Activity Classification Using Realistic Data From Wearable Sensors

Juha Pärkkä, Miikka Ermes, Panu Korpipää, *Member, IEEE*, Jani Mäntyjärvi, Johannes Peltola, and Ilkka Korhonen, *Member, IEEE*

***Abstract*—Automatic classification of everyday activities can be used for promotion of health-enhancing physical activities and a healthier lifestyle. In this paper, methods used for classification of everyday activities like walking, running, and cycling are described. The aim of the study was to find out how to recognize activities, which sensors are useful and what kind of signal processing and classification is required. A large and realistic data library of sensor data was collected. Sixteen test persons took part in the data collection, resulting in approximately 31 h of annotated, 35-channel data recorded in an everyday environment. The test persons carried a set of wearable sensors while performing several activities during the 2-h measurement session. Classification results of three classifiers are shown: custom decision tree, automatically generated decision tree, and artificial neural network. The classification accuracies using leave-one-subject-out cross validation range from 58 to 97% for custom decision tree classifier, from 56 to 97% for automatically generated decision tree, and from 22 to 96% for artificial neural network. Total classification accuracy is 82% for custom decision tree classifier, 86% for automatically generated decision tree, and 82% for artificial neural network.**

***Index Terms*—Activity classification, context awareness, physical activity, wearable sensors.**

I. Introduction

PHYSICAL inactivity is a health risk that many people in both developed and developing countries are facing today. According to World Health Organization (WHO), at least 60% of the world's population fails to achieve the minimum recommendation of 30 min moderate intensity physical activity daily [13]. The main reason for not achieving this basic level of physical activity is that the level of activity required in work, in travel, and at home is decreasing with sedentary work and with the advent of technologies that are designed to ease home activities and traveling. The physical activities on free time are insufficient or too irregular to achieve the weekly goal. Physical inactivity is known to contribute in many chronic diseases, such as cardiovascular disease, type 2 diabetes, and possibly certain types of cancer and osteoporosis [2], [3]. As the population is rapidly aging in many countries, promotion of a healthier lifestyle, especially for the elderly population, can provide substantial savings in future health care costs.

By following the minimum recommendation, many health benefits can be obtained, when compared with completely inactive people [4]–[6]. The basic level of physical activity helps, for example, in managing weight, in lowering blood pressure, in increasing the level of the good high-density lipoprotein (HDL) cholesterol, in improving sugar tolerance, and in changing hormone levels to a direction more suitable for preventing cancer [3], [7]. The basic level of physical activity can be achieved by everyday activities like walking at work, shopping, gardening, cleaning, etc. The 30-min daily physical activity targets to at least 1000 kcal energy expenditure weekly. The only limitation in achieving the goal is that the daily 30-min physical activity must be collected in continuous periods of a minimum 10 min.

Level of daily physical activity can be measured objectively by measuring energy expenditure. The accelerometer signal has been used previously to estimate energy expenditure, and the estimate has been shown to correlate well with true energy expenditure [8]. Although achieving the minimum recommendation of physical activity brings many health benefits, even more health benefits can be achieved by taking part in a more vigorous [5] and a wider spectrum of physical activities. For example, endurance-enhancing activities and activities maintaining flexibility and muscular strength bring health benefits that are not achieved with basic activity [3]. Endurance can be enhanced, e.g., with energetic walking, jogging, cycling, and rowing. Activities maintaining functions of the musculoskeletal system are, e.g., ball games, gym, and dancing. Thus, in addition to daily energy expenditure, activity types play an important role in overall well being and health.

Accelerometers have been shown to be adequate for activity recognition. The studies using accelerometry for monitoring human movement have been recently reviewed in [9] and [10]. In laboratory settings, the most prevalent everyday activities (sitting, standing, walking, and lying) have been successfully recognized with accelerometers [11]–[15]. However, applicability of these results to out-of-lab monitoring is unclear. For example, in [15] the recognition accuracy of nine patterns decreased from 95.8% to 66.7% as the recordings were shifted outside the laboratory. Also, recognition of different activities involving dynamic motion has not yet been studied thoroughly. In a few studies data have been collected outside the laboratory. In [15] 24 subjects spent approximately 50 min outside laboratory. Accelerometers were placed on sternum, wrist, thigh, and lower leg. Nine patterns (sitting, standing, lying, sitting and

Manuscript received December 21, 2004; revised June 16, 2005. This work was supported in part by National Technology Agency of Finland, Nokia, Suunto and Clothing+.

J. Pärkkä, M. Ermes, and I. Korhonen are with VTT Information Technology, 33101 Tampere, Finland (e-mail: juha.parkka@vtt.fi; miikka.ermes@vtt.fi; ilkka.korhonen@vtt.fi).

P. Korpipää, J. Mäntyjärvi, and J. Peltola are with VTT Electronics, 90571 Oulu, Finland (e-mail: jani.mantyjarvi@vtt.fi; panu.korpipaa@vtt.fi; johannes.peltola@vtt.fi).

Digital Object Identifier 10.1109/TITB.2005.856863

talking, sitting and operating PC, walking, stairs up, stairs down, and cycling) were recognized from presegmented data with an overall accuracy of 66.7%. In [16] five biaxial accelerometers attached to hip, wrist, arm, ankle, and thigh were used to recognize 20 everyday activities such as walking, watching TV, brushing teeth, vacuuming, etc. From 82 to 160 min of data were collected from 20 subjects and a decision tree classifier was used for classification. Recognition accuracies ranged from 41 to 97% for different activities.

Many research groups have recently studied activity recognition as part of context awareness research [16]–[22]. Context sensing and use of context information is an important part of the ubiquitous computing scenario [23]–[25]. Context sensing aims at giving a computing device (e.g., cellular phone, wrist-top computer, or a device integrated into clothes) senses, with which it becomes aware of its surroundings. With the senses the device is capable of measuring its user and environment and it becomes context aware. The context describes the situation or status of the user or device. Different devices can use the context information in different ways, e.g., for adapting its user interface, for offering relevant services and information, for annotating digital diary (e.g., energy expenditure), etc. Location and time belong to the group of the most important contexts and the use of these contexts has been studied extensively. However, to recognize the physical activities of a person, a sensor-based approach is needed.

Our vision in automatic classification of physical activities is to contribute to long-term monitoring of health and to a more active lifestyle. The application we have in mind can be called an "activity diary". The diary would show the user which activities he did during the day and what were the daily cumulative durations of each activity. When the user is shown this information, he can draw the conclusions himself and adjust his behavior accordingly. This model is called the *behavioral feedback model* [26]. This model is being successfully used, e.g., in weight management programs. On the other hand the activity diary information can be utilized by context-aware services and devices that offer adapted information or adapt their user interface (UI) based on the user's activity type.

In this work our aim was to study activity classification, which are the most information-rich sensors and what kind of signal processing and classification methods should be used for activity classification. We took a data-oriented and empirical approach and collected a large data library of realistic data. In this paper, we describe methods for automatic activity classification from data collected with body-worn sensors.

II. Methods

A. Data Collection

The goal of our data collection was to assess the feasibility and accuracy of context recognition based on realistic data. We collected a large data library of realistic context data with many different sensors (accelerometers, physiological sensors, etc.) and with many test persons. The collected data were then used in development of context recognition algorithms. A data collection system was developed for sensing and storing context-related data in real-life conditions. Only the sensors are small in size that were applicable to ambulatory measurements were used. The data were stored on a rugged, compact PC (Databrick III, Datalux Corporation, Winchester, VA, USA) and on a flash-card-memory-based, 19-channel recorder (Embla A10, Medcare, Reykjavik, Iceland). Additionally, two stand-alone devices were used: Global Positioning System (GPS) recorder (Garmin eTrex Venture, Garmin Ltd., Olathe, Kansas, USA) and wrist-top computer that measured heart rate and altitude (Suunto X6HR, Suunto Oy, Vantaa, Finland). The PC and recorder were placed into a normal rucksack (dimensions: 40 cm $\times$ 30 cm $\times$ 10 cm, weight 5 kg with the equipment) that the test persons carried during the measurement sessions. The sensors were put on the test person with help of an assistant before the start of the measurement session. The system measured 18 different quantities from the user and his environment (Table I). Some of the quantities were measured with multiple sensors, which resulted in altogether 22 signals and 35 channels of data.

During the measurement sessions, the test persons followed a scenario (Table II) that describes the tasks they should at least do and locations they should at least visit. The scenario consists of visits to several everyday places (bus, restaurant, shop, and library) and of several physical activities (lying, sitting, standing, walking, Nordic walking, running, rowing, cycling). Nordic walking is fitness walking with specifically designed poles to engage the upper body.

Because the signals have large interindividual difference in different activities, we recruited 16 volunteers (13 males, 3 females, age 25.8 $\pm$ 4.3 years, body mass index [BMI] 24.1 $\pm$ 3.0 kg/m^2) to gather a representative dataset for algorithm development. The volunteers were recruited by using bulletin board and news advertisements at a local university. The duration of each measurement session was about 2 h. The durations varied between measurement sessions because of the loose scenario, which was not supposed to be followed strictly. Because the goal was to collect realistic data, the test persons were given a lot of freedom during the measurement session. For example, they could choose the restaurant and shop they preferred. Also the order of places visited and time spent in each place depended on the test person.

The test person was accompanied by an annotator (same person for all cases), who used a personal digital assistance (PDA) to mark changes in context for reference purposes. An annotation application (Fig. 1) was written for a PDA using C#.NET. The annotation application provides a UI for visualizing and changing the currently selected and active contexts. In the UI, the contexts are organized hierarchically into upper level context types, e.g., *activity* and lower level context values, e.g., *lie* or *sit*. The context values are mutually exclusive. As a context value changes, the annotator taps on the name of the new context value with the PDA pen. The software stores the new state together with a timestamp on PDA

TABLE I
SIGNALS AND SENSORS OF DATA COLLECTION SYSTEM

Signal	Sensor	Measurement site	Fs
Altitude	Air Pressure (Suunto X6HR)	Wrist	0.5
Audio	Microphone (AKG C417)	Chest, on rucksack strap	22000, mono, 16 bit
Body Position	Metal ball moves between resistors (ProTech Position)	Chest	200
Chest Accelerations	3D acceleration (2 x Analog Devices ADXL202)	Chest, on rucksack strap	200
Chest Compass	3D compass (Honeywell HMC-1023)	Chest, on rucksack strap	200
EKG	Voltage between EKG electrodes (Blue Sensor VL, Embla A10)	Below left armpit, on breastbone	200
Environmental Humidity	Humidity (Honeywell, HIH-3605-B)	Chest, on rucksack strap	200
Environmental Light Intensity	Light sensor with two output dynamics (Siemens SFH 203P)	Chest, on rucksack strap	200
Environmental Temperature	Temperature sensor (Analog Devices TMP36)	Chest, on rucksack strap	200
Event Button	Switch (Embla XN Oximeter)	Chest, on rucksack strap	-
Heart Rate	IR light absorption (Embla XN oximeter)	Finger	1
Heart Rate	IR light reflectance (Nonin XPOD)	Forehead	3
Heart Rate	Voltage between chest belt electrodes (Suunto X6HR)	Chest	0.5
Location	GPS satellite receiver (Garmin eTrex Venture)	Shoulder, on rucksack strap	Based on location
Pulse Plethysmogram	IR light reflectance (Nonin XPOD)	Forehead	75
Respiratory Effort	Piezo sensor (Pro-Tech Respiratory Effort)	Chest	200
SaO2	IR light absorption (Embla XN Oximeter)	Finger	1
SaO2	IR light reflectance (Nonin XPOD)	Forehead	3
Skin Resistance	Resistance between two metal leads (Custom-made)	Chest	200
Skin Temperature	Resistive temperature sensor (YSI 409B)	Upper back, below neck	200
Wrist Accelerations	3D acceleration (Analog Devices, ADXL 202E)	Wrist, dominant hand	40
Wrist Compass	2D compass (Honeywell HMC-1022)	Wrist, dominant hand	40

TABLE II
SCENARIO FOR DATA COLLECTION

Location	Task
Home	Sitting at home
	Lying
	Sitting & reading newspaper
	Putting clothes on, going out
Bus	Walking to a bus stop
	Waiting for bus
	Traveling in bus
Restaurant	Walking to restaurant
	Queuing
	Eating, drinking, talking
Library	Walking to library
	Sitting in library, reading
Shop	Walking to shop
	Walking in shop, shopping
Home	Walking back home
Outdoor activities	Nordic Walking
	Running
Indoor activities	Rowing (rowing machine)
	Walking
	Bicycling (exercise bike)
	Sitting, drinking

Fig. 1. Annotation Software on PDA. Checkboxes on the left are used to expand and collapse between the title line and full view. Radio buttons are used to mark the active context value. Eating and Drinking can be active simultaneously. The asterisk is used to mark the context value "other."

memory. Data collection start and end markers were manually added to annotation data and all context data to allow synchronization of the data. The accuracy of manual markers is ± 0.5 s. In 2-h data collection this was considered an adequate accuracy.

B. Context Data Library

After the measurements, the data were synchronized, calibrated, re-sampled, and converted into suitable formats [27] for visualization. All the data (31 h) were collected into *Palantir Context Data Library 2003*.

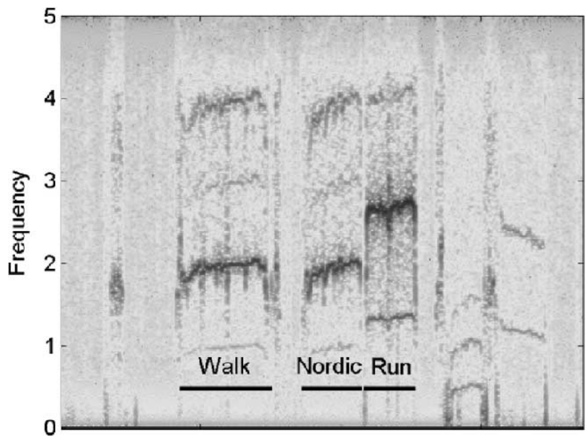

Fig. 2. Spectogram of vertical acceleration on chest during walking, Nordic walking and running. Horizontal axis is time.

C. Signal Processing and Feature Extraction

The goal in context recognition is to develop algorithms that can automatically infer the annotated context from the collected signals. The signals were first visually inspected and compared against the annotated contexts. This gave us the first impression on which signals are more useful than others. Feature signals (1-Hz sampling rate) were calculated from the raw data.

A priori information was used to select which features to calculate. For example, walking and running (measured in realistic circumstances) have constant frequency, which did not vary much between test persons either. Walking is seen as 2 Hz and running as 2.5–3 Hz oscillation in the signal (Fig. 2).

Time-domain features calculated were mean, variance, median, skew, kurtosis, 25% percentile and 75% percentile counted using a sliding window. *Frequency-domain features* were spectral centroid, spectral spread, estimation of frequency peak, estimation of power of the frequency peak, and signal power in different frequency bands. For acceleration both 4-s and 10-s windows were used. For blood oxygen saturation (SaO_2) the window was 10 s, for respiratory effort it was 60 s, and for all others it was 1 s.

Time-domain features were calculated for 1) body position; 2) humidity; 3) blood oxygen saturation SaO_2 ; 4) skin resistance; 5) skin temperature; and 6) environmental temperature. Both time and frequency domain features were calculated for 1) accelerations; 2) magnetometer signals; 3) environmental light intensity; and 4) respiratory effort.

In addition to the basic time- and frequency-domain features, the following features were calculated. Speech was detected from *audio* signal using a modified version of a speech/music discriminator [28]. Radius and two angles describing the vector of magnetic field as well as ratio between frequency bands 1–1.5 Hz and 0–5 Hz were calculated from *magnetometer signals*. R-peaks were detected and different features related to heart rate variability (e.g., R-R interval) were calculated from the electrocardiogram [29]. Speed was calculated from *GPS location data*. Power on frequency band 80–100 Hz was calculated from a *light-intensity signal*. Respiratory frequency, tidal volume, frequency and amplitude deviations, rate of ventilation and spectral entropies were estimated and calculated from the *respiratory effort signal*.

D. Feature Selection

Feature selection was based on visual and statistical analysis. The features were visually compared against annotation to find good candidate features. Distribution bar graphs of each feature signal during different activities were plotted for comparison (Fig. 4). The plots show how the distribution of each feature signal changes between different activities. The more the distribution moves between activities and the less the distributions overlap, the better it is for discrimination of activities.

A priori information was used in the quest for the best features. For example, during running there is more up-down movement and thus more energy in acceleration signal than during other activities. Based on *a priori* information, some new features were calculated from raw data. The best features were selected based on the distribution bar graphs. If there were more than one feature that could have been used for a specific decision, the feature with best discrimination power was selected.

As a result of the feature selection process, six features (Fig. 4) were selected for classification: 1) peak frequency of up-down chest acceleration F_{peak} *(chestacc,y)*; 2) median of up-down chest acceleration *Med(chestacc,y)*; 3) peak power of up-down chest acceleration P_{peak} *(chestacc,y)*; 4) variance of back-forth chest acceleration *Var(chestacc,z)*; 5) sum of variances of three-dimensional (3-D) wrist accelerations $\sum$ *Var(wristacc, 3 D)*; 6) power ratio of frequency bands 1–1.5 Hz and 0.2–5 Hz measured from left-right magnetometer on chest P_1 *(chestmagn,x)*.

E. Classification

During the feature selection process it was noticed that with the selected sensor setup, it was not possible to discriminate sitting and standing from each other (see Discussion for more details). Thus sitting and standing were combined into one class, resulting in seven target classes for classification: 1) lying; 2) sitting/standing; 3) walking; 4) Nordic walking; 5) running; 6) rowing (with a rowing machine); and 7) cycling (with an exercise bike). Three different classifiers were used in classification. All of them were given the same set of six features as inputs.

For classification, two decision trees were applied, namely a custom decision tree and an automatically generated decision tree. Also, an artificial neural network (ANN) was used as a reference classifier. Decision trees have been successfully applied to activity recognition earlier [16]. The custom decision tree was selected to represent a simple and transparent approach based on human rationalization. The automatically generated decision tree was selected to see how well the automatic tree generation algorithm performs compared with human-made rules. An advantage of the decision trees is that the problem of context recognition is divided in to smaller subproblems, which are tackled one by one very intuitively.

The recorded data were used for context recognition on a second-by-second basis by using the feature signals as inputs

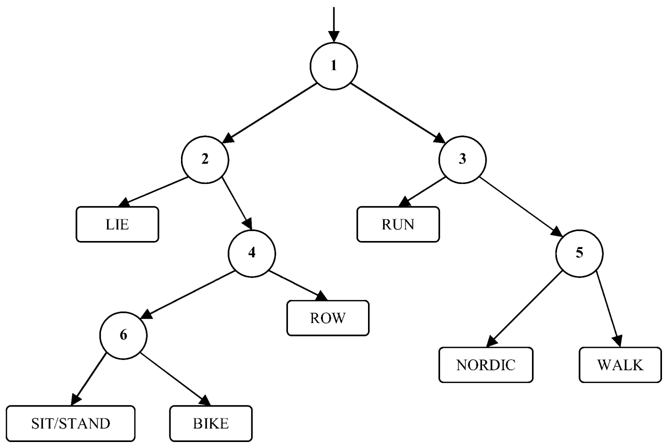

Fig. 3. Custom decision tree.

and PDA annotations as targets. For all three classifiers the results were acquired by 12-fold leave-one-subject-out cross-validation.

1) Custom Decision Tree: The custom decision tree (Fig. 3) was built by using domain knowledge and visual inspection of the signals. The tree has 13 nodes, 7 of which are leaf nodes and 6 of which represent a binary decision. The decisions can be named with questions: 1) footsteps?; 2) lying?; 3) running?; 4) rowing?; 5) Nordic walking?; 6) cycling?. The numbering refers to numbers of the six selected features. Leaf-nodes "sitting/standing" and "walking" can be considered as classes "other," because everything that is not recognized as any of the activities in upper levels of the tree falls into these categories. This is in line with the data, because the context value "other" was not used in annotations either. Fig. 4 depicts the decisions made in the nodes: it shows the distributions of feature data during each activity. The circled activities are relevant for the node; others have been ruled out in the upper level decisions. For each branch of the tree, the threshold value was defined by using a 12-fold leave-one-subject-out cross-validation. The threshold value for each node was chosen to be the average of the acquired 12 thresholds. The threshold values remained unchanged during the whole validation process.

2) Automatically Generated Decision Tree: An automatically generated decision tree was generated using a Matlab (MathWorks Inc, Natick, MA) Statistics Toolbox function called "treefit." The rule for splitting was Gini's index [30], which is one of the standard options. It progressively looks for the largest class in the data set and tries to isolate it from the rest of the data. The results were obtained by using leave-one-subject-out cross-validation resulting in separate training/validation sessions for each subject. In each training/validation session the tree was built using the training data (containing data from all but one subject), pruned to an optimum level (the level with the lowest error rate in the training set) using cross-validation within the training data, and the obtained tree was used to classify the data of the left-out subject. It should be noted that the size of the tree may be different in each training/validation session. In average the tree had 9.7 branches (minimum, 7; maximum, 14) and 10.7 leafs (minimum, 8; maximum, 15).

3) Artificial Neural Network: A multilayer perceptron with resilient backpropagation as the training algorithm was used as the artificial neural network classifier. The sizes of input, hidden and output layers were 6, 15 and 7, respectively. The output that had the highest value was selected as the classification result.

F. Postprocessing

Classification was made for each second of the data independently, and no temporal connections were considered. This resulted in rapid changes of the classification results especially at transitions between two activities. For instance, getting up from a sitting or lying position produced high acceleration peaks that caused misclassification. Activities that only last for a few seconds are not realistic. Thus, median filtering was used on the results of all three classifiers to use simple temporal logic to filter out short-duration misclassifications. The median filter replaces short activities with the surrounding longer duration activity. After several experiments, a median filter of 31 s was selected. A median filter this long may prevent the recognition of some short periods of activities (such as short walks) but it improves the overall classification. Both causal and anticausal versions were tested, and with the selected filter length their results were very close to each other. Anticausal filtering worked slightly better. Fig. 5 demonstrates the difference between filtered and unfiltered results.

III. Results

A. Data Quality

Data from 12 of 16 cases were used in classification. Data of four cases were left out because of missing wrist acceleration signals. The wrist acceleration signals were lost because of a hardware problem.

B. Classification Results

Tables III–V show the confusion matrices for the three different classifiers. In the tables, each sample represents 1 s. Table VI summarizes the classification accuracies of different activities.

IV. Discussion

We classified activities from realistic, out-of-lab context data using three different classifiers and six feature signals as inputs. Classification was done with 1-s time resolution; thus each second of the data was classified and compared with annotated data. Only a few previous studies have recognized activities from data measured in the out-of-lab environment. Rowing and Nordic walking have not been recognized in previous studies. Lying, sitting, standing, walking, running, and cycling have also been recognized in previous studies.

Bao and Intille [16] achieved recognition accuracy of 94.96% for lying down and relaxing, 94.78% for sitting and relaxing, 95.67% for standing still, 89.71% for walking, 87.,68% for running, and 96.29% for bicycling. Their data were measured in

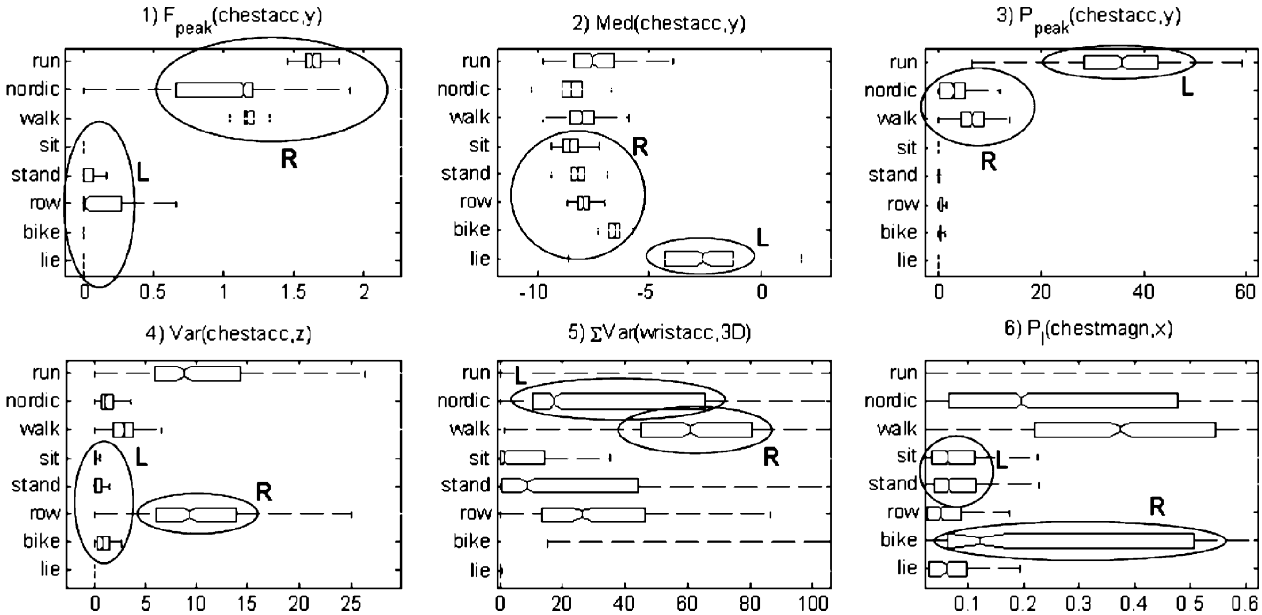

Fig. 4. Nodes of custom decision tree. Figure depicts distributions of feature signals during different activities. Activities marked with an R fall into right branch and activities marked an L fall into left branch of node. Circled activities are relevant for the node, others have been ruled out in upper level nodes.

Fig. 5. Classification results before (top) and after (bottom) median filtering. During black time intervals on the timeline sitting is classified as active $(\mathrm{sitting} = \mathrm{true})$. During white time intervals, sitting is not classified as active $(\mathrm{sitting} = \mathrm{false})$. Most of the sitting intervals that are shorter than 15 seconds are replaced with the dominant activity by median filtering.

TABLE III
CONFUSION MATRIX OF CUSTOM DECISION TREE

Annotation	Recognized Activity Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk
Lie	1417	0	0	205	0	0	0
Row	0	1646	0	717	0	0	23
ExBike	0	0	2461	612	0	0	29
Sit/ Stand	121	40	53	34083	4	340	962
Run	0	0	0	44	2284	21	5
Nordic walk	0	1	0	256	39	4507	194
Walk	0	16	4	5412	15	3964	12797

TABLE IV
CONFUSION MATRIX OF AUTOMATICALLY GENERATED DECISION TREE

Annotation	Recognized Activity Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk
Lie	1354	0	0	268	0	0	0
Row	0	1327	0	1028	0	0	31
ExBike	0	0	2552	508	0	0	42
Sit/ Stand	86	109	66	33928	8	13	1385
Run	0	0	0	45	2293	0	16
Nordic walk	0	0	0	250	36	3581	1130
Walk	0	27	4	4249	24	642	17262

TABLE V
CONFUSION MATRIX OF ARTIFICIAL NEURAL NETWORK

Annotation	Recognized Activity Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk
Lie	1206	0	0	357	0	0	59
Row	0	1414	0	874	63	0	35
ExBike	0	0	2336	561	0	0	205
Sit/ Stand	41	131	27	34032	0	22	1345
Run	0	250	0	40	517	1070	477
Nordic walk	0	2	0	210	0	2597	2188
Walk	0	18	4	4620	0	109	17457

a naturalistic environment, which is comparable to our setting. They used five acceleration sensors on the hip, wrist, arm, ankle, and thigh. They concluded that the thigh and wrist could be the ideal locations for activity recognition.

Absence of an accelerometer on the lower body is a limitation in our study. An extra accelerometer on the lower body would probably improve classification accuracy. Placing an accelerometer on the thigh was also considered in our study, but the thigh was not seen as a feasible sensor placement for a consumer product and this placement was ignored.

Foerster *et al.* [15] achieved recognition accuracy (subactivities combined) of 89% for lying, 100% for sitting, 88% for standing, 99% for walking, and 100% for cycling. Their data were collected in an out-of-lab environment. They segmented the data manually into 20 s or longer segments according to the behavior observation. Results were obtained by classifying the selected segments only (466 segments). About the segmentation they mention: "The classification can be improved by lengthening segments..." They used four sensor placements (chest, wrist, thigh, and lower leg). In our study, 1-s segments were used (72 272 segments).

TABLE VI
CLASSIFIER RESULTS [%]

	Custom Decision Tree	Automatic Decision Tree	Artificial Neural Network
Lie	**87**	83	74
Row	**69**	56	59
ExBike	79	**82**	75
Sit/ Stand	**96**	95	**96**
Run	**97**	**97**	22
Nordic walk	**90**	72	52
Walk	58	78	**79**
TOTAL	82	**86**	82

A. *Confusions*

Much of our classifiers' confusion seen in the results can be explained with transitions from one activity to another. The annotator was not given the choice to annotate "transition," but he had to switch from one activity to another instantly at some point during the transition. The transition is sometimes gradual, for example, when sitting changes to lying. The resulting inaccuracy is especially visible in the recognition of lying, which should be detected almost perfectly from the direction of gravity. Because lying periods were short, the uncertainty caused by transition periods in the beginning and end of lying became significant.

Lying is detected by the custom-made decision tree as a combination of decisions "no footsteps" and "lying." Duration of each lying period was only 2 min per case (total, 27 min) and confusion is 13 s per case (total, 3.5 min). The inaccuracy in annotation and duration of transition from sitting/standing to lying was in practice in this order. The artificial neural network additionally confuses lying with walking.

Recognition of *running* combines decisions "footsteps" and "running." Recognition of footsteps is rather clear (Fig. 4, node 1). The distributions of activities including footsteps and not including footsteps do not overlap much. The total amount of annotated running is about 39 min. The custom decision tree and the automatic decision tree recognize running very well. About 1 min of running is confused with standing and a few seconds with walking. Again, at least on part, the classifiers can be more accurate than the annotation and part of the confusion is not really confusion at all. Running started from the standing position and because of cars, slippery weather, etc. some walking and stops are included in the period annotated as "running." Artificial neural network confuses running heavily with other activities, especially with Nordic walking and walking.

Rowing is recognized as combination of decisions "no footsteps," "no lying." and "rowing." The custom decision tree recognizes 27 min of the total 40 min annotated as rowing. Because this includes data from 12 cases and rowing was started and ended by sitting, some sitting may indeed have been annotated as rowing. However, the amount of confusion toward sitting is rather large, so some classification error is also present. In addition, confusion with walking cannot be explained with annotation inaccuracy. The automatic decision tree similarly confuses rowing with sitting/standing and with walking. The artificial neural network commits the same error and further confuses 1 min of rowing as running.

Walking is one of the most common activities in this data set as in everyday life. Walking is recognized as combination of decisions "footsteps," "no running," and "no Nordic walking." Distributions of walking and Nordic walking partly overlap when using the feature in node 5. Both decision trees confuse walking mostly with Nordic walking and with sitting/standing. The artificial neural network confuses walking mostly with sitting/standing. Confusion with sitting/standing can be explained with inaccuracies in annotation. Activity annotated as walking often includes short periods of standing. Very short periods of walking between other dominant activities, even if annotated correctly and classified correctly by the decision tree, are replaced with dominating activity by the postprocessing method in the classification results. This degrades the performance when lots of short periods of walking are present.

Nordic walking was detected from increased arm motion. This approach is successful when the poles are used as effectively as they should be used. People not familiar with Nordic walking tend to use the poles very little and smoothly. Such use of the poles creates problems for recognition because the accelerations measured from the wrist have very low amplitude. This fact can be utilized, e.g., in teaching effective Nordic walking.

Sitting/standing is the most dominant activity in this data library and for most people in their everyday lives. It is recognized by combining decisions "no footsteps," "no lying," "no rowing," and "no cycling." All of the three classifiers classify sitting/standing rather well, mostly confusing them with walking. This is partly due to annotation inaccuracy. For example, in a library the activity annotated as standing includes very short periods of walking, which has not always been annotated correctly. Also, if annotated and classified correctly, very short periods of standing are replaced with the dominating activity by postprocessing.

Cycling with an exercise bike is detected from the left-right movement of chest by using the magnetometer signal. The distribution of cycling in this feature overlaps slightly with sitting/standing and thus some confusion with sitting/standing is inevitable. A small amount of annotation inaccuracy can be present mostly in the beginning and end of activities annotated as exercise biking. These are because the test person does not start or stop cycling exactly at the same time with annotation.

B. *Classifiers*

The custom decision tree treats the different activities more equally than the other classifiers because it optimizes performance of one node at a time, not the overall performance as the other classifiers do. That is why it has the best recognition accuracy for more than half of the activities, but the overall accuracy is not the best. The automatically generated tree had the best overall performance. This is in line with earlier studies [16]. For artificial neural network classification, the everyday data are rather noisy. Thus the artificial neural network easily overfits. Noticeable in neural network results is the poor recognition of running, which was well recognized by both of the decision tree classifiers.

C. Physiological Signals

Physiological signals such as heart rate and respiration were expected to have a larger role in activity recognition. Although they have been previously used together with accelerometers in ambulatory monitoring [14], they did not provide very useful data for activity recognition in our setup, in part, because they react to activity changes with a delay. The physiological signals correlate with the intensity level of the activity, but they do not reflect the type of activity (e.g., cycling versus walking), nor the duration of activity very accurately. With physiological signals (e.g., heart rate), the interindividual difference is also large, which creates extra challenge for algorithm development.

D. Sensors

In this study, accelerometers proved to be the most information-rich and most accurate sensors for activity recognition. They react fast to activity changes and they reflect well the type of activity. Placement of accelerometers in this study on rucksack straps and on wrists did not make it possible to separate sitting and standing from each other, because there was no clear change in the signal properties between these two activities. Different approaches were tried for detection of these activities. For example, it was assumed that the direction of a test person's body would stay more stable during sitting than during standing. However, the recorded data did not show such behavior. In the future, we will consider placing one accelerometer on the waist to enable discrimination of sitting from standing.

Even though gravity and magnetic flux are fundamentally different measures (e.g., direction), our data showed that for activity recognition, the information content of accelerometer and magnetometer signals is similar. Our 3D magnetometer and 3D accelerometer were located in one box, attached on a rucksack strap. When visually comparing the signals recorded during different physical activities, the magnetometer signal looks like a low-pass-filtered version of the accelerometer signal.

E. Temporal Connections Between Activities

Temporal connections between activities were not thoroughly studied in this work. In this study a median filter was used to remove very short activities from classifier results. Use of median filtering degrades the classification accuracy of the short-duration activities, which may be a problem in some applications. However, when aiming for a daily summary of activities, this is not a major problem. Utilizing the temporal history of activities might improve accuracy of activity recognition. Probabilistic models can be used to help in classification process, especially in transition from one recognized activity type to another. In [20] Markov chains have been used to assign probabilities to state transfers from one activity to another. The model is used to inhibit class change based on raw data only. If the transition has low probability, more requests from raw data classification are required before the change is accepted by the overall classification system. The drawback of this approach is that it requires a lot of realistic training data and probably also user-specific training data. However, in the future we will consider using a probabilistic model to reduce the number of (short-duration) false recognitions.

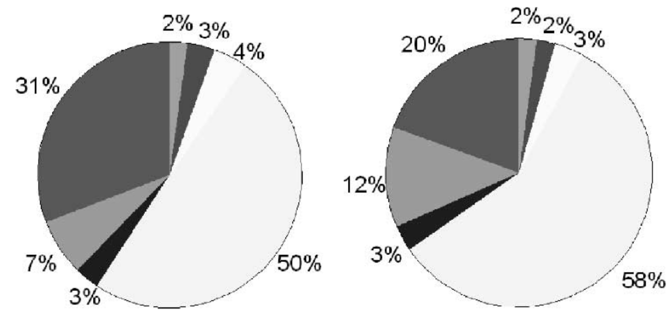

Fig. 6. Portions of activities in annotation (left) and results of custom decision tree (right). Activities clockwise from 12 o'clock: lying, rowing, cycling, sitting/standing, running, Nordic walking, walking.

F. Rucksack

Weight of the rucksack with the equipment was approximately 5 kg. This felt like a normal rucksack. Before selecting the rucksack, we also tried a belt bag, but compared with the rucksack, it felt uncomfortable with the equipment. In the data collection, the rucksack may have some effect on the activities, but it was not considered disturbing by the volunteers. Note that in this study placement of the chest acceleration sensor on the rucksack strap may affect the signal during dynamic activities, like running, because the rucksack moves slightly. However, in overall activity classification, the effect caused by rucksack movement is not significant.

G. Application: Activity Diary

Automatic classification of everyday activities can be used for promotion of a healthier lifestyle, e.g., with an "activity diary." The user could, e.g., in the evening check what kind of activities he has done during the day and how much time he has spent on each activity. Fig. 6 depicts the portions of our data in the form of an "activity diary."

V. Conclusion

Results of activity recognition were encouraging. With careful selection and placement of sensors, several everyday activities can be automatically recognized with good accuracy by using feature extraction and classification algorithms. Information about the daily activities can be used in consumer products to show the user his daily activity diary. This would increase the user's awareness of his daily activity level and promote a more active lifestyle.

Acknowledgment

The authors would like to thank A. Ylisaukko-oja, L. Cluitmans, J. Vilmi, P. Välkkynen, M. van Gils, and S.-M. Mäkelä for their work and help during the project. The authors would also like to thank Finnish Institute of Occupational Health for letting us use some of their equipment. Special thanks also to all volunteers for collecting a valuable data library with us.

REFERENCES

[1] P. Puska, H. Benaziza, and D. Porter (2004, Oct. 15). Physical Activity, WHO Information Sheet on Physical Activity, WHO [Online]. Available: http://www.who.int/ dietphysicalactivity/media/en/gsfs_pa.pdf

[2] H. Hagendoorn, I. Vuori, and P. Oja (2004, Oct. 15). Guidelines for the development of national policies and strategies for promoting health through physical activity, The European Network for the Promotion of Health-Enhancing Physical Activity [Online]. Available: http://www.who.int/gb/ebwha/pdf_files/WHA57/A57_R17-en.pdf

[3] R. R. Pate, M. Pratt, S. N. Blair, W. L. Haskell, C. A. Macera, C. Bouchard, D. Buchner, W. Ettinger, G. W. Health, A. C. King, A. Kriska, A. S. Leon, B. H. Marcus, J. Morris, R. S. Paffenbarger, K. Patrick, M. L. Pollock, J. M. Rippe, J. Sallis, and J. H. Wilmore, "Physical activity and public health: a recommendation from the centers for Disease Control and Prevention and the American College of Sports Medicine," *JAMA*, vol. 273, no. 5, pp. 402–407, Feb. 1995.

[4] 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," *Eur. Heart J.*, vol. 25, pp. 2204–2211, 2004.

[5] J. E. Manson, P. Greenland, A. Z. LaCroix, M. L. Stefanick, C. P. Mouton, A. Oberman, M. G. Perri, D. S. Sheps, M. B. Pettinger, and D. S. Siscovick, "Walking compared with vigorous exercise for the prevention of cardiovascular events in women," *N. Engl. J. Med.*, vol. 347, no. 10, pp. 716–725, Sep. 2002.

[6] M. Martinez-Gonzalez *et al.*, "Physical inactivity, sendentary lifestyle and obesity in the European Union," *Int. J. Obes. Relat. Metab. Disord.*, vol. 23, no. 11, pp. 1192–1201, Nov. 1999.

[7] M. Fogelholm, P. Oja, M. Rinne, J. Suni, and I. Vuori, "Riittääkö puoli tuntia kävelyä päivässä" *Suomen Läökärilehti*, vol. 59, no. 19, pp. 2040–2042, May 2004.

[8] C. Bouten, A. Sauren, M. Verduin, and J. Janssen, "Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking," *Med. Biol. Eng. Comput.*, vol. 35, no. 1, pp. 50–56, Jan. 1997.

[9] M. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiol. Measure.*, vol. 25, no. 2, pp. R1–R20, Apr. 2004.

[10] K. Aminian and B. Najafi, "Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications," *Comput. Animation Virtual Worlds*, vol. 15, no. 2, pp. 79–94, May 2004.

[11] M. Mathie, A. Foster, N. Lovell, and B. Celler, "Detection of daily physical activities using a triaxial accelerometers," *Med. Biol. Eng. Comput.*, vol. 41, no. 3, pp. 296–301, May 2003.

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

[13] F. Foerster and J. Fahrenberg, "Motion pattern and posture: Correctly assessed by calibrated accelerometers," *Behav. Res. Methods Instrum. Comput.*, vol. 32, no. 3, pp. 450–457, Aug. 2000.

[14] J. Ng, A. Sahakian, and S. Swiryn, "Accelerometer-based body-position sensing for ambulatory electrocardiographic monitoring," *Biomed. Instrum. Technol.*, vol. 37, no. 5, pp. 338–346, Sep. 2003.

[15] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Comput. Hum. Behav.*, vol. 15, no. 5, pp. 571–583, 1999.

[16] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. 2nd Int. Conf. Pervasive Computing*, 2004, pp. 1–17.

[17] V.-M. Mäntylä, J. Mäntyjärvi, T. Seppänen, and E. Tuulari, "Hand gesture recognition of a mobile device user," in *Proc. Int. IEEE Conf. Multimedia Expo (ICME)*, 2000, pp. 281–284.

[18] J. Mäntyjärvi, J. Himberg, and T. Seppänen, "Recognizing human motion with multiple acceleration sensors," in *Proc. Int. IEEE Conf. Systems, Man Cybernetics (SMC)*, 2001, pp. 747–752.

[19] S.-W. Lee and K. Mase, "Activity and location recognition using wearable sensors," *Pervasive Computing*, vol. 1, no. 3, pp. 24–32, Jul.–Sep. 2002.

[20] K. van Laerhoven and O. Cakmakci, "What shall we teach our pants?," in *Proc. 4th Int. Symp. Wearable Comput.*, 2000, pp. 77–83.

[21] P. Korpipää, M. Koskinen, J. Peltola, S.-M. Mäkelä, and T. Seppänen, "Bayesian approach to sensor-based context awareness," in *Personal Ubiquitous Comput. J.*, vol. 7, no. 4, pp. 113–124, Jul. 2003.

[22] A. Flanagan, J. Mäntyjärvi, and J. Himberg, "Unsupervised clustering of symbol strings and context recognition," in *Proc IEEE Int. Conf. Data Mining (ICDM'02)*, 2002, p. 171.

[23] A. Dey, G. D. Abowd, and D. Salber, "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications," *Human-Computer Interaction*, vol. 16, no. 2–4, pp. 97–166, 2001.

[24] N. Streitz and P. Nixon, "The disappearing computer," *Commun. ACM*, vol. 48, no. 3, pp. 32–35, Mar. 2005.

[25] J. Coutaz, J. L. Crowley, S. Dobson, and D. Garlan, "Context is key," *Commun. ACM*, vol. 48, no. 3, pp. 49–53, Mar. 2005.

[26] J. Pärkkä, M. van Gils, T. Tuomisto, I. Korhonen, and R. Lappalainen, "A wireless wellness monitor for personal weight management," in *Proc. IEEE ITAB-ITIS2000*, Arlington, VA, 2000, pp. 83–88.

[27] J. Pärkkä, L. Cluitmans, and P. Korpipää, "Palantir Context Data Collection, Annotation and PSV File Format," in *Pervasive 2004 Workshop Proc.: Benchmarks and Database Context Recognition*, 2004, pp. 9–16.

[28] J. Penttilä, J. Peltola, and T. Seppänen, "A speech/music discriminator-based audio browser with a degree of certanity measure," in *Proc. Infotech Oulu Int. Workshop Information Retrieval (IR-2001)*, 2001, pp. 125–131.

[29] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, F. Lombardi, A. Malliani, A. J. Moss, J. N. Rottman, G. Schmidt, P. J. Schwartz, and D. H. Singer, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use," *Eur. Heart J.*, vol. 17, pp. 354–381, Mar. 1996.

[30] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, MAr. 1984.

Juha Pärkkä received the M.Sc. (Tech.) degree in information technology (digital signal processing) from Tampere University of Technology in Tampere, Finland, in 1997.

Currently, he is working as a research scientist at VTT Information Technology in Tampere, Finland. His main research interests are biomedical signal processing and ubiquitous computing.

Miikka Ermes is studying for the M.Sc. degree in Tampere University of Technology. He is majoring in signal-processing methods and he is currently finishing his M.Sc. thesis concerning activity recognition.

Currently, he is working as a Research Trainee at VTT Information Technology in Tampere, Finland.

Panu Korpipää (M'05) received his M.Sc. (Tech.) degree in electrical engineering (software engineering) from the University of Oulu, Finland, in 1996.

Currently, he is working as a senior research scientist at VTT Electronics, Advanced Interactive Systems, Oulu, Finland. His professional interests include mobile and context aware computing, novel user interfaces, and visualization.

Jani Mäntyjärvi received the M.Sc. degree in biophysics and the Ph.D. degree in information processing from the University of Oulu in 1999 and 2004, respectively.

Currently, he is a senior research scientist at the Technical Research Center of Finland, VTT Electronics. His current professional interests include pervasive and context-aware computing for handheld devices, wearable sensing, and technologies for adaptive interaction.

Johannes Peltola received the M.Sc. (Tech.) degree in electrical engineering from the University of Oulu, Finland, in 1998.

Currently, he is working as a research scientist in VTT Electronics, Oulu. His research topics are multimedia signal-processing algorithms for source compression, content analysis, and joint source channel coding. He is also leading a research team in the field of these topics.

Ilkka Korhonen (M'98) received the M.Sc. and Dr.Tech. degrees in digital signal processing from Tampere University of Technology in 1991 and 1998, respectively.

Currently, he is currently working as a research professor for intuitive information technology at VTT Information Technology. He is a docent in medical informatics (with a speciality on biosignal processing) in Ragnar Granit Institute at Tampere University of Technology. His main research interests include biosignal interpretation methods and pervasive health care technologies, especially their application in critical care patient monitoring, wearable biomedical monitoring, and home health monitoring. He has published more than 60 original papers in international scientific journals and conference proceedings.