errors where symbols afterwards would need to be merged are avoided. Secondly, by not taking the length of a segment into account when mapping the segments to symbols, more importance is placed on patterns in the data where strong peak sequences occur. With a symbolic representation of the time series now completed, the next section will discuss the method for the finding of motifs.

Extracting Motifs by means of Suffix Trees

Having mapped the raw acceleration data to a symbol sequence, motif discovery can now be done by finding substrings that occur multiple times in the target class. This is above all an efficiency problem: searching for all occurrences of every motif in a long string in an exhaustive fashion will result in a slow discovery process that is not scalable, as large sets of motifs are expected to be present.

To significantly speed up this search procedure of motifs, a technique is applied that transforms the string from a long array of symbols to a tree representation. This data structure, called *suffix tree*, generally requires more storage space than the string array, but in return allows searching for all substrings up to a certain length in linear time. Furthermore, suffix trees can be constructed in linear time using an algorithm by Ukkonen [25]. The motifs are found by checking the number of leaves for all suffixes up to a certain depth in the tree, which then corresponds to the number of the substring's occurrences in the data.

Figure 5 depicts the generalized suffix tree for the string `mississippi`. The generalized suffix tree is produced by adding a unique terminating character (such as commonly used $ or #) to the original string. With a generalized suffix tree created, this structure can be used for a multitude of different applications. The most common application is searching for query substring occurrences, for example, those of the substring `issi` in the example above: First, verifying whether the query is present in the original string at all can be answered by traversing the edges `2:i` and `3:ssi` of the tree from its root. The fact that this path can be traversed, means that the query is present in the original string. In this case, the places and number of occurrences are found by counting the leaf nodes in the sub-tree and looking at the leaf node indices: 2 occurrences with positions 2 and 5 are found after traversing the edges `6:ssippi$` and `9:ppi$`.

Suffix trees are used for the discovery of motifs that are likely descriptors for a target activity class. Motifs are found by searching the suffix tree to a certain depth, and accumulating those motifs that occur at least two times and have a minimum length (a trade-off evaluation on our data set showed a minimum length of 5 to still produce sufficient motifs for the bag-of-words classifier). The most discriminant motifs are then selected for classification, from all discovered ones by removing those that appear frequently in the background data provided during the training. After thus finding a set of motifs that tend to represent a particular activity class, these can be used as weak detectors in classification by evaluating the density of their occurrences.

Bag-of-Words Classification

Using the most discriminant motifs during a training phase as described in the previous paragraphs, classification is performed by local evidence of all motifs that support an activity. This straightforward bag-of-words classifier uses a sliding time window over the time series and accumulates local evidence by counting occurrences of motifs. As the activities tend to last at least 30 minutes and up to an hour and a half, a window size of 10 minutes was chosen.

EVALUATION

The entire approach as described in the previous section is tested in this section under conditions from the motivation scenario of psychiatric monitoring. After presenting the wrist-worn sensor prototype, as well as the test subjects and the chosen activities, a comparison of the proposed approach is given with two common activity recognition techniques that have been chosen as a benchmark. Finally, we are discussing the performance results of our and the other methods.

Participants and Target Activities

The data used in the following experiment comes from a group of volunteers who have no known psychiatric disorders and for whom a leisure activity was known before the recording phase (a key leisure activity, regularly performed as it would be chosen by a psychiatrist), which they would do once each day, for a full working week. For most, this turned out to be a leisure activity, for some a daily activity that was part of their regular schedule. Table 1 gives an overview of all participants, specifying their gender, age and their personally chosen target activity which will be used for testing detection accuracy. Additionally, the amount of raw sensor data as well as the total number of segments used in the evaluation are given. Here, the first data reduction step (modified

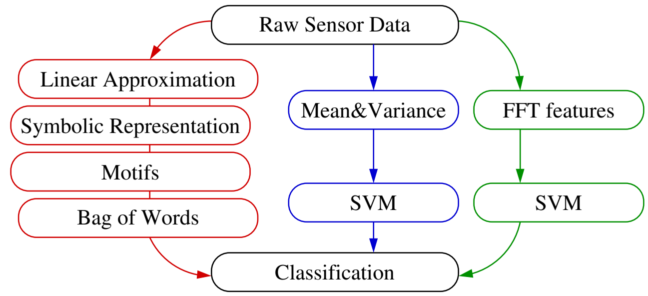

Figure 6. Overview of the detection evaluation: The dense motif classifier (red) is compared with two strong classifiers that rely on mean and variance features (blue) and FFT-derived features (green) respectively.

Table 1. The list of participants, specifying their gender, age, and the leisure activity they performed once a day, along with the number of recorded data samples and the number of linear segments produced from raw data by the segmentation process.

subject	gender	age	target activity	3D samples	segments
1	female	30	zumba	26 927 159	2 011 826
2	male	35	cycling	42 841 897	2 259 414
3	male	30	badminton	44 244 417	2 758 001
4	female	27	guitar	43 230 164	2 825 311
5	male	28	gym	34 822 499	2 480 707
6	female	26	flamenco	43 101 537	2 980 562

SWAB algorithm executed with approximation threshold of 10 and buffer size of 80) is also shown to have a significant effect, resulting in more than thirteen times less data points.

The data set from each participant was split into separate blocks of about a full day (24 hours ±50 minutes) each to facilitate 5-fold cross validation. Each activity instance lasted approximately one hour. The target activity thus holds ±5% of the entire fold, with the rest being other daily activities.

Benchmarking the Performance

To evaluate the classification performance of our approach, a comparison to two standard activity recognition techniques was done. For the latter, several classifiers were identified, with the Support Vector Machine (SVM) as the best performing, as well as different feature sets to abstract the raw data. Due to their coverage in the activity recognition community, e.g. in [9] or [29], mean and variance were identified as one combination. An additional set of features based on Fast Fourier Transform (FFT) coefficients were chosen as another: the 16 FFT features that have been suggested and evaluated in [29] consist of the absolute, real valued FFT coefficients grouped into 4 logarithmic bands, 10 Cepstral Coefficients, the spectral entropy and energy of the signal.

One imbalance in this comparison is illustrated in Figure 6: Since the dense motif approach aims at extracting characteristic motion patterns for target activities from the symbolic representation of the original sensor data, more resources are spent on pre-processing the sensor data, and less on the classification. Although Figure 6 details just the required steps, and not their time complexity, it is clear that the approaches differ significantly in how the processing steps are weighted.

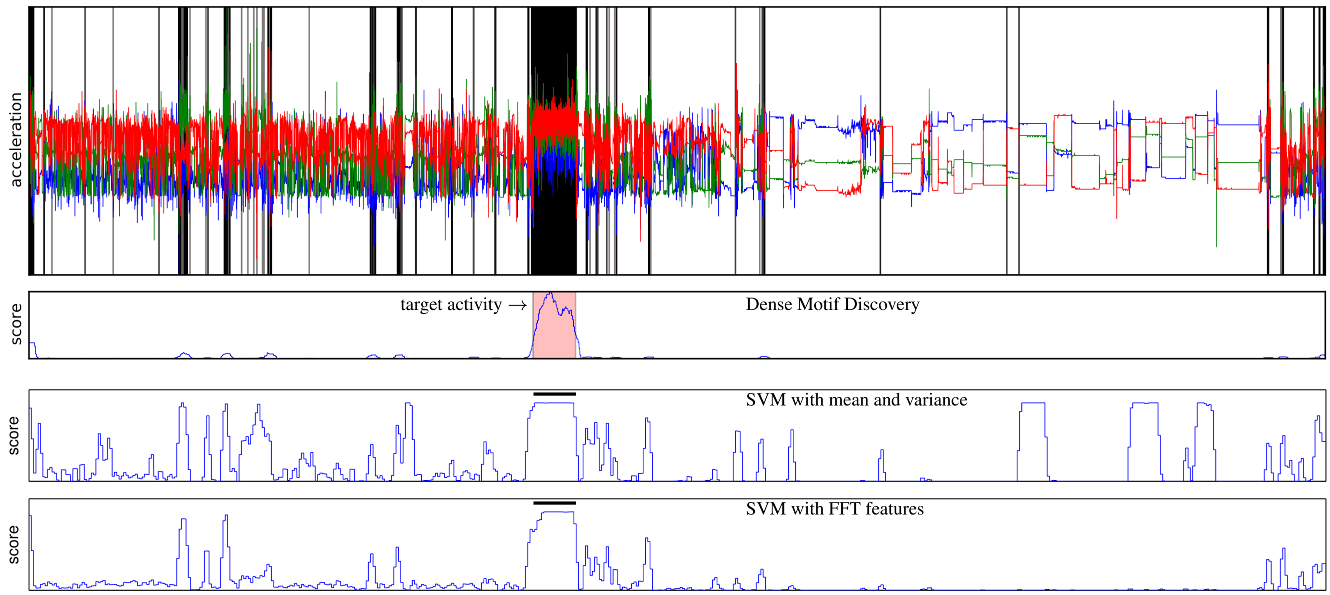

Figure 8. One day of the experiment data in which one of the participants cycled for about one hour. The topmost plot shows the original 3D sensor data, along with motif occurrences highlighted by black markers. The three plots below give the corresponding score plots produced by the different classification approaches during the evaluation: The first plot shows aggregated motif occurrences, while the two plots below show the smoothened SVM classification for mean and variance and FFT-based features respectively, with all three approaches using the same sliding window length. After combining all such results for all participants' data, the precision and recall figures show overall performance of the three approaches in Figure 9.

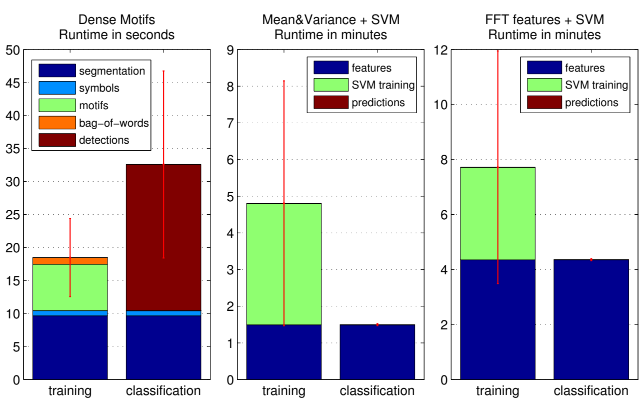

Figure 7. Mean execution times for training and detecting on 24 hours with the three approaches, with upper and lower quartiles (red lines): The dense motif method is especially faster in training, with segmentation and discovery of motif occurrences taking up most of the time. For the SVM-based approaches, most of the time is spent on calculating the features on the sliding window, with the classification done in a few seconds. Used parameters are the same as in the classification analysis.

Figure 7 shows the average times gathered during the 5-fold cross validations with our dense motifs approach for the best-performing set of parameters (as also shown in later evaluation plots, approximation threshold and buffer size: 10/80; symbols mapping to 5 bins). For one day worth of data the two abstraction steps (producing segments and converting them to symbols) require about 10 seconds. Depending on the activity, the time required for extracting the characteristic motifs from the training part of the dataset ranges from 3 up to 12 seconds. Obtaining motif occurrences for the classification on the fifth part of the dataset and computing the score needs from 18 up to 46 seconds, using a standard laptop setup and with the source code written in Python.

Both mean and variance, as well as the FFT-based features, are computed on a sliding window over the raw data, with window sizes varying from 1 to 30 seconds. For classification, the `svmtrain` and `svmclassify` methods from the Matlab Bioinformatics Toolbox were used. The performance of the features with the SVM classifier was evaluated by the same 5-fold cross validation as for the dense motifs approach. The detections produced by the SVM classifier are smoothened by a sliding window of 10 minutes to filter out outlier false detections, resulting in a score. At this stage, by evaluating the obtained classification versus the the ground truth annotations, precision and recall are computed for our dense motifs as well as for the features with SVM approaches. Figure 8 shows an example illustrating how the different classification techniques performed on the third day of the cycling dataset during the evaluation phase. The score plots below the raw data show the aggregated motif occurrences for the dense motif method, and the normalized results of the windowed filter after SVM.

Experiment Results and Discussion

This section presents the experiment results for the leave-one-day-out 5-fold cross validations: For every activity one day is left out for testing purposes, while training (obtaining the motifs that tend to represent the activity) on the other four days. Since the evaluation considered a wide range of possible parameter combinations (abstraction thresholds, buffer lengths, window sizes, etc.), only a few prolific figures are shown to discuss the experiment results.

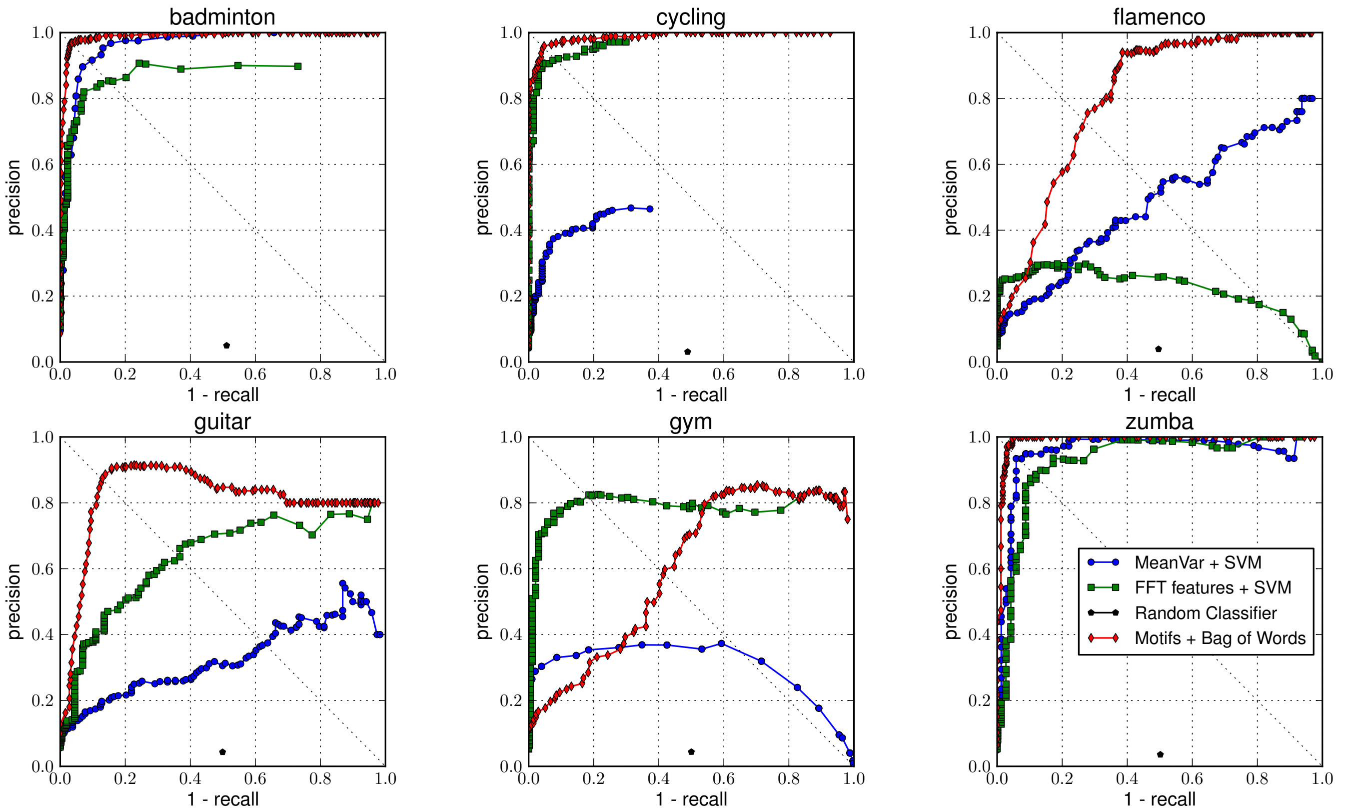

Figure 9. Precision and recall performance results on the six different activities obtained through the leave-one-day-out 5-fold cross validation, averaged over the number of folds. The dense motif approach outperforms the SVM classifier trained with mean & variance or FFT features on five out of six activities, while performing significantly worse to the FFT features trained SVM classifier on the gym dataset (fifth plot from the top-left).

Figure 9 shows a comparison of the best performing average precision and recall figures of our approach and the SVM classifier that has been trained with mean and variance, and the FFT features. Additionally, the performance of a random 'guessing' classifier is depicted for completeness. Precision and recall are averaged over the number of folds, while for each activity and classification method the choice of parameters with the best classification performance is chosen.

The SVM classifier trained with mean and variance features performs well on activities that involve a lot more motion, with especially the variance of the signal playing a significant role, as can be seen by comparing the activities badminton or zumba with gym, cycling or flamenco. While the first two activities exhibit very high accelerations due to sharp hand motions, the three latter activities lack such high accelerations. The dense motif approach is in many cases the best-performing, in some cases even significantly. To illustrate its strengths, the performance on the badminton data is shown in Figure 10, zooming in on a short time span of 50 seconds with motifs occurrences matching the underlying characteristic motion patterns. Motifs here often overlap, with their dense occurrences making the detection of the activity more reliable.

When investigating the impact of the abstraction, we noticed that the parameters that control the first abstraction step for

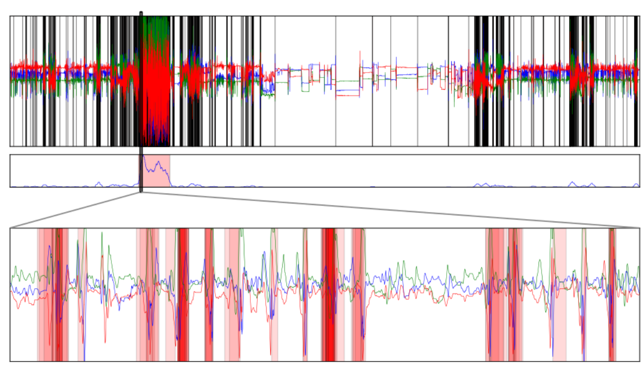

Figure 10. Dense motifs performance on one fold of the badminton data over one day (upper plot, with badminton activity marked in red), and a 50-seconds fraction thereof with motif occurrences (lower plot). Characteristic motions such as forehand, backhand, smashes, are often marked by motifs, while areas in between tend to be left out.

the badminton and zumba data have almost no impact on the classification performance. This can be explained by the importance of high acceleration peaks in the signal that are preserved even with a coarse grained approximation. The approximation threshold plays a more important role for the flamenco or cycling datasets, or generally for activities without extreme accelerations where coarse grained approximation results in a worsened classification. With the best per-

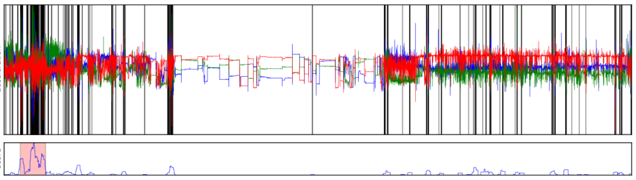

Figure 11. Example of dense motifs on a day with the flamenco activity (highlighted in red). Note the drops in the score for the target class.

forming parameters, the dense motif approach reaches over 95% in precision and recall on the badminton, zumba or cycling datasets. For the other datasets, namely flamenco, guitar and gym, the dense motifs approach needs a more detailed discussion.

The flamenco dataset (see also Figure 11) shows a wide performance difference between the chosen approaches. While mean and variance features give average results, FFT features perform surprisingly poorly. This could be explained by the fact that flamenco dancing is an activity with lots of irregular motions that are often complex. The dense motifs approach benefits the most from the characteristically short motions, which are not equally distributed over the whole activity, hence the slight drop in the recall. The equal error rate reaches 73% for the flamenco data set.

The dense motifs approach on the gym activity data was surprisingly low. The exercises consisted of different weightlifting workouts, with gestures being much slower compared to the other activities in this evaluation. Figure 12 shows the dense motif performance on 24 hours and a sub-sequence lasting for about 2 minutes. Inspection of different folds during the evaluation and comparison to other activities shows that the initial number of motifs is not very high in the first place, and is heavily reduced as motifs that appear in the background data are discarded (equal error rate of 60%). The slow motions are also the reason why the mean & variance trained SVM classifier fails at classifying the activity correctly. The FFT-based features, on the other hand, computed on a window of 5 seconds, were able to profit from the frequency domain characteristics of the gym exercise activity: The SVM classifier trained with these features performs considerably good reaching over 80% in equal error rate.

The playing guitar data turned out to be captured well with motifs, with the approach gaining a significant advantage over a classifier on the traditional features. While different ways to play were observed, including hitting or plucking the strings depending on musical genre, the dense motif approach still detected much of the activity, reaching 87% in equal error rate. Figure 13 illustrates one day, where the participant took a short break in activity. Such reduced performances due to the different ways to play the guitar, as well as breaks of varying durations, might be of particular interest to the psychiatric analysis and might provide additional hints regarding their patients' mood. Implementing detection for such events requires further investigations though and are left as future work.

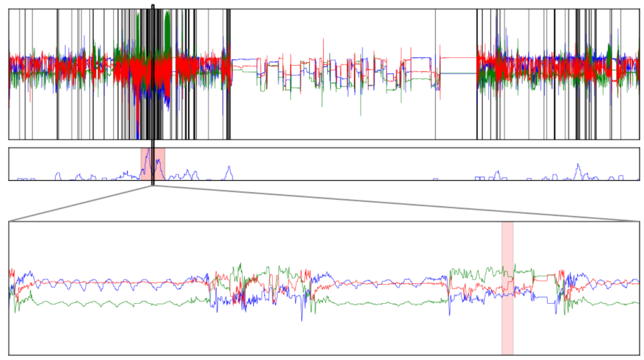

Figure 12. Dense motifs for a day of the gym data (upper plot, with gym activity marked in red in the middle plot) and a sub-sequence lasting 2 minutes (lower plot, with motifs marked in red). Two of the exercises can be recognized by periodically signals in the lower plot.

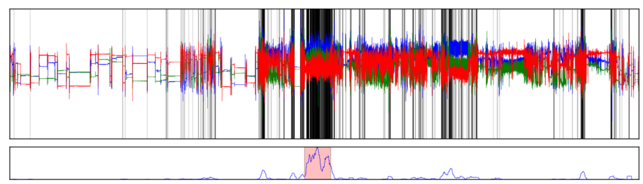

Figure 13. Dense motifs on one day with playing guitar for an hour: the gap in the middle of the activity (see bottom plot) was found to be due to a short bathroom break of the participant.

CONCLUSIONS AND FUTURE WORK

This paper presented a practical activity detection system to spot leisure activities in long-term datasets, that is based on finding parts in the data that contain frequent matches with a set of motifs. These dense motifs are discovered in exemplar training data by finding the most descriptive motifs for the activity against the large amount of background data. The approach has been designed for continuous deployment in psychiatry monitoring, and was evaluated on a data set with similar constraints, containing more than a month of data taken from a custom-built wrist-worn sensor unit that records 3D accelerometer data at 100Hz.

Experiments show that the approach is able to detect many physical activities, on par with standard approaches, reaching an equal error rate performance of 95% for 3 of the 6 activities, and only being significantly outperformed on one. It was demonstrated to be able to work on large and long-term sets of inertial data and can, unlike many traditional approaches, be expected to be scalable for weeks or months of such data. A remaining weakness identified is the method's reliance on short gestures, such that slower movements (such as weight lifting exercises) are not always picked as motifs.

While this work aimed at efficiency and therefore focused on extracting characteristic features, future work is underway to replace the bag-of-words classifier by more powerful models. The data set and code used in this paper are publicly available[2], to encourage reproduction of these results.

[2] at http://www.ess.tu-darmstadt.de, or by contacting the first author

ACKNOWLEDGEMENTS
This work was sponsored by the Project "Long-Term Activity Recognition with Wearable Sensors" (LA 2758/1-1) from the German Research Foundation (DFG).

REFERENCES
1. O. Amft, H. Junker, and G. Tröster. Detection of eating and drinking arm gestures using inertial body-worn sensors. In *ISWC '05*, pages 160–163. IEEE, 2005.

2. P. Bech, T. Bolwig, P. Kramp, and O. Rafaelsen. The bech-rafaelsen mania scale and the hamilton depression scale. *Acta Psych. Scand.*, 59(4):420–430, 1979.

3. K.-H. Chang, M. Chen, and J. Canny. Tracking free-weight exercises. In *UbiComp '07*, pages 19–37, Berlin, Heidelberg, 2007. Springer.

4. B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. In *KDD '03*, pages 493–498, New York, NY, USA, 2003. ACM.

5. T. Choudhury, G. Borriello, S. Consolvo, D. Haehnel, B. Harrison, et al. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7:32–41, 2008.

6. P. Corkum, R. Tannock, H. Moldofsky, S. Hogg-Johnson, and T. Humphries. Actigraphy and parental ratings of sleep in children with adhd. *Sleep*, 24(3):303, 2001.

7. R. Hamid, S. Maddi, A. Bobick, and M. Essa. Structure from statistics - unsupervised activity analysis using suffix trees. In *ICCV '07.*, pages 1–8. IEEE, 2007.

8. T. Holleczek, J. Schoch, B. Arnrich, and G. Tröster. Recognizing turns and other snowboarding activities with a gyroscope. In *ISWC '10*, pages 75–82, 2010.

9. T. Huynh and B. Schiele. Analyzing features for activity recognition. In *sOc-EUSAI '05*, pages 159–163, New York, NY, USA, 2005. ACM.

10. S. S. Intille, K. Larson, E. M. Tapia, J. S. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson. Using a Live-In Laboratory for Ubiquitous Computing Research. In *Pervasive*, 2006.

11. E. J. Keogh, S. Chu, D. Hart, and M. J. Pazzani. An online algorithm for segmenting time series. In *ICDM '01*, pages 289–296. IEEE, 2001.

12. K. V. Laerhoven, E. Berlin, and B. Schiele. Enabling efficient time series analysis for wearable activity data. In *ICMLA '09*, pages 392–397. IEEE, 2009.

13. J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *IJCAI '05*, pages 766–772. Morgan Kaufmann Publishers, 2005.

14. J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *DMKD '03*, pages 2–11, 2003.

15. M. Mahdaviani and T. Choudhury. Fast and scalable training of semi-supervised CRFs with application to activity recognition. In *NIPS*, December 2007.

16. D. Minnen, T. Starner, I. A. Essa, and C. L. I. Jr. Discovering characteristic actions from on-body sensor data. In *ISWC '06*, pages 11–18. IEEE, 2006.

17. N. Oliver and E. Horvitz. A comparison of HMMs and dynamic bayesian networks for recognizing office activities. In *UM 2005*, volume 3538, page 199, 2005.

18. J. Partridge, K.; Begole. Activity-based advertising: Techniques and challenges. In *Workshop on Pervasive Advertising 2009*, 2009.

19. D. Patterson, D. Fox, H. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *ISWC*, pages 44–51, 2005.

20. M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Fox, and H. Kautz. Inferring Activities from Interactions with Objects. *IEEE Pervasive Computing*, 3(4):50–57, October 2004.

21. T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, 7:42–50, 2008.

22. T. Stiefmeier, D. Roggen, and G. Troster. Fusion of string-matched templates for continuous activity recognition. In *ISWC '07*, pages 41–44. IEEE, 2007.

23. D. Tacconi, O. Mayora, P. Lukowicz, B. Arnrich, G. Tröster, and C. Haring. On the feasibility of using activity recognition and context aware interaction to support early diagnosis of bipolar disorder. In *Ubiwell Workshop 2007*, pages 206–209, 2007.

24. M. H. Teicher. Actigraphy and motion analysis: New tools for psychiatry. *Harvard Review of Psychiatry*, 3(1):18–35, 1995.

25. E. Ukkonen. Constructing suffix trees on-line in linear time. In *IFIP World Computer Congress on Algorithms, Software, Architecture - Information Processing '92*, pages 484–492. North-Holland Publishing Co., 1992.

26. A. Vahdatpour, N. Amini, and M. Sarrafzadeh. Toward unsupervised activity discovery using multi-dimensional motif detection in time series. In *IJCAI '09*, pages 1261–1266. Morgan Kaufmann Publishers Inc., 2009.

27. K. Van Laerhoven, H. Gellersen, and Y. Malliaris. Long term activity monitoring with a wearable sensor node. In *BSN 2006*, pages 171–174. IEEE, 2006.

28. F. H. Wilhelm, M. C. Pfaltz, and P. Grossman. Continuous electronic data capture of physiology, behavior and experience in real life. *Interacting with Computers*, 18:171–186, March 2006.

29. A. Zinnen, U. Blanke, and B. Schiele. An analysis of sensor-oriented vs. model-based activity recognition. In *ISWC '09*, pages 93–100. IEEE, 2009.