An Analysis of Sensor-Oriented vs. Model-Based Activity Recognition

Andreas Zinnen[1,2] Ulf Blanke[2] Bernt Schiele[2]

[1]SAP Research

zinnen@mis.tu-darmstadt.de

[2]Computer Science Department
TU Darmstadt
{*blanke,schiele*}*@cs.tu-darmstadt.de*

Abstract

Model-based activity recognition has been recently proposed as an alternative to signal-oriented recognition. Such model-based approaches seem attractive due to their ability to enable user-independent activity recognition and due to their improved robustness to signal-variation. The first goal of this paper is therefore to systematically analyze the benefit of body-model derived primitives in different sensor settings for multi activity recognition. Furthermore we propose a new body-model based approach using accelerometer sensors only thereby reducing the sensor requirements significantly. Results on a 20 activity dataset indicate that body-model based approaches consistently improve results over signal-oriented approaches.

1. Introduction

Context awareness is a central issue in wearable computing. While context information may consist of any information characterizing the user's situation, activity recognition using body-worn sensors is of particular importance for a diverse range of areas such as industrial applications and medical care. As a consequence many different activity recognition approaches have been proposed. Despite the continuous research efforts, the field does not seem mature enough to allow activity recognition to be commonly used in real-world settings or even real-world applications. To this end the primary goal of this paper is to contribute a systematic and in-depth analysis and comparison of model-based with sensor-oriented activity recognition considering different types and number of sensors.

Interestingly, most activity recognition methods using body-worn sensors follow a similar two-stage pattern. In the first stage signal-oriented features (such as FFT-coefficients or mean/variance [14]) are extracted from the continuous data stream. In the second stage a generative model (such as HMMs [13]), a discriminative classifier (such as AdaBoost [5]) or a hybrid generative-discriminant model [14] is trained and used for classification. These types of approaches (called *signal-oriented methods* in the following) have achieved good performance on various datasets. Recently [27] proposed a different type of approach (which we call *model-based method*). The authors first estimate a body-model from five inertial measurement units (IMU) providing their orientations in a global reference frame and then use the human body-model to derive high-level primitives such as moving the arms up or down or turning the wrist. On a 20-activity dataset [20] the authors report good performance and while [20] recognizes activities in a user-dependent manner, [27] achieves similar results in a user-independent setting.

The first goal of this paper is therefore to answer the question if such model-based methods have indeed potential to advance the state-of-the-art in activity recognition. For this we systematically compare the model-based approach of [27] to more traditional signal-oriented approaches. The experimental results in this paper indeed indicate that model-based approaches enable more robust activity recognition than signal-oriented approaches and that their combination can further improve recognition performance. As the original approach [27] requires the use of five relatively expensive and power-hungry IMUs we extend the approach to reduce sensor requirements. To this end we propose an alternative model-based approach that does not require the use of IMUs but instead uses accelerometer sensors only. We explore the possibility to reduce the number of sensors required to merely two sensors attached to the wrists of the human. Last but not least we also analyze the importance of location information for activity recognition.

Section 2 summarizes related work, Section 3 describes the evaluated features and Section 4 introduces Joint Boosting to classify multiple activities. Section 5 describes the dataset and the evaluation procedure and Section 6 summarizes the experimental results. Finally, the main contributions of the paper are discussed in Section 7.

2. Related Work

Current research covers a wide range of approaches to activity recognition. Different types and combinations of

sensors are used [14, 18]. In particular, inertial sensors such as accelerometers, gyroscopes and magnetometers are popular [3, 14, 20, 26]. While acceleration sensors attached to the user's wrist seem sufficient to allow recognition of long term activities [9, 11], inertial measurement units (IMU) like XSens [2] have been successfully employed to spot short activities amid background data [6, 8, 20, 24]. Also the number of sensors varies between approaches. With one or two sensors good results can be obtained for activities like jogging or walking and long term activities like shopping or working [16, 9, 10, 7]. A larger number of sensors is required to recognize more complex activities like workshop or maintenance activities [20, 27].

In addition to accelerometer and inertial sensors, location sensors (e.g., Ubisense [1]) have proven helpful for activity recognition [12, 15, 20]. Since activities are often executed at different locations, location information can help to increase the detection rate significantly. In exchange, a complex infrastructure has to be installed and maintained, especially for indoor positioning. Often, users must wear additional sensors to determine their position.

3. Segmentation, Body-Model and Features

This section introduces *signal-oriented* features as well as *body-model derived* primitives used in our evaluation in Section 5. Based on [27] Section 3.1 briefly describes how a human body-model can be estimated from five IMUs, which primitives are derived from this model and also describes the segmentation procedure used throughout the paper. To reduce the sensor requirements of [27] Section 3.2 introduces and discusses a novel method to estimate a human body-model based on acceleration sensors only. Section 3.3 briefly summarizes features based on a Ubisense location environment and Section 3.4 recapitulates common signal oriented features as used, for example by [14].

3.1. Body-Model by IMUs (BM_IMU)

To derive features or primitives like moving the hands up, turning the arm, or keeping the arm in a specific posture, [27] introduces a 3D human body-model (BM_IMU). For this the orientation information of five IMUs (Inertial Measurement Units) located at the user's upper and lower arms and the torso is concatenated to a kinematic chain. Figure 1 illustrates snapshots of the resulting 3D body-model (depicted as stick figures) while closing the hood of a car.

With this upper body-model, the authors segment a continuous data stream using short but fixed positions of the hands and turning points of hand movements. Since the variance over the hand positions is lower for fixed positions and turning points, local minima within the variance of the hand positions can be detected separately for both hands. The lower picture shows the variance of the user's right

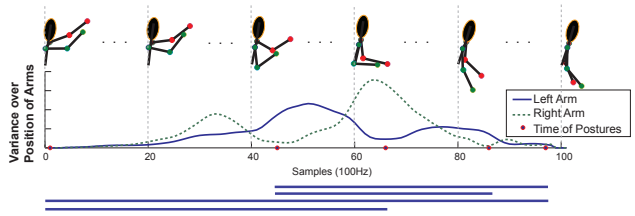

Figure 1. Body-model (top) and variance of hand positions (bottom) including the segmentation while closing a hood based on 5 IMUs.

(dashed green) and left (blue) hand positions in the course of the activity. Local minima are illustrated as red circles on the x-axis. Pairwise combinations of those minima can be used to extract segments of a specific minimum and maximum length (see blue lines below the figure).

In the following, the feature calculation as well as the activity classification is done for all segments of interest.

In this section the features characterizing sub-actions and postures with motion and posture primitives are presented as proposed in [27]. Based on BM_IMU, several data streams for each point in time are calculated including the hands' height, the arms' twist, the torso's bending or twisting, the distance between the left and right hand as well as the arms' orientation towards gravity. Furthermore, the torso's global orientation as well as the angles included by the connection lines of the torso to the hands and elbows can be estimated. This data provides the basis for the subsequent extraction of motion and posture primitives.

Motion Primitives Motion primitives are defined to characterize basic arm movements like up or down *(height primitives)*, push and pull *(push-pull primitives)*, the bending of the body forward and back *(bending primitives)* and arm twisting *(twist primitives)*. The following exemplary summarizes the extraction for height primitives (for details see [27]). Applying a sequential minimum-maximum search on the arms' height, a segment is divided into areas of up and down movements. To calculate temporal features over primitives, the segment is divided into 20 equally spaced bins. Each bin is assigned the height difference of the associated height primitive (20 features). Furthermore, the number, average, maximum and minimum of up and down primitives for the analyzed segment (8 features) are added. Finally, a histogram of the primitives length, normalized by the segments overall length, is included (5 features).

Push-pull primitives are based on the torso's twisting as well as the angles included by the connection lines of the torso to the hands and elbows. Bending primitives and twist primitives consider the torso's bending and the arms' twist in a similar manner as height primitives. Overall, 66 height features for both hands, 165 push-pull features (66 hands, 66 elbows, 33 torso), 33 bending features and 66 twist features are included in the subsequent training and classifica-

tion steps.

Additionally, histograms on the movement direction of both hands in a global reference system (direction histograms) have proven to be valuable features. Therefore, the hands' data points of the 3D body-model are projected onto the plane defined by the gravity vector as the normal. Next, directional vectors of succeeding hand positions are calculated and assigned to eight bins of a histogram represented by eight equally distributed directions (8 features).

Posture Primitives In addition to motion primitives, [27] consider postures that help to distinguish between activities and the background. Postures based on following dimensions are included in the feature set: the arms' orientation towards gravity (6 dimensions), the distance between the two hands (1 dimension) and the hands' height (2 dimensions), as well as the torso's direction in a global reference system (2 dimensions). For each dimension, the minimum, maximum, mean and variance of the postures (44 features) are added to the feature set.

In total we get a feature dimension of 382 on BM_IMU. As motivated before, we explore the possibility to reduce the number of sensors required to merely two sensors attached to the human wrists. When calculating feature vectors for an evaluation of this setting (*Reducing Number of Sensors*), not all features are calculated. Whereas features on height primitives and rotation features as well as postures can be estimated considering only the wrist sensors, all other primitives (push-pull, bending, and twist) cannot be detected. Furthermore, the distance between the two hands as well as the torso's direction cannot be estimated. Instead, the average orientation of the two wrist sensors approximates the torso's direction when using two sensors only.

3.2. Body-Model by Acceleration (BM_ACC)

As previously shown [27] a body-model can be used to extract abstract features to improve recognition results. Having access to precise data from inertial measurement units (IMUs), respectively to the global orientation of the sensor, it is straightforward to determine the angles of joints and hereby the configuration of the body. However, IMUs come with the price of power, are harder to embed into wearable items and are still expensive. While the results are less accurate than using IMUs, accelerometer-based approaches also allow to estimate the sensor's orientation [19] and have been used to create low cost and power efficient motion capture systems [4, 22, 21]. The following introduces a novel method to estimate a human body-model from acceleration sensors only and then briefly discusses difficulties and drawbacks compared to the BM_IMU. Section 5 analyzes the impact of this novel model on recognition performance.

We use the effect of the earth gravity vector G on the 3D acceleration values as reference to estimate the orientation of each sensor individually. Additional acceleration caused by human movement obviously influence the estimate of the direction of G. To reduce this dynamic motion component we smooth the signal by calculating the mean of the acceleration on a sliding window. A window length of 120ms turned out to be a suitable value. The direction of the normalized acceleration vector is taken as estimate of the earth gravity vector. This vector is then used as the sensor orientation with respect to the ground plane.

Some of the remaining ambiguities can be resolved by adding simple constraints, before creating the kinematic chain of the human body. The constraints are described in the following and disallow unnatural poses.

Upper Arm When using acceleration only to estimate the orientation of a limb, unnatural postures of the upper arm can occur. Figure 2 (a) shows the upper body from the front view. The gray-colored hemispheres approximate the space of natural postures of the upper arms. An example of an unnatural pose is illustrated as dashed line. Dynamic motion affects the estimate of the orientation around G and might lead to those postures. Also certain sequences of arm movements might cause these poses. To disallow these unnatural poses we simply mirror the arm vertically back into a hemisphere as shown in Figure 2 (a) .

Lower Arm Bending Figure 2 (b) shows the right view of the upper body and the allowed lower arm postures within the rectangle. We obtain the lower arm's direction toward gravity by observing the axis (blue bigger arrow) along the elbow. If the arm is directed towards the ground, the full earth gravity affects the sensor. The higher the arm moves, the less the sensor is affected by gravity. We constrain the lower arm to bend to the front perpendicular to the shoulder from 0 to 180 degrees. That way we do not get unnatural poses illustrated by the dashed line.

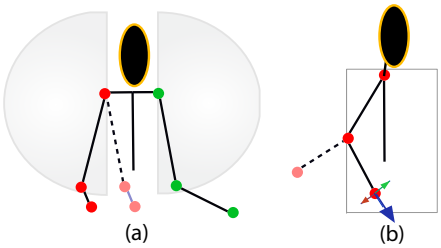

Figure 2. (a) The upper arm rotational space. (b) The lower arm constraint: The rotational space of the lower arm around is constrained to the front.

Lower Arm Rotation To represent the rotation along the elbow, we use the two axes perpendicular to the lower arm. This works best on the ground plane and is less precise the closer the rotation occurs around the axis of gravity due to the fact that rotation around G does not change the acceleration.

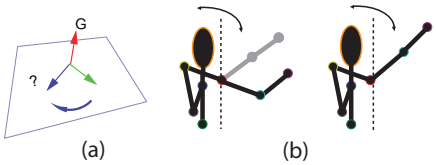

Figure 3. (a) Rotation around the gravity vector G leaves vectors on the ground plane undetermined. (b) As consequence the push-pull-primitive (left) using the BM_ACC on horizontal plane cannot be determined (right).

After applying the constraints we concatenate the obtained orientation vectors as in Section 3.1, starting from the torso to the hand, to estimate the body configuration. As a result, using acceleration as basis we can achieve a similar kinematic chain as using IMUs. One drawback using this approach is its limitation to activities with large and fast movement. Fast motion strongly influences the acceleration values and hereby the estimate of the gravity vector.

Lacking the rotational information around G, we introduce a systematic error. Figure 3 (a) illustrates a rotation in the plane. As can be seen rotating on a plane perpendicular to G does not change the gravity field, i.e., it does not affect the acceleration. Thus, the orientation of the sensor in the plane cannot be precisely estimated. This affects the body-model as follows. Figure 3 (b) illustrates on the left the push and pull primitive in the plane. This primitive is characterised a by vertical rotation around the shoulder and a counter rotation at the elbow. The right side of Figure 3 (b) shows the result of our model which hardly reveals any vertical rotation around the shoulder, if the upper arm is aligned to the horizontal plane.

In this section the BM_ACC was described. To enable direct comparison of this approach with BM_IMU we use the same motion and posture primitives as described in Section 3.1.

3.3. Location Features

For many tasks, the user's location while performing the current activity can be a powerful cue. The location system Ubisense [1] provides continuous 3D coordinates with an accuracy up to 15cm. As the activities in our scenario do not vary significantly in height, we use the 2D position on the ground plane only. For each segment the means of the (x,y) co-ordinates are calculated. The two location features are considered when evaluating the importance of location information for activity recognition (*Adding Location*).

3.4. Signal Features

An evaluation of body-model features compared to signal-oriented features was motivated in Section 1. Whereas Sections 3.1 and 3.2 introduce features based on two body-models, we briefly explain common signal-oriented features in the following. Most research on activity recognition successfully base their feature calculation directly on a sensor signal [3, 24, 9]. Typically, the features are calculated separately for each dimension. Using 5 IMUs, features for 45 dimensions are calculated (3D orientation, 3D acceleration, 3D gyroscope per sensor). Using 5 acceleration sensors, only 15 dimensions will be considered. Note that for the evaluation *Benefit of Location*, two more dimensions will be added. In case of *Reducing Number of Sensors*, the dimensions will decrease.

For each dimension, for example, a sensor's first acceleration dimension, a set of features in frequency and time domain is calculated. First, the Fast Fourier Transform (FFT) maps the incoming signal into the frequency spectrum. We group the FFT-coefficients into five logarithmic bands (5 features). The resulting absolute and real values are added to the feature set. Furthermore, we calculate the cumulative energy of the Fourier series. In addition, 10 cepstral coefficients are calculated modeling the spectral energy distribution. The spectral entropy (1 feature), which gives a cue about the complexity of the signal is also added. As features in the time domain we calculate the mean and variance of the signal (2 features). Hence we obtain 19 features per dimension. Given 5 IMUs and the additional 2D location, the feature calculation yields a total number of 865 features.

4. Multi-Activity Recognition using Joint Boosting

This section describes a method [27] using Joint Boosting [23] to classify activities based on the features introduced in the previous section.

Boosting [5] is a widely used discriminative machine learning technique for classification. In each boosting round an optimal weak classifier is trained, often based on a single feature, and the overall boosting algorithm combines these weak classifiers into a final strong classifier for binary categorization. For multi-class categorization, Torralba et al. [23] extend the idea of boosting looking for common weak classifiers and features that can be shared across multiple categories. In our case groups of similar activities are separated during the initial rounds and are disambiguated among each other in later boosting rounds. Along the way, Joint Boosting reduces the computational complexity by finding common features that are shared across several classes. Input for the training and test phase of the boosting are features on segments as described above.

In the dataset used below activities are hand annotated and all features can directly be calculated on the positive training segments. Boosting uses directly the features as described in Section 3. The respective feature vectors are then used to train Joint Boosting. Calculation of feature vectors for the negative training and test data is as follows. As dis-

cussed in Section 3.1, segments of a specific minimum and maximum length are extracted using short but fixed positions of the hands and turning points of hand movements. For all resulting segments, the features are calculated and they are either used as negative training samples for Joint Boosting or classified in the test case.

5. Dataset and Evaluation

To evaluate the features and the algorithm presented in the previous section we have chosen a quality control scenario from a car production process. The following describes the publicly available dataset [20] and the evaluation procedure.

Dataset The dataset contains 20 activities that are performed during a typical car quality check. Example activities are checking gaps of the car's frame or inspecting movable parts, for example, by opening and closing doors. See Table 1 for a complete list of the 20 activities. Besides displaying a high variability of motion patterns, most of them are short activities with an average duration of 3s per activity. As a result the ratio of activity versus non-activity data is $1 : 135$ for each activity making the spotting of the activities a challenging task in this dataset.

The dataset was recorded within the scope of an industrial project focusing on wearable technology in production environments [17]. During the experiments, data of a wearable system composed of seven motion sensors and four ultra-wide band [1] tags for tracking user position were collected. The sensors are located at the wearer's torso, the upper and lower arm and the hand. In this paper we use either five sensors (leaving out the sensors mounted on the hands) or just two wrist-worn sensors. In total 12h of data were recorded. The activities were performed by 8 subjects, each of them repeating the procedure about 10 times on average.

Evaluation Procedure We evaluate the distinctiveness of each activity individually with respect to the collected background data and the remaining activities. As a result the number of false positives increases while increasing the recall for an activity. We perform a leave-one-user-out cross-validation to enable user-independent activity recognition. In each cross-validation round, we calculate the probability for all detected segments. A segment T will be counted as true positive if the ground truth segment A has the same activity label and if the following Equation 1 holds true:

$$start(A) \leq center(T) \leq stop(A) \tag{1}$$

with $start(A)$ and $stop(A)$ correspond to the begin and end times of the ground truth segment A and $center(T)$ indicates the central time of segment T. This ensures that the analyzed activities are spotted at the right time location. Only if the central time of a segment intersects with the annotated activity, the segment is counted as a true positive. Ideally both precision and recall are 100%. Typically however, the precision decreases when increasing the recall for a particular activity. For brevity we use a single point of the precision-recall curve namely the commonly used equal error rate (EER) where recall and precision are equal. Additionally, we report the mean equal-error-rate for each setting in our evaluation.

As mentioned earlier the main focus of this paper is a systematic evaluation of different aspects and their impact on activity recognition performance. In order to make the results as comparable as possible we use the same segmentation procedure introduced above (Section 3.1) as a prefilter of all algorithms. Please note that the remaining segments contain all annotated activities so that all algorithms can obtain 100% recall. Although other segmentation procedures based on two sensors (acceleration or IMUs) [26] exist, the same segmentation procedure for all algorithms is used to make the results directly comparable for the purpose of this paper.

6. Experimental Results

As motivated before, this paper contributes a systematic evaluation of various aspects of activity recognition algorithms. The presentation is structured in three parts. The first (*Benefit of Body-Model*) compares results from body-model based features to signal-based features. Here we consider two body-models (BM_IMU and BM_ACC) that are either based on IMUs or acceleration sensors only. Results of incorporating location (*Benefit of Location*) are presented in the second part and the third part reports on the results of reducing the number of sensors (*Reducing Number of Sensors*).

Benefit of a Body Model Section 3 introduced two body-models either using five IMUs or five acceleration sensors. The top three rows in Table 1 contain a comparison of the algorithm for BM_IMU derived features with signal-oriented features for the IMU sensors. On average, BM_IMU performs better on body-models with an EER of 0.92 than signal-based features with 0.88. Only on four activities, the signal-oriented approach is marginally better. The best result is achieved by combining both approaches which improves the average EER to 0.93. In combination 18 activities are recognized better than using signal-oriented features only. Solely for one activity (*check trunk gaps*), the signal-oriented features perform marginally better.

Row four to six in Table 1 show results using the BM_ACC and signal-based features on acceleration sensors only. An EER of 0.57 is obtained using signal oriented features. The BM_ACC outperforms the former with 0.61. Again the combination of the two types of features performs best with an EER of 0.64. For five out of twenty activities, the signal-based approach is slightly better.

		Average EER	open hood	close hood	open trunk	check trunk	close trunk	fuel lid	open left door	close left door	open right door	close right door	open two doors	close two doors	mirror	check trunk gaps	lock check left	lock check right	check hood gaps	open swl	close swl	writing
w/o location	BM_IMU + signal, inertial	**.93**	.99	1.00	.92	.97	.92	.96	.87	.92	.94	.88	.94	.82	1.00	.97	.92	.89	.92	.97	.94	.95
	BM_IMU	**.92**	.99	.99	.91	.99	.92	.94	.86	.89	.94	.87	.88	.77	1.00	.96	.89	.86	.92	.97	.91	.92
	signal, inertial	**.88**	.95	.97	.86	.91	.82	.92	.74	.85	.86	.88	.87	.76	.99	.99	.88	.90	.77	.95	.90	.93
	BM_ACC + signal, acceleration	**.64**	.65	.59	.72	.79	.59	.63	.44	.44	.44	.38	.65	.30	.95	.82	.71	.69	.71	.79	.62	.94
	BM_ACC	**.61**	.78	.47	.64	.77	.62	.68	.43	.35	.40	.35	.62	.28	.88	.81	.69	.65	.56	.65	.58	.93
	signal, acceleration	**.57**	.56	.44	.46	.86	.59	.71	.32	.44	.28	.31	.62	.24	.88	.82	.78	.70	.72	.44	.27	.91
with location	BM_IMU + signal, inertial	**.92**	.97	.96	.94	1.00	.94	.96	.85	.89	.92	.83	.92	.76	.99	.97	.88	.87	.93	.97	.94	.95
	BM_IMU	**.92**	.99	.99	.94	1.00	.94	.94	.86	.90	.92	.81	.90	.81	.97	.99	.90	.83	.95	.96	.91	.94
	signal, inertial	**.90**	.96	1.00	.90	.95	.82	.91	.83	.88	.90	.90	.88	.75	.99	.95	.84	.82	.92	.94	.91	.94
	BM_ACC + signal, acceleration	**.81**	.94	.97	.87	1.00	.88	.90	.71	.60	.60	.50	.76	.54	.97	.90	.78	.79	.89	.92	.79	.95
	BM_ACC	**.81**	.96	.99	.86	.99	.88	.83	.73	.58	.62	.46	.77	.56	.99	.91	.78	.71	.88	.91	.74	.95
	signal, acceleration	**.71**	.85	.72	.81	.99	.68	.86	.51	.55	.46	.42	.65	.40	.99	.90	.80	.78	.88	.54	.46	.95

Table 1. Top group rows show EER for activity recognition without using location. The bottom group incorporates location. For each case we use body-model, signal-oriented or their combination either based on IMU-sensor data or on acceleration

From the results we can conclude that using body-models does indeed improve results by about 4% in both cases with respect to signal-oriented features alone. Combining sensor-oriented features with the body-model features further increases performance. A larger difference in performance however is observed between the precise IMUs and the acceleration only approaches, where the performance drops substantially (from 0.93 to 0.64, using the combination of signal-based features and the body-model).

Benefit of Location In a second setting we incorporate location information as described in Section 3.3. The results are given in rows seven to twelve of Table 1. On average, the BM_IMU performs better than the signal-based approach on the IMU with an EER of 0.92 respectively 0.90. Only for three activities, the signal-oriented approach is marginally better. Combining the two feature types performs similar to the body-model only approach with an average EER of 0.92. Only for two activities, the signal-oriented features perform slightly better.

For the acceleration-based approach, we obtain an EER of 0.71 using signal-oriented features. The approach using BM_ACC again outperforms the signal-oriented approach with 0.81. A combination of both yields no significant improvement in average remaining at an EER of 0.81. For three out of twenty activities, the signal-based approach is slightly better.

For the IMU-based approach using location information does not have a significant effect. In case of acceleration sensors however integrating location information helps to improve the results narrowing the difference of the EER to the IMU-based approach to about 10%.

Reducing Number of Sensors In a third step we reduced the number of sensors from 5 to 2 sensors worn at the left and right wrist. Using the IMU approach the recognition drops about 7% from 0.93 (5 sensors) to 0.86 (2 sensors). All activities are better recognized with 5 sensors. The largest drop in performance of about 30% can be observed for the activities *opening/closing 2 doors*. The reason behind this are magnetic disturbances caused by the moving doors, which heavily influence the orientation estimation of IMUs.

Reducing the number of acceleration sensors the performance also drops significantly from 0.81 (5 sensors) to 0.61 (2 sensors). Whereas the IMU only looses 7%, the decrease using acceleration sensors is about 20%. Using two IMUs without location information achieves 0.86 EER whereas using two acceleration sensors achieves only 0.61 EER together with location.

Discussion In this section the effects on the recognition performance regarding different aspects, namely the *Benefit of Body-Model*, *Comparison between IMUs and Acceleration*, the *Benefit of Location* and the *Reduction of Number of Sensors* are reported. Incorporating a body-model improves consistently the results for all settings. As the analyzed activities are still relatively simple we expect that for more complex activities the margin between signal based approaches and body-model based approaches will become even more pronounced.

Replacing IMUs with acceleration sensors always results in a significant drop in performance. This is an interesting result in itself as most systems using body-worn sensors rely on accelerometer data only. By incorporating additional information (such as location and combining signal and body-model based features) in the best case the performance difference between using IMU-sensors only versus accelerometer sensors is still 10%. For activities such as *writing, check trunk, mirror* we obtain constantly good recognition rates for the BMU_ACC.

On rotation changes perpendicular to earth gravity which can be found in activities like *opening hood, closing hood, opening trunk, closing trunk* the acceleration based approach works expectedly well as the orientation towards ground can be estimated accurately using accelerometers. Figure 4 shows a sequence of snapshots of an *opening hood*

location		Average EER	open hood	close hood	open trunk	check trunk	close trunk	fuel lid	open left door	close left door	open right door	close right door	open two doors	close two doors	mirror	check trunk gaps	lock check left	lock check right	check hood gaps	open swl	close swl	writing
w/o	BM_IMU + signal, inertial	**.86**	.99	.99	.88	.95	.90	.92	.75	.85	.88	.81	.75	.51	.96	.94	.89	.86	.86	.82	.81	.94
	signal, inertial	**.82**	.97	.95	.88	.92	.88	.86	.68	.78	.77	.79	.67	.35	.95	.99	.86	.89	.63	.90	.71	.94
	signal, acceleration	**.43**	.28	.44	.51	.79	.45	.67	.23	.32	.35	.00	.35	.03	.90	.54	.70	.73	.35	.00	.09	.91
with	BM_IMU + signal, inertial	**.89**	.97	.99	.90	.97	.92	.97	.75	.88	.84	.85	.82	.56	.96	.97	.88	.88	.94	.91	.79	.95
	signal, inertial	**.86**	.95	1.00	.88	.94	.87	.91	.73	.84	.78	.79	.77	.51	.97	.97	.82	.88	.90	.90	.83	.96
	signal, acceleration	**.61**	.60	.50	.77	.96	.49	.85	.32	.45	.44	.38	.53	.22	.88	.90	.69	.68	.76	.50	.37	.95

Table 2. EER for activity recognition with 2 sensors (both IMUs and acceleration sensors) with and without location

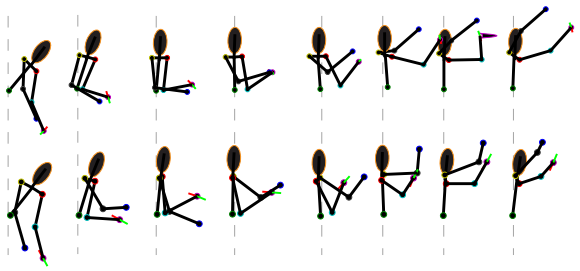

Figure 4. The Acceleration-based sequence (top) approximates the IMU-based sequence (bottom) well for the *opening hood* activity

activity. The user lowers himself to grab the hood and lifts it above his head. It shows that the BM_ACC (top) approximates the BM_IMU (bottom) visually well for this activity.

Figure 5. *Writing.* (Left) The wrong acceleration based body configuration. (Right) IMU-based body configuration.

As we limit the lower arm to rotate to the front, we may obtain wrong horizontal positions for the hands, like illustrated in Figure 5 – the left side illustrates *writing* using the acceleration based approach, the right side the more accurate IMU-based approach. Though the lack of full orientation information using acceleration leads to wrong postures, it might not hurt the classification task, as long as the posture and its features stay consistent within the same activity and distinctive enough between different activities. The good results of recognizing *writing* supports this assumption.

Without location information the acceleration based approach worsens significantly. Similar activities, e.g., *open/close hood, open/close trunk* get confused, as these are similar in motion and thereby distinguishable only by location. The IMU-based approach does not profit from location. As IMUs yield a global unique orientation and hereby encode the orientation of the wearer with respect to the car, it contains enough information to distinguish for instance *opening hood* or *opening trunk*. However as the sensor fusion of IMUs includes magnetometers BM_IMU suffers from magnetic disturbances found in activities like *closing/opening the door* and *checking the locks*. This influence is more intense using the wrist sensors only.

Obviously, there is a correspondence between the user's orientation with respect to the car and his absolute location in this specific scenario. This fact raises the question if the improved performance using IMUs instead of acceleration sensors only results from the global orientation given by the IMU sensors. To this end, we evaluated the setting combining BM_IMU and signal-oriented features for five and two sensors with a restricted feature set discarding global features, namely, the 3D global orientation. Table 3 shows the results applying five and two sensors.

Discarding all global features of the BM_IMU approach combined with signal-based features with five IMUs obtain an average EER of 0.81. In fact, the results are slightly worse compared with the global approach (0.93). They still outperform the approach combining BM_ACC with signal-oriented features on acceleration data only (average EER of 0.64). Regarding two sensors, the performance drops from 0.86 to 0.69 (0.43 for acceleration sensors only). From the results we can conclude that using IMUs instead of acceleration sensors only, we can still improve the results significantly without considering global features.

By exploiting the feature selection property of Joint Boosting, a combination of all features (BM_ACC, BM_IMU and signal-based) achieves a minor improvement to an average EER of 0.94

7. Conclusion and Outlook

This paper provides a systematic analysis regarding different activity representations (model-based vs. signal oriented) and sensor settings such as type and number of sensors. It is shown that model-based approaches enable more robust activity recognition than signal-oriented approaches. The improvement can be observed in all considered settings. Already a relatively simple construction of a body model using acceleration sensors helps to improve our results compared to the signal based approach. In future work, we will analyze a more sophisticated approach applying double integration to estimate the hand's position using the low variance phases to recalibrate and to keep hereby the

	Average EER	open hood	close hood	open trunk	check trunk	close trunk	fuel lid	open left door	close left door	open right door	close right door	open two doors	close two doors	mirror	check trunk gaps	lock check left	lock check right	check hood gaps	open swl	close swl	writing
5 senors	**.81**	.86	.86	.81	.87	.77	.85	.71	.81	.79	.79	.83	.72	.87	.86	.77	.79	.76	.85	.79	.84
2 sensors	**.70**	.82	.72	.77	.87	.71	.85	.63	.66	.58	.60	.60	.33	.83	.81	.77	.79	.54	.54	.58	.92

Table 3. EER for activity recognition with 2 and 5 sensors (BM_IMU + signal) without global features

accumulation error low. With a more sophisticated model we hope to lower the gap between the body-model using acceleration only and the body-model using precise but expensive IMUs.

Interestingly, promising results can be obtained using two wrist-worn IMUs only without any additional information. Whereas additional location turns out to be important information for activity recognition with low-cost and power-efficient acceleration sensors, the benefit of location for the IMU based approach is limited. As shown, different sensor requirements do not always lead to a high performance difference. Depending on the activities and the scenario, a prior effort to choose a feasible sensor setting is crucial for successful activity recognition.

In addition to our experiments using Joint Boosting as discriminant classifier, we evaluated a generative approach using Hidden Markov Models (HMMs). We experimented with different topologies of HMMs, more specifically an ergodic topology and a left-right topology. Additionally, we analyzed different numbers of states ranging from 5 to 15. We tested both HMM types with 2 different feature sets. First, we provided as input the sequence of primitives, which are also used by Joint Boosting. As HMMs have the property to capture continuous time series well, we also evaluated directly on the continuous values. Preliminary results show that the discriminant approach outperforms the generative approach for all settings by a significant margin. In future work, we plan to analyze the performance of HMMs in more details considering insights and methods of [14, 25] combining discriminative and generative approaches.

References

[1] http://www.ubisense.com.

[2] Xsens http://www.xsens.com.

[3] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. *Pervasive*, 2004.

[4] E. Farella, L. Benini, B. Riccò, and A. Acquaviva. Moca: A low-power, low-cost motion capture system based on integrated accelerometers. *Advances in Multimedia*, 2007.

[5] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 2000.

[6] E. Heinz, K. Kunze, M. Gruber, D. Bannach, and P. Lukowicz. Using wearable sensors for real-time recognition tasks in games of martial arts - an initial experiment. In *CIG*, 2006.

[7] T. Huynh, U. Blanke, and B. Schiele. Scalable recognition of daily activities with wearable sensors. In *LoCA*, 2007.

[8] H. Junker, O. Amft, P. Lukowicz, and G. Tröster. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition: The Journal of the Pattern Recognition Society*, 2008.

[9] J. Kela, P. Korpipäää, J. Mäntyjäarvi, S. Kallio, G. Savino, L. Jozzo, and D. Marca. Accelerometer-based gesture control for a design environment. *Personal Ubiquitous Comput.*, 2006.

[10] P. Klasnja, B. Harrison, L. LeGrand, A. LaMarca, J. Froehlich, and S. Hudson. Using wearable sensors and real time inference to understand human recall of routine activities. In *Ubicomp*, 2008.

[11] K. V. Laerhoven, M. Borazio, D. Kilian, and B. Schiele. Sustained logging and discrimination of sleep postures with low-level, wrist-worn sensors. In *ISWC*. 2008.

[12] S.-W. Lee and K. Mase. Activity and location recognition using wearable sensors. *Pervasive Computing*, 2002.

[13] J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. 2006.

[14] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *IJCAI*, 2005.

[15] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition. In *NIPS*, 2005.

[16] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. S. Intille. A long-term evaluation of sensing modalities for activity recognition. In *Ubicomp*, 2007.

[17] P. Lukowicz, A. Timm-Giel, M. Lawo, and O. Herzog. Wearit@work: Toward real-world industrial wearable computing. *Pervasive Computing*, 2007.

[18] P. Lukowicz, J. Ward, H. Junker, M. Staeger, G. Troester, A. Atrash, and S. Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive*, 2005.

[19] D. Mizell and I. Cray. Using gravity to estimate accelerometer orientation. In *ISWC*, 2003.

[20] G. Ogris, T. Stiefmeier, P. Lukowicz, and G. Tröster. Using a complex multi-modal on-body sensor system for activity spotting. In *ISWC*, 2008.

[21] R. Slyper and J. Hodgins. Action capture with accelerometers. In *Eurographics Symposium on Computer Animation*, 2008.

[22] J. Tiesel and J. Loviscach. A Mobile Low-Cost Motion Capture System Based on Accelerometers. *Advances in Visual Computing*, 2006.

[23] A. Torralba, K. Murphy, and W. Freeman. Sharing visual features for multiclass and multiview object detection. In *CVPR*, 2004.

[24] J. Ward, P. Lukowicz, G. Tröster, and T. Starner. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2006.

[25] P. Yin, I. Essa, T. Starner, and J. M. Rehg. Discriminative feature selection for hidden markov models using segmental boosting. In *ICASSP*, 2008.

[26] A. Zinnen and B. Schiele. A new approach to enable gesture recognition in continuous data streams. In *Proceedings of the 12th IEEE International Symposium on Wearable Computers*, 2008.

[27] A. Zinnen, C. Wojek, and B. Schiele. Multi activity recognition based on bodymodel-derived primitives. In *LoCA*, 2009.