

Activity Recognition Using Biomechanical Model Based Pose Estimation

Attila Reiss, Gustaf Hendeby, Gabriele Bleser, and Didier Stricker

German Research Center for Artificial Intelligence (DFKI)
Department of Augmented Vision
Trippstadter Str. 122, 67663 Kaiserslautern, Germany

Abstract. In this paper, a novel activity recognition method based on signal-oriented and model-based features is presented. The model-based features are calculated from shoulder and elbow joint angles and torso orientation, provided by upper-body pose estimation based on a biomechanical body model. The recognition performance of signal-oriented and model-based features is compared within this paper, and the potential of improving recognition accuracy by combining the two approaches is proved: the accuracy increased by 4–6% for certain activities when adding model-based features to the signal-oriented classifier. The presented activity recognition techniques are used for recognizing 9 everyday and fitness activities, and thus can be applied for e.g., fitness applications or ‘in vivo’ monitoring of patients.

1 Introduction

With recent progress in wearable sensing it is reasonable to detect the wearer’s context or location in real life situations. Within this area of research, activity recognition has become important in recent years. The applications of activity recognition vary from industrial scenarios to fitness applications and health care.

Physical activity monitoring is not only used in therapies for various chronic diseases, but also in preventive health care. Physical inactivity is a major risk factor for diseases including diabetes, cardiovascular diseases and certain types of cancer. Wearable devices that can monitor the type and duration of activities over a patient’s day can not only be used for appropriate preventive actions, but could also motivate the patient itself and so promote a more active lifestyle. The basic daily physical activity level (30 minutes) can be achieved by everyday activities like walking or ascending/descending stairs instead of using the elevator. Performing a wider spectrum of activities, e.g., endurance-enhancing activities like Nordic walking, jogging or cycling, can help to achieve even more health benefits.

Previous work has shown that the recognition of basic physical activities, such as resting, walking, running and cycling, is possible even with just one 3D-acceleration sensor [2,3,11,12]. However, when increasing the number of activities, the classification performance drops significantly. The number of sensors must then be increased and new classification techniques introduced. A sensor

placed by the wrist is preferable when trying to distinguishing everyday or fitness activities with similar lower-body, but significantly different upper-body movement, [1,4], e.g., distinguish walking from Nordic walking [15]. Sensors placed on both upper and lower body improve the classification accuracy [1,16]. Using other sensors than acceleration or inertial sensors was also investigated in e.g., [15], but it was shown that 3D-acceleration sensors are the most powerful sensors for activity recognition.

Recently model-based methods have been introduced as an alternative to signal-oriented methods for activity recognition ([22,23]). In [22] a model-based approach is compared to signal-oriented approaches and it is shown that model-based methods have the potential to increase the recognition rate by introducing new features. However, the data set in [22] was recorded from a very specific industrial scenario: 20 well-defined activities were performed during a quality inspection of a car production process. Therefore, further analysis is needed to prove the benefits of body-model derived features, especially for everyday and fitness activity recognition (resting, walking, cycling, etc.).

For model-based activity recognition, it is fundamental to use a reliable and accurate human body pose estimation. Therefore, in this paper an upper-body pose estimator is introduced from which upper-body related features are deduced and used in the activity classifier. The estimator fuses a biomechanical model of the upper body with measurements from inertial sensors to estimate the pose of the torso and the arms. A similar approach is described in [21] where the authors use inertial sensors and a kinematic model to monitor stroke patients rehabilitation process. The papers [17,19] describe motion capturing systems using inertial sensors with support of ultrasound and magnetic sensors for capturing the body pose.

This paper presents a model-based method for recognizing 9 everyday and fitness activities, which can be used for e.g., ‘in vivo’ monitoring of patients. A model-based and a signal-oriented approach is compared, and it is shown that combining the two methods can significantly improve the recognition accuracy for different activities. Furthermore, as for deriving body-model related features, this paper goes beyond previous work: instead of just concatenating the orientation information of the inertial sensors, as presented in [22] and [23], a kinematic upper-body model is used for the upper-body pose estimation.

2 Data Collection

To obtain data, 6 Colibri inertial measurement units (IMU) from Trivisio [18] were used. The sensors are lightweight (22 g without cable) and small ($30 \times 30 \times 13$ mm). Each IMU contains a 3-axis MEMS accelerometer, a 3-axis MEMS gyroscope, and a 3-axis magneto-inductive magnetic sensor, all sampled at 100 Hz. The accelerometers have a resolution of 0.038 ms^{-2} in the range $\pm 16g$, the gyroscopes $0.37^\circ/\text{s}$ in the range $\pm 1500^\circ/\text{s}$, and the magnetometers have a scale of $\pm 1100 \mu\text{T}$ with a resolution of $0.4211 \mu\text{T}$. The inertial sensors and magnetometers are individually calibrated for offsets, scale factors and alignment errors using the method of [5] and [7], respectively. Of the 6 IMUs used for data

collection, 5 were placed on the upper body (cf. Section 4.1 for exact sensor placement) and one was shoe-mounted.

As pointed out in Section 1, data was collected from everyday and fitness activities. A criterion for selecting activities was to have activities differing mainly only in upper-body movement (like walking and Nordic walking), so that there is a high demand on good features derived from the upper-body pose estimation. Finally, data of 9 different activities were recorded: lying, sitting, standing, walking, Nordic walking, running, cycling, ascending stairs, and descending stairs. Sitting mainly consists of computer work, while standing consists of standing still or standing still and talking. Walking and running were performed both indoors and outdoors, while cycling was performed outdoors with a real bike. Nordic walking was performed outdoors on asphaltic terrain, using asphalt pads. Approximately 2 h of data was collected from the 9 different activities with one subject (male 26 years, BMI 25.66).

3 Signal-Oriented Activity Recognition

3.1 Feature Extraction

For the activity classifier that uses signal-oriented features only, acceleration data derived from the 3-axis accelerometers is used. Standard signal features were calculated over a window of 512 samples (about 5 s of data), in both time and frequency domain. Time-domain features were mean, median, standard deviation, peak acceleration and energy. For the frequency-domain features, the DC component of the signal was first removed, then the power spectral density (PSD) was calculated. Frequency-domain features were peak frequency of the PSD, power ratio of the frequency bands 0–2.75 Hz and 0–5 Hz, energy of the frequency band 0–10 Hz and spectral entropy of the normalized PSD on the frequency band 0–10 Hz.

The extracted features of the 3-axis accelerometers are computed for each axis separately, and for the 3 axes together, too. As for the window size, to receive a characteristic window, at least 2–3 s of data is necessary for certain activities, such as walking or ascending stairs. Hence, a window size of 512 samples was chosen — which also enables efficient implementation of the FFT — with a window shifting of 0.25 s.

3.2 Feature Selection

Features were selected based on both visual and statistical analysis. To compare and select features for distinguishing different groups of activities (like resting vs. non-resting activity classes), distribution bar graphs as presented in e.g., [15] were used. To identify for each activity the feature having the best performance in discriminating the corresponding activity from all the other activities, the measure presented in [8] was utilized. The K -means algorithm with $k = 100$ clusters was used for clustering the different features. The fraction for each cluster and activity was then computed, and the cluster precisions for each activity were obtained from the fractions, as presented in [8].

The cluster precisions can also be used to identify the best features to separate 2 groups of activities, or just to find the best feature to distinguish 2 different activities. Questions like “what is the best feature to separate running from all other activities including footsteps?” or “what is the best feature to distinguish lying and sitting?” can be easily answered by using the cluster precisions method only on the 2 groups of activities as a binary classifier. The selected signal features for classification are the following:

1. peak absolute value of the up-down acceleration measured on the shoe-mounted sensor (this feature is basically used to indicate footsteps)
2. mean value of the up-down (transversal) acceleration measured on the chest sensor
3. energy of the forward-backward (horizontal) acceleration measured on the shoe-mounted sensor
4. median value of up-down (transversal on initial position) acceleration measured on the right wrist sensor
5. energy of the frequency band 0–10 Hz of the up-down (transversal) acceleration measured on the chest sensor
6. standard deviation of the forward-backward (coronal on initial position) acceleration measured on the right wrist sensor
7. spectral entropy of the normalized PSD on the frequency band 0–10 Hz of the accelerations summarized for the 3 axes measured on the right wrist sensor
8. mean value of left-right (sagittal on initial position) acceleration measured on the right wrist sensor

3.3 Activity Classification

The selected features are used as feature vectors in a classifier to determine the currently performed activity. For activity classification, different approaches exist and yielded good results, such as SVMs, Naive Bayes classifiers, HMMs, ANNs or Boosting. Here the decision fell on custom decision trees, as they also have been successfully applied to activity recognition in previous work [1,2,3,4,11,15]. The advantages of decision trees include low computation requirements and a simple implementation. A custom decision tree can also be used to observe the effect of selecting different features by keeping the structure of the tree and only changing the features in the decision nodes. This method was used to compare the signal-oriented features with the model-based ones (cf. Section 5).

The custom decision tree is built using a priori knowledge about the activities, inspection of the distribution bar graphs, and the results of the cluster precisions method. The structure of the tree is depicted in Figure 1. The tree has 8 binary decision nodes and 9 leaf nodes, the latter representing the activities. The numbers in the decision nodes correspond to the selected features for each decision, as described in Section 3.2. The first decision node divides all activities into activities with and without footsteps, all other decisions are used to separate one activity from the remaining other activities. The tree has been constructed

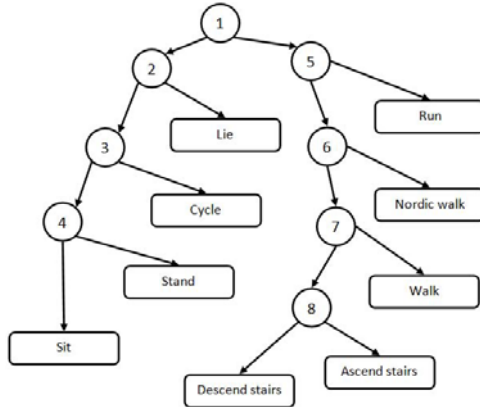


Fig. 1. Structure of the custom decision tree

so that “Sit” and “Descend stairs” are the default activities, so if the current activity shows no matching to any others while passing the decision tree, it falls through to this default classes. Classification results are shown and discussed in Section 6.

4 Upper-Body Pose Estimation

The pose and motion of the upper body are contained in the measurements from the five IMUs that are attached to it (cf. Section 2). The measured accelerations, angular velocities, and magnetic fields are compared to predictions based on a simple biomechanical body model, and this way the information is obtained with an extended Kalman filter (EKF) [10]. First the biomechanical model used for this and the sensor calibration needed are described. Then, it is explained how these are brought together to determine the pose kinematics using model based sensor fusion.

4.1 Biomechanical Body Model

A schematic of the used biomechanical upper-body model is depicted in Figure 2 (left). It consists of five rigid bodies (bones: upper arms, forearms, and torso) connected by anatomically motivated restricted joints. The right side of the figure shows how an IMU is attached to each segment on a test subject. The torso and shoulder joints are modeled with three degrees of freedom (DOF), whereas the elbow is restricted to two DOF.

Within the global reference frame, G , each individual joint constitutes its own coordinate system: the torso frame, T , the left and right shoulder frames, S , and the elbow frames, E . In the nominal pose, where all coordinate systems are aligned with the global frame, the upper body is upright, with the arms vertically

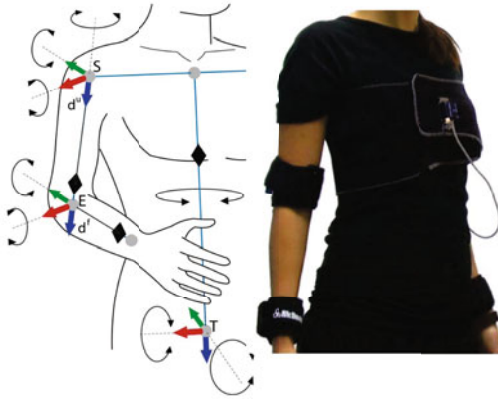


Fig. 2. Biomechanical body model and fixation of IMUs on a test subject

and the thumbs pointing forward. A coordinate system is also attached to each IMU: the torso IMU, I^t , the upper arm IMUs, I^u , and the forearm IMUs, I^f ¹.

The relative positions that are important for the upper-body pose estimation are: the positions of the elbows, d^e , and the upper arm IMUs, d^u , relative to the shoulders, and the positions of the forearm IMUs, d^f , relative to the elbows. The torso is during estimation modeled as dimensionless. Hence, both the shoulders and the center of rotation in the torso are assumed to be collocated in one point.

The procedure for calibrating the relative rotations of the different IMUs with respect to the body is described in Section 4.2. The relative positions are determined by measuring the distances along the arm segments.

The biomechanical model does not take into account the effects of soft tissue, or the non-orthogonality of the two rotation axes in the elbow. However, it has proven to be a sufficient description of the human body for the purpose of this paper.

4.2 Calibration

To be able to use the IMU measurements to estimate the body pose, the way the sensors are attached to the segments must be known. The relative sensor orientations can be calibrated in two steps. First the torso sensor is calibrated, and then the orientation of the arm sensors aligned.

The orientation of the torso sensor is obtained using two accelerometer measurements, y_1^a and y_2^a , standing in the reference pose and bending forward, respectively. The direction of the z axis in the sensor system is then given by y_1^a (the gravity). Bending forward is a rotation around x , hence given by $y_2^a \times y_1^a$. The y axis then follows in order to obtain a right handed coordinate system. The three directions specify $R_{I^t T}$.

¹ The abbreviations are used as subscripts to indicate, in which frame a quantity is resolved. For transformations, subscripts with two letters denote the mapping.

Once the torso has been calibrated, the sensors on the arms can be calibrated. This is done with a measurement in the nominal pose, and then aligning the z axes with gravity and using the magnetometer measurement of the torso as a reference direction to determine the x and y axes. This yields the rotations R_{I^uS} and R_{I^fE} .

As an alternative to using the magnetometer measurement to calibrate the arm sensors it is possible to use a two step procedure only using acceleration measurements [13]. The arms then need to be held horizontally in the second step. A drawback with this method is that there is no guarantee that the arms get aligned to each other or the torso.

4.3 Sensor Fusion

The EKF is an elegant algorithm for fusing the angular velocity, acceleration, and magnetic field measurements provided by the five IMUs. Moreover, it allows for recovering the upper-body pose based on forward kinematics only.

In this work, three separate EKFs are used: One EKF estimates the absolute torso orientation, R_{TG} , from the torso IMU measurements. It uses a standard attitude and heading reference system approach [6] as implemented in most of the commercially available IMUs [20,9].

The remaining two EKFs estimate the movements of the left and the right arm from the respective shoulder and forearm IMUs. These EKFs are loosely coupled with the torso EKF, insofar as they use its results as control input to the measurement model. The two arms are estimated independently using the same state-space model, which is detailed in the following two sections. Afterwards, the initialization of the complete EKF structure is described.

The upper body tracking assumes gravity being the only force acting on the IMUs. If gravity is not the dominant force over longer periods of time, the estimated pose may be negatively affected.

Arm system model. Knowing the segment lengths, the arm pose is fully specified by the shoulder rotation, R_{ST} , and the elbow rotation, R_{ES} . Together with the calibrated quantities (cf. Sections 4.1 and 4.2) and the orientation obtained from the torso EKF, R_{TG} , this information is sufficient to compute the IMU orientations:

$$R_{I^uG} = R_{I^uS}R_{ST}R_{TG} \quad (1a)$$

$$R_{I^fG} = R_{I^fE}R_{ES}R_{ST}R_{TG} \quad (1b)$$

and positions:

$$I_G^u = R_{TG}^T R_{ST}^T d^u \quad (2a)$$

$$I_G^f = R_{TG}^T R_{ST}^T (d^e + R_{ES}^T d^f) \quad (2b)$$

with respect to the global frame.

In order to obtain a minimal parametrization and to be able to constrain the elbow joint to two DOF, Euler angles are used to represent the configuration

of the joints. Hence, the system state, $x = [\theta, \dot{\theta}, \ddot{\theta}]$, comprises the joint angles, $\theta = [\theta_{ST}, \theta_{ES}]^T$, with $R_{ab} = \text{rot}(\theta_{ab})$, their velocities, $\dot{\theta}$, and their accelerations, $\ddot{\theta}$. Assuming constant angular acceleration, the dynamic model is:

$$\begin{bmatrix} \theta \\ \dot{\theta} \\ \ddot{\theta} \end{bmatrix}_{t+T} = \begin{bmatrix} I & T & \frac{T^2}{2} \\ 0 & I & T \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \theta \\ \dot{\theta} \\ \ddot{\theta} \end{bmatrix}_t + \begin{bmatrix} \frac{T^2}{2} \\ T \\ I \end{bmatrix} v_t \quad (3)$$

where v_t denotes zero-mean Gaussian process noise and T denotes the sample time. Variances of 0.01 rad/s^2 for the shoulder joint and 0.04 rad/s^2 for the elbow joint worked well for the motions conducted in the experiments.

Arm measurement model. The measurement models relate the measured angular velocities, y^ω , accelerations, y^a , and magnetic fields, y^m , in the local IMU frames, I , to the state.

The accelerometers measure a combination of body acceleration, \ddot{I} , and acceleration due to gravity, g , in the local IMU frame. Hence, the acceleration measurement model is:

$$y^a = R_{IG}(\ddot{I}_G(\theta, \dot{\theta}, \ddot{\theta}) - g_G) + e^a. \quad (4a)$$

Here, the body acceleration in the global frame, \ddot{I}_G , is a function of θ , $\dot{\theta}$, and $\ddot{\theta}$. It results from differentiating (2) with respect to time twice and transforming the result to the local IMU frame using R_{IG} . The latter is obtained from (1).

The gyroscopes measurement model is:

$$y^\omega = \omega_I(\theta, \dot{\theta}) + e^\omega, \quad (4b)$$

where the angular velocity in the IMU, ω_I , is obtained by transforming $\dot{\theta}$ to the local frame. The transformation can be derived from the relation $S(\omega) = (R_{IG} \dot{R}_{IG}^T)$, where $S(\omega)$ is the skew-symmetric matrix of ω [14].

When using only the inertial sensors, the rotation around the global z axis is not observable. Hence, sensor noise, and calibration and model errors cause a drift in the estimate. In order to correct for this, the magnetometers are used as aiding sensors. They provide a common forward direction, m_G . The respective measurement equation is:

$$y^m = R_{IG}(\theta)m_G + e^m. \quad (4c)$$

To simplify the equation and lessen the influence of magnetic disturbances (4c) is reduced to the heading direction. This is achieved by comparing the $\arctan(y, x)$ of both sides of the relation.

In (4), e denotes mutually independent zero-mean Gaussian measurement noise. Reflecting the actual noise of the inertial sensors, 0.002 rad/s variance is used for the gyroscopes and 0.02 m/s^2 for the accelerometers. For the heading direction the variance was set to 9° , which worked well in the experiments.

Initialization. Assuming that the subject is standing still, the initial rotation of the torso is fully determined by the acceleration and magnetic field measurements of the torso IMU. During initialization, the global heading direction is reset to the current one.

By further assuming that the rotation around the upper arm is zero, the joint angles of the arms are fully determined by the accelerometer measurements of the upper and forearm IMUs.

5 Model-Based Activity Recognition

The upper-body pose estimation presented in Section 4 provides 3-DOF torso orientation, 3-DOF shoulder joint angles and 2-DOF elbow joint angles. Features from the pose estimation are computed, selected and used for activity recognition in this section. A classifier using both signal-oriented and model-based features is described, too.

5.1 Feature Extraction and Selection

From the torso orientation and joint angles mean, median, standard deviation, and minimum and maximum value were computed as time-domain features, and energy and spectral entropy of the frequency band 0–10 Hz as frequency-domain features. These features were computed for each DOF separately and in the five joints for all dimensions together. Furthermore, features were also computed from the corresponding left and right shoulder and elbow joint angles together. Moreover, the position of the elbows and hands in the torso reference system were computed using the joint angles, and the aforementioned features extracted on the position information in all 3 dimensions. Finally, as an additional high-level feature, an energy equivalent of the two arm movements was introduced and added to the feature set. All features were computed over a window of 512 samples, as defined in Section 3.1.

The same feature selection methods were applied on the model-based features as on the signal-oriented ones, as described in Section 3.2. The structure of the custom decision tree also remained the same, only the features used in the decision nodes were changed. The selected features for the model-based classifier, numbered accordingly to the decision nodes of Figure 1, are the following:

1. energy equivalent of the arm movements
2. median value of the torso orientation in sagittal plane (bending forward-backward)
3. median value of the sagittal elbow joint angle added for the left and right arm
4. median value of the horizontal position of the right hand
5. mean value of the sagittal elbow joint angle added for the left and right arm
6. spectral entropy of the normalized PSD on the frequency band 0–10 Hz of the coronal torso orientation (bending left-right)

7. energy equivalent of the arm movements
8. standard deviation of the sagittal elbow joint angle added for the left and right arm.

5.2 Combining Signal-Oriented and Model-Based Features

When analyzing the selected model-based features, it is reasonable that for certain decision nodes they can improve the performance on separating the two groups of activities. For instance, the model-based feature of the first decision node uses a completely different characteristic of the activities to be separated: instead of using the information whether footsteps were detected or not, it uses the fact, that significant movement of the arms is present while performing walking related activities. Although the accuracy of the separation decreases from 99.21% to 98.29% compared to when using the signal-oriented feature, when using the two features together, an improvement of the classification performance is expected.

The combined classifier uses the same structure for the custom decision tree as the classifiers of Section 3.3 and Section 5.1. However, instead of just training for one, either signal-oriented or model-based feature, the decision in e.g., the first node is as following: use the model-based feature (the energy equivalent of the arm movements) if its value is under a trained limit and classify the sample for the group of the activities lying, sitting, standing, and cycling, otherwise use the signal-oriented feature (in this case the peak absolute value of the up-down acceleration measured on the shoe-mounted sensor) for the classification of the current sample. Section 6 shows the results of the combined classifier, and discusses for which decision nodes the incorporation of the model-based features improved the classification.

6 Results and Discussion

Table 1 shows the confusion matrix of the signal-oriented classifier. The overall performance of the classifier is 96.08%. For certain activities, such as lying or running, this classifier already provides a highly reliable classification. Also, the separation of cycling from the resting activities is accurate by using the shoe-mounted sensor data. However, more than 7% of the cycling-samples wrongly indicate footsteps, which lead this samples to the right side (cf. Figure 1) of the decision tree and thus resulting a performance of only 92.61% for the activity cycling.

The confusion matrix also shows some misclassifications between sitting and standing, and between ascending and descending stairs, where improvement is expected by using model-based features in addition. Ascending and descending stairs also introduced difficulties in the classification of walking. The selected signal feature for discriminating walking uses the fact, that is usually a more regular movement than going upstairs or downstairs. However, the detection accuracy could benefit from introducing model-based features.

Table 1. Confusion matrix of the signal-oriented activity classification

Annotation	Recognized activity								
	Lie	Sit	Stand	Walk	Nordic walk	Run	Cycle	Ascend stairs	Descend stairs
Lie	3106	0	0	0	0	0	0	0	0
Sit	0	4235	18	0	0	0	0	0	0
Stand	0	42	3007	0	0	0	0	0	0
Walk	0	0	0	3402	0	0	0	97	103
Nordic walk	0	0	0	0	1364	0	0	244	0
Run	0	0	0	0	0	2304	0	0	0
Cycle	0	0	0	0	1	0	1943	0	154
Ascend stairs	0	0	0	103	93	0	28	1531	0
Descend stairs	0	0	0	4	0	0	0	18	1299

Table 2. Confusion matrix of the model-based activity classification

Annotation	Recognized activity								
	Lie	Sit	Stand	Walk	Nordic walk	Run	Cycle	Ascend stairs	Descend stairs
Lie	2977	0	0	0	4	3	122	0	0
Sit	32	3874	32	0	0	0	315	0	0
Stand	0	42	2971	0	1	0	28	5	2
Walk	0	0	0	2873	407	0	0	147	175
Nordic walk	0	0	0	286	1321	0	0	0	1
Run	0	0	0	0	0	2304	0	0	0
Cycle	0	194	25	55	9	38	1687	0	90
Ascend stairs	0	0	0	284	64	46	0	1361	0
Descend stairs	0	0	189	142	224	0	0	0	766

Nordic walking was detected with 84.83% and walking with 94.45%. For discriminating Nordic walking from the remaining activities, the increased arm motion in the forward-backward direction was used. However, the characteristic of ascending stairs overlap in this respect, which implies the need for a better performing feature. Interestingly, no misclassifications were detected between walking and Nordic walking, what conforms to the requirement of a significantly increased arm movement while performing Nordic walking compared to normal walking.

The confusion matrix of the model-based classifier is shown in Table 2. The overall performance of the classifier is 87.18%. Before further analyzing the detection accuracy, it has to be mentioned, that the upper-body pose estimation failed on one part of the collected data, mainly locomotion activities performed indoors were affected. The magnetometers of the IMUs are sensitive to magnetic disturbances, which explains this inaccurate pose estimation. However, even for the indoor activities, the estimated torso orientation and joint angles were reliable in the majority of the samples, and thus useful for activity classification. Another limitation of the pose estimation method described in Section 4 is, that the used state-space model assumes that the upper body is not subjected to severe body accelerations. However, this barely effected the pose estimation on the collected data.

The detection accuracy of walking was 79.76% and for Nordic walking 82.15%. Most of the misclassified samples can be explained by the above mentioned inaccurate pose estimation while performing walking indoors. Interestingly, the best feature to discriminate Nordic walking (the feature of decision node number 6) was found the spectral entropy of the left-right torso orientation. This conforms to the requirement of an increased torso movement while performing Nordic walking, which results in a regular left-right bending of the torso, thus a low entropy of the the PSD.

Table 3. Confusion matrix of the combined activity classification

Annotation	Recognized activity								
	Lie	Sit	Stand	Nordic Walk	walk	Run	Cycle	Ascend stairs	Descend stairs
Lie	3106	0	0	0	0	0	0	0	0
Sit	0	4235	18	0	0	0	0	0	0
Stand	0	42	3007	0	0	0	0	0	0
Walk	0	0	0	3402	0	0	0	79	121
Nordic walk	0	0	0	0	1364	0	0	8	236
Run	0	0	0	0	0	2304	0	0	0
Cycle	0	0	0	0	0	0	2065	0	33
Ascend stairs	0	0	0	78	46	0	28	1603	0
Descend stairs	0	0	0	4	0	0	0	0	1317

Difficulties were encountered when distinguishing cycling and sitting, resulting in a detection accuracy of 80.41% and 91.09%, respectively. These difficulties are understandable, since the upper-body pose of these two activities is similar, and thus clearly highlight the need of a sensor placed on the lower body for recognizing everyday and fitness activities, in contrast to the upper-body dominant activities of the industrial scenario presented in [22,23].

The confusion matrix of the combined classifier is shown in Table 3. The overall performance of the classifier is 97.00%. For this classifier, the reliability of the pose estimation was first tested for each sample, due to the problems encountered for locomotion indoor activities. This investigation relies on the elbows' and hands' position, verifying first if the current positions are anatomically possible, and for each decision node if they are in the expected area. If the pose estimation turns out to be not reliable, only signal-oriented features are used to classify that sample.

Three decision nodes of the custom decision tree of the combined classifier only use signal-oriented features, the other five decision nodes benefit from model-based features. The former three nodes are, where lying, cycling and running are discriminated from remaining activities. As mentioned before, for discriminating cycling from resting activities (especially from sitting), a feature calculated on data of a lower body sensor is preferable. As for running, both the selected signal-oriented and the model-based feature performed highly reliable, no improvement was needed. Since the signal-oriented feature is applicable under all circumstances, even indoor when high magnetic disturbances are present, the combined classifier uses this feature.

For distinguishing standing and sitting, both the signal-oriented and the model-based features are used, but no improvements were achieved in the recognition accuracy as one can see by comparing the corresponding rows of Table 1 and Table 3.

Similarly, both features were used when distinguishing ascending and descending stairs, which solved the misclassifications compared to the signal-oriented approach, and thus increased the recognition accuracy of descending stairs from 98.33% to 99.70%.

Even more improvement on the classification results can be observed when incorporating model-based features for the decision nodes number 6 and 7. This results in less misclassification of ascending stairs into Nordic walking and walking, thus increasing the classification accuracy of ascending stairs from 87.24% to 91.34%.

The most significant improvement in recognition accuracy was observed when combining the footstep detection (signal-oriented feature) with the energy equivalent of the arm movements (model-based feature) to separate the activities in the first decision node. This led to a nearly perfect separation in this node, since resolving most of the misclassifications of the cycling-samples. The recognition accuracy of cycling was thus increased by nearly 6%, from 92.61% to 98.43%.

7 Conclusion and Future Work

In this paper, a model-based activity recognition method was introduced and applied to recognizing everyday and fitness activities. Model-based and signal-oriented approaches were compared and the benefits of body-model derived features for activity classification were proved. It was also shown, that the recognition accuracy can be significantly improved when adding model-based features to a signal-oriented classifier. The overall classification performance of the presented combined classifier showed an improvement of approximately 1% compared to the signal-oriented classifier. For certain activities the classification results improved even more significantly, with up to 6%. The presented results can be regarded as a proof-of-concept of improving activity recognition accuracy when incorporating body-model features. For the future, data collection over a longer period, and with multiple test subjects is planned. It also has to be mentioned, that the custom decision tree used as classifier does not necessarily provide the most suitable classifier for a given set of features, thus in future work, the solution space also should be exhaustively explored.

The model-based features were calculated from shoulder and elbow joint angles and torso orientation, provided by a biomechanical model based upper-body pose estimation. This pose estimation turned out to be reliable in most cases, and the classifier based only on this estimator showed a high recognition accuracy of over 87%. However, data collected while performing indoor activities also showed some limitations to the applicability of this approach, due to the sensitivity of the magnetometers to magnetic disturbances caused by ferrous material in the environment. Hence, if mainly indoor activities should be classified, hard and soft iron effects resulting from such disturbances need to be handled in the model.

Acknowledgements. This work has been performed within the project PAMAP funded under the AAL Joint Programme (AAL-2008-1). The authors would like to thank the project partners and the EU and national funding authorities for the financial support. For more information, please visit the website www.pamap.org

References

1. Bao, L., Intille, S.: Activity recognition from user-annotated acceleration data. In: Proc. 2nd Int. Conf. Pervasive Comput., pp. 1–17 (2004)
2. Bieber, G., Peter, C.: Using physical activity for user behavior analysis. In: PETRA 2008 (2008)

3. Ermes, M., Pärkkä, J., Cluitmans, L.: Advancing from offline to online activity recognition with wearable sensors. In: 30th Annual International IEEE EMBS Conference, pp. 4451–4454 (2008)
4. Ermes, M., Pärkkä, J., Mäntyjärvi, J., Korhonen, I.: Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Trans. Inf. Technol. Biomed.* 12(1), 20–26 (2008)
5. Ferraris, F., Grimaldi, U., Parvis, M.: Procedure for Effortless In-Field Calibration of Three-Axis Rate Gyros and Accelerometers. *Sensor and Materials* 7, 311–330 (1995)
6. Harada, T., Mori, T., Sato, T.: Development of a Tiny Orientation Estimation Device to Operate under Motion and Magnetic Disturbance. *The International Journal of Robotics Research* 26, 547–559 (2007)
7. Hu, X., Liu, Y., Wang, Y., Hu, Y., Yan, D.: Autocalibration of an Electronic Compass for Augmented Reality. In: ISMAR (2005)
8. Huynh, T., Schiele, B.: Analyzing features for activity recognition. In: sOc-EUSAI 2005, pp. 159–163 (2005)
9. InterSense, <http://www.intersense.com>
10. Jazwinski, A.H.: Stochastic Processes and Filtering Theory. Mathematics in Science and Engineering, vol. 64. Academic Press, Inc., London (1970)
11. Karantonis, D., Narayanan, M., Mathie, M., Lovell, N., Celler, B.: Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Trans. Inf. Technol. Biomed.* 10(1), 156–167 (2006)
12. Long, X., Yin, B., Aarts, R.M.: Single-accelerometer based daily physical activity classification. In: 31st Annual International IEEE EMBS Conference, pp. 6107–6110 (2009)
13. Luinge, H.J., Veltink, P.H., Baten, C.T.M.: Ambulatory measurement of arm orientation. *Journal of Biomechanics* (40), 78–85 (2007)
14. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S.: An Invitation to 3-D Vision. Springer, Heidelberg (2003)
15. Pärkkä, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J., Korhonen, I.: Activity classification using realistic data from wearable sensors. *IEEE Trans. Inf. Technol. Biomed.* 10(1), 119–128 (2006)
16. Pirttikangas, S., Fujinami, K., Nakajima, T.: Feature selection and activity recognition from wearable sensors. In: Youn, H.Y., Kim, M., Morikawa, H. (eds.) UCS 2006. LNCS, vol. 4239, pp. 516–527. Springer, Heidelberg (2006)
17. Schepers, H.M., Roetenberg, D., Veltink, P.H.: Ambulatory human motion tracking by fusion of inertial and magnetic sensing with adaptive actuation. *Medical and Biological Engineering and Computing* (48), 27–37 (2010)
18. Trivisio, <http://www.trivisio.com>
19. Vlastic, D., Adelsberger, R., Vannucci, G., Barnwell, J., Gross, M., Matusik, W., Popovic, J.: Practical Motion Capture in Everyday Surroundings. In: SIGGRAPH (2007)
20. XSens, <http://www.xsens.com>
21. Zhou, H., Hu, H.: Inertial sensors for motion detection of human upper limbs. *Sensor Review* 27(8), 151–158 (2007)
22. Zinnen, A., Blanke, U., Schiele, B.: An Analysis of Sensor-Oriented vs. Model-Based Activity Recognition. ISWC, pp. 93–100 (2009)
23. Zinnen, A., Wojek, C., Schiele, B.: Multi activity recognition based on bodymodel-derived primitives. LoCA, pp. 1–18 (2009)