All for one or one for all? Combining Heterogeneous Features for Activity Spotting

Ulf Blanke, Bernt Schiele
Computer Science Department, TU Darmstadt, Germany
{blanke,schiele}@cs.tu-darmstadt.de

Matthias Kreil, Paul Lukowicz
Embedded Systems Lab (ESL), University of Passau, Germany
{matthias.kreil,paul.lukowicz}@uni-passau.de

Bernhard Sick, Thiemo Gruber
Computationally Intelligent Systems Lab (CIS), University of Passau, Germany
{thiemo.gruber,bernhard.sick}@uni-passau.de

***Abstract*—Choosing the right feature for motion based activity spotting is not a trivial task. Often, features derived by intuition or that proved to work well in previous work are used. While feature selection algorithms allow automatic decision, definition of features remains a manual task. We conduct a comparative study of features with very different origin. To this end, we propose a new type of features based on polynomial approximation of signals. The new feature type is compared to features used routinely for motion based activity recognition as well as to recently proposed body-model based features. Experiments were performed on three different, large datasets allowing a thorough, in-depth analysis. They not only show the respective strengths of the different feature types but also their complementarity resulting in improved performance through combination. It shows that each feature type with its individual and complementary strengths and weaknesses can improve results by combination.**

***Keywords*-activity recognition, wearable computing, feature analysis**

I. Introduction

Activity recognition is a broad, active research area within the pervasive computing community. The type of activities that have been targeted range from modes of locomotion (walking, standing, running, etc.), through interaction with objects and devices (e.g., opening a drawer) to complex high-level approaches (e.g., preparing breakfast).

This paper deals with a specific subproblem of activity recognition: the spotting of sporadic actions using wearable motion sensors. Spotting means that we aim to locate a set of relevant, often very subtle actions in a continuous data stream. In general, the relevant actions are arbitrarily distributed and mixed with a large body of non-relevant actions. The problem is significant for two reasons. First, many complex activities can be decomposed into such isolated actions. Being able to spot and classify them is a key component of the more complex *composed activity* recognition problem. Second, it is known to be a hard problem that has not been satisfactorily solved so far. The main difficulties are ambiguities in the sensor signal, a high percentage of "NULL" class events in a typical signal, a lack of appropriate models for the "NULL" class, and high variability in the duration of relevant events.

There has been much previous work on activity recognition with wearable sensors (see related work).We build on this work to investigate an aspect that, in our opinion, has not received sufficient attention so far: feature definition.

The first contribution of this paper is the introduction of a new type of features adapted from time series approximation. As such, they are quite different in nature to the features currently used for activity recognition. The second contribution is an extensive comparative evaluation of the new polynomial features type, standard signal-oriented features (statistical parameters, frequency, etc.), and recently proposed body-model based features [1]. This includes an examination of the complementarity of the different feature types by analyzing their performance in different feature combination schemes. It shows that while the new polynomial features can not replace the other two types of features, they provide significant added value.

The evaluation is performed on three data sets with a total of 44 different activities collected from a total of 22 subjects and containing about 30 hours of data. Together, the three data sets provide a comprehensive test suite that allows for a thorough examination of the strengths and weaknesses of the different feature types. We want to emphasize that the feature comparison is not about comparing specific individual features but rather about different approaches to feature definition. Which specific features from which approach are actually used is determined automatically by a Joint Boosting algorithm.

First, we review related work (Section II) and then introduce all features used in this paper in Section III. Section IV explains the overall approach consisting of a common segmentation and a spotting procedure. Section V introduces the three datasets. Section VI describes the evaluation procedure and section VII discusses the experimental results. Finally, we conclude our findings in Section VIII.

II. Related Work

Previous research covers a variety of approaches to activity recognition. Besides different types of sensors and machine learning techniques (e.g., SVM, Boosting, or HMM), different types of features are employed. The predominant type of features are—what we call—signal-oriented features such as mean and variance [2] or frequency based features [3]. While the mean captures the posture during an activity the variance indicates how much motion is present in the signal. The combination of computational efficacy and the

descriptive power made them widely used in different studies. In [4], high level activities are successfully recognized using a fixed sliding window and mean and variance as features. Frequency based features prove to work well on activities including reoccurring motion patterns [5]. In [6], the signal is discretized and modeled by symbols. Then, similar subsequences which model motifs of activities are discovered.

While most of related work bases its detection on sliding windows with a fixed size, others segment the continuous data stream into windows of interest. Algorithms for time series segmentation can be found in a wide range of applications, for example medical and biological diagnostics, analysis of financial time series, speech processing, or sensor analysis. A commonly used method for representing and segmenting time series is piecewise linear approximation [7], used for example in [8] in conjunction with time-warping to segment ECG signals. In [9], SWAB (cf. [10], [11]) is used in a first detection stage.

Another type of feature and segmentation has been used in [1]. Here, atomic primitive features (such as moving the hands up, turning the arm, or keeping the arm in a specific posture) are derived intuitively from body movement for each activity. A comparison [12] to signal oriented features reveals that such a body model can leverage performance.

While [12] has used a single dataset for comparison only, the present paper extends their study by using three datasets. Furthermore, an alternative feature type based on piecewise linear approximation [13] of inertial sensor data is introduced, compared, and analyzed.

III. Features for Activity Recognition

It is often a manual and intensive task to choose or discover features that are suited for certain scenarios. The quality and appropriateness of this choice translates directly to recognition performance. In the following, we describe three fundamentally different feature types. First, we outline common signal based features used in this paper and which are most widely used in the community. Then we describe an alternative feature type based on motion primitives during activities. Finally, we describe a new feature type based on polynomial approximations.

A. Signal oriented features

For each sensor signal, e.g., a sensor's acceleration dimension (x, y or z), a set of features in the frequency and time domains are calculated. First, the Fast Fourier Transform (FFT) maps the incoming signal into the frequency domain. We use 10 coefficients and group them into five logarithmic bands by summation. 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, which gives a cue about the complexity of the signal is also added. Features in the time domain are mean and variance of the signal.

B. Body model derived features

Substantial variability in most activities requires the definition of invariant features to varying performances. Unlike in the signal based approach, the sensors are set in a relation to each other. With knowledge about the sensors' placement [1], we calculate a body-model as depicted in Fig 1. The orientation information of the user's upper and lower arms and the torso are concatenated to a kinematic chain starting at the torso and ending at the hand.

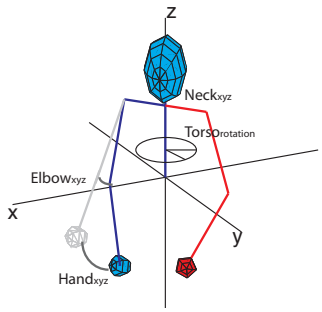

Figure 1. Calculated body model using 5 inertial measurement units placed at the back, the upper and lower arms. Given a global reference system by the sensor, the absolute direction of the torso can be estimated.

Using this body model, primitives are derived such as *moving the arms up or down*, *push-pull the hands*, *bend*, *twist the torso*, or *twist the arms*. For each kind of primitive, a temporal representation of a fixed size, the number of primitive occurrence, maximum, minimum, and average are considered as features. Furthermore, histograms of the primitive's length as well as directional vectors of subsequent hand positions are estimated and added. In addition to movement primitives, postures turn out to be a valuable cue for activity recognition. Here, the maximum, minimum, mean, and variance over the arms' orientation towards gravity, the distance between two hands, the hands' height, and the torso's direction in a global reference system are added to the feature set.

C. Novel Features Describing Trends in Time Series

In this section we describe a new kind of features that describe essential trends in time series (or segments of time series). These features can be used to determine the similarity of time series efficiently. For that purpose, we search for polynomials that approximate the time series and use coefficients of an *orthogonal expansion* of the approximating polynomial as features.

Assume we are given a time series (or segment) consisting of $N+1$ real-valued observations y_n at points in time x_n with $n \in \{0,\dots,N\}$. An optimally (in the least-squares sense) approximating polynomial $p_{\boldsymbol{a}}$ of degree K can be represented by a linear combination of $K+1$ basis polynomials p_k ($k \in \{0,\dots K\}$):

$$p_{\boldsymbol{a}}(x) = \sum_{k=0}^{K} a_k p_k(x), \tag{1}$$

with a weight vector $\boldsymbol{a} \in \mathbb{R}^{K+1}$, $\boldsymbol{a} = (a_0, a_1, \dots a_K)^{\mathrm{T}}$.

The basis polynomials must have the following properties:

1) They must have ascending degrees $0, \dots, K$.
2) The coefficient of the monomial with the highest degree of each basis polynomial must be one.
3) Each pair of basis polynomials p_{k_1} and p_{k_2} (with $k_1 \neq k_2$) must be *orthogonal* with respect to a certain inner product, That is, for all $k_1 \neq k_2$,

$$\sum_{n=0}^{N} p_{k_1}(x_n) p_{k_2}(x_n) = 0. \tag{2}$$

The choice of these basis polynomials depends on the points in time when samples are observed. In the context of a representation with orthogonal basis polynomials, the a_k are called *orthogonal expansion coefficients*.

Techniques can be applied allowing for an efficient computation of approximating polynomials in either sliding or growing time windows. Assume, we are given an approximating polynomial for a time series (or segment) of $N+1$ real-valued observations y_n at points in time x_n with $n \in \{0, \dots, N\}$ and a new observation y_{N+1} at x_{N+1}. Then, the approximating polynomial for either the observations y_n with $n \in \{1, \dots, N+1\}$ or the observations y_n with $n \in \{0, \dots, N+1\}$ can be computed with low computational effort which is independent from N and only depends on the polynomial degree K. In the case of a sliding window, the basis polynomials remain unchanged, in the case of growing windows they are updated "on the fly" in each step. This makes this technique well-suited for time-critical applications. More details can be found in [14], [13].

In [14], [15] we have shown that a_0, a_1, a_2, a_3, etc. can be interpreted as the optimal estimators of average, slope, curve, change of curve, etc. of the time series. Thus, they express the essential behavior of the time series in a few values (usually $K \ll N$) and can be used as features for time series classification. It is quite simple to choose an appropriate polynomial degree in a real application: The recommendation is simply to select a degree that is higher than an assumed one. If the degree was too high, the respective coefficients contain no information. This could easily be detected by an appropriate feature selection technique.

With a given representation of a time series by orthogonal expansion coefficients, two time series can be compared simply by taking the Euclidian distance (or the scaled Euclidean distance) of two orthogonal expansion coefficient vectors [15]. Doing so, the temporal effort to compare precomputed coefficient vectors is marginal. However, depending on the application we may also wish to compare two time series neglecting constant offsets in the target domain or different lengths. This can be done easily with some simple transformations as shown in [15].

In addition to the coefficients, the lengths of the time series themselves and the approximation errors turned out to be useful features in various applications [15].

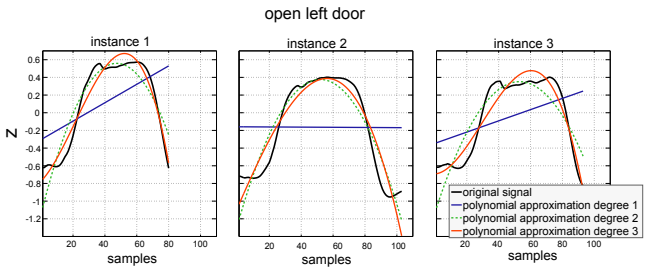

Figure 2. Example for polynomial approximation of degrees 1 to 3 of the right hand's trajectory (z-axis) during the gesture of *openening a car's door* from the car quality inspection dataset.

In this paper, polynomial features are used for activity spotting for the first time. An example of approximating polynomials is shown in Fig. 2: the approximation smoothens the signal but preservs its original form.

IV. Spotting Activities

Many recognition tasks are based on two steps. First, the data is segmented using a sliding window with a fixed size to calculate features on a continuous data stream. Then, a classifier is trained on each window given a class label. In the classification task a sliding window is applied again to unknown data, returning scores for all activities. The following describes the method used to spot activities.

To reduce the amount of potential windows we adopt the segmentation procedure suggested in [1]. The segmentation is based on the observation that the movement of the hand slows down at the starting and ending point of an interaction. Since the variance over the hand positions is lower at these points, local minima within the variance of the hand positions can be detected separately for both hands.

Boosting [16] is a state-of-the-art machine learning algorithm for binary classification and has been used successfully for activity recognition [3], [17], [2], [1], [18]. In each boosting round, a weak classifier is trained, often based on a single feature. Weak classifiers are then combined to a final strong classifier for categorization. Torralba et al. [19] propose an extension called Joint Boosting that trains weak classifiers shared across multiple categories. In activity recognition, groups of similar activities are separated during the initial boosting rounds. In the following rounds, activities in the same group are discriminated with additional weak classifiers. Joint Boosting reduces the computational complexity by sharing features across several classes. Here, it uses the segment features described Section III. Given the annotations, features on the positive training segments can directly be calculated and used as inputs for the training phase. When calculating the features on the test data and the negative instances of the training data, all possible combinations of segments with specified minimum and maximum lengths are considered. In the test phase all segments are classified and activities are spotted by finding local maxima in the streams of the classifier score.

V. Datasets

As motivated in Section I, we want to benchmark the novel feature type for spotting human activities. We conduct our studies on three different datasets, namely the *car quality inspection*, the *woodshop*, and the *drink and work* datasets. Each contains real-world challenges such as the high variability in executing such activities. While the first two datasets contain a large variety of classes, the latter is specifically interesting for its fairly short drink activities amid a large amount of background data.

Across all datasets inertial measurement units (IMU) [20] are used to collect the data. Each sensor integrates 3D- acceleration, rate of turn, and magnetic field data. A sensor fusion algorithm is used which allows the sensors to accurately estimate absolute orientation in a three-dimensional space in real-time. The result is the estimated orientation of the sensor-fixed coordinate system with respect to a Cartesian earth-fixed coordinate system. The sensors are located at the wearer's torso and the upper and lower arm. While performing the activities, the subjects are recorded on video for later annotation.

First, the next section introduces a new *woodshop* dataset. Then, we describe the *car quality inspection* dataset, which was previously used in [1], [12], [21]. Finally, the new *drink and work* dataset is outlined.

A. Woodshop

We asked 8 different people to perform the overall task of building two wooden book boxes. Fig. 3 (a) shows the book box from the front and from the side. Building such a bookshelf consists of a variety of manual activities, for instance, *sawing*, *drilling*, or *screw driving*. In total the complete procedure covers 22 activities. The procedure took roughly 45–70 min per person, resulting in approximately 9 h of data in total.

In the following, challenges in recognizing different physical activities are motivated. In many scenarios, selected activities differ significantly in their constitution. Often, activities are characterized by repetitive movements such as turning the arm when screwing or moving the arm up and down while hammering. Beyond these activities of longer duration (>10 s), very short activities (<3 s) such as drilling, marking holes, or hanging up boxes are of interest. Not only the short duration complicates the detection of activities. In addition, short activities often do not contain discriminant arm movements. Whereas hammering or turning screws can be identified by noticeable arm movements, the arm position hardly changes for activities such as cutting the template or marking holes for drilling. The dataset includes activities of diverse complexity as illustrated in Fig. 4. In addition to repetitive activities such as sawing, hammering, or screwing, the recognition of short activities such as drilling, fixing a back support, marking, cutting, or hanging up the boxes are of major importance.

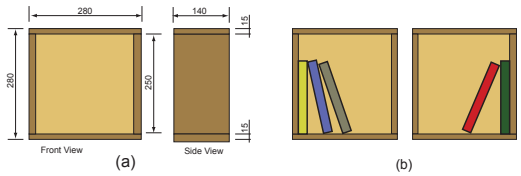

Figure 3. Description of boxes in the manual. (a) front (b) side view

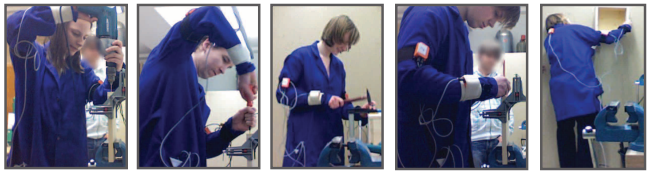

Figure 4. Diverse class complexity: drilling, screwing, hammering, mark, hang up (left to right).

Beyond the difficulty to find discriminant characteristics of these activities, the execution often differs significantly between subjects.. Fig. 5 illustrates four different people while *screwing side parts*. Although the subjects perform the same activity, a high variance in execution (intra class variability) can be observed. A rotation of the screw driver can be enforced either by hand or turning the whole arm (see subject in the left images). Whereas the subject in the third picture uses his left hand, the last subject clasps the screw driver in a different way than the three other subjects.

B. Car quality inspection

The dataset contains 20 activities that are performed during a typical car quality inspection [22]. Example activities are checking gaps of the car's body or inspecting movable parts, for instance, by opening and closing doors. Besides a high variability of motion patterns, such activities are short (on average 3 s per activity). As a result, the ratio of the activities versus the "NULL" class is 1:135 for each activity. This makes activity spotting a challenging task.

The dataset was recorded within the scope of an industrial project focusing on wearable technology in production environments [23]. In total, 12 h of data were recorded. 8 subjects performed the activities, repeating the procedure about 10 times on average.

C. Drink and work

This dataset consists of several drinking events embedded in daily scenarios. The subject drinks from four different drinking vessels (glass, bottle, beer mug, cup) while completing four scenarios in a typical daily routine: *office work*, *eating*, *gaming* and *leisure*. Within these routines, activities such as *using the computer* and *printing*, *preparing*

Figure 5. Intra-class variability when performing the activity screwing.

a sandwich, *scratching head*, *answering phone call* etc. occur. The presence of ambiguities (e.g., drinking vs. eating or scratching the head) and a great percentage of NULL class makes this dataset particularly interesting for analysis.

Altogether, six subjects were recorded, each 50–60 min, including about 12 min of drinking.

VI. Evaluation

In all datasets we evaluate the performance of activity spotting for each activity individually. A leave-one-user-out cross-validation is performed to enable user-independent activity recognition. In each cross-validation round, scores for all detected segments (see Section IV) are calculated. A segment S is counted as true positive if the annotated ground truth segment A has the same activity label and if the following equation holds:

$$start(A) \leq center(S) \leq stop(A), \tag{3}$$

where $start(A)$ and $stop(A)$ correspond to the begin and end times of the ground truth segment A and $center(S)$ indicates the central time of segment S. In other words: Only if the central time of a spotted 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 with increasing recall for a particular activity. For the description of the results we use the equal error rate (EER), a characteristic point in the precision-recall curve.

Unlike in previous work, we use 20% of background data only during training, due to computational reasons.

VII. Results

As mentioned earlier, we compare three different feature types on three different datasets. In the following, results are given for each dataset individually. First, we describe the results on the one-class dataset *drink and work*. Then, results on the multi-class datasets *woodshop* and *car quality inspection* are given.

A. Drink and Work

Spotting drinking activities while the user resides in different working scenarios results in an EER of 90% using *signal oriented* features and 89% using *polynomial* features. By combining both feature types, we achieve an EER of 91% and by combining all features 92%. Using *body model* features only, we achieve an EER of 88%.

To understand the performance drop using *body model* features only, we examined the features selected by Joint Boosting. It turned out that *hand height* features are predominantly represented with 45.7%. This corresponds to the intuition, that the height of the hand is a strong cue for drinking events. However, the dataset also contains activities such as *eating* and *scratching the head* which cannot be discriminated by the height only. Hence, more specific features are necessary.

B. Woodshop

The overall results are depicted in Table I. The rows are sorted by the average EER.

Using feature types individually (row 4 to 6), the performance is worse than using their combination (row 1 to 3). For individual activities, *signal oriented* features perform better for 11 activities. Using *polynomial* features, better results are achieved for 5 activities. *Body model* features achieve better results for one activity only. While the average performance of *signal oriented* features and *polynomial* features is similar (53% and 51%), the *body model* features perform significantly worse with 41%. Combining different feature types (row 1 to 3) performs better than each feature type individually. The combination of all three features types performs best at about 59% EER. Here, the combination of all feature types (row 1) performs better on 8 activities. The combination of *polynomial* and *signal oriented* features is better on 6 activities and combining *signal* with *body model* features performs better on 7 activities.

Near perfect recognition is gained for *sawing* using *signal oriented* features. Intuitively, the variance, respectively the recurrent movement in the hand is a good cue to discriminate this activity from others. As one expects for multi-class activity recognition, a significant number of similar activities are not well recognized. More specifically, the following activities are not recognized well: *marking positions of the back part*, *marking template*, *marking holes in template*, and *cut template*. All those activities are very short and almost no arm movements are involved. Furthermore, executing those activities allows for a high variance, for example when cutting the template. Here only the interplay between all features is able to recognize the activities to a certain extend (40–60% EER). Interestingly, on those activities the polynomial features seem to contain important information (e.g., on *marking positions of the back part* both, *body model* and *signal oriented* features, have an EER of 0% while *polynomial* features yield 19%). The best performance on those classes always involves polynomial features.

Altogether, the *body model* features have a significantly worse performance. Note that the *body model* features were designed on the *car quality inspection* dataset. Postures and movements are quite different in both scenarios. For instance, while the person is standing straight and bowing for an activity in the *car quality inspection*, the subjects are bending over for almost the entire procedure of building a bookshelf (excluding activities at the wall). Fig. 6 shows the different nature of the scenarios. While activities in the *car quality inspection* dataset are characterized by an obvious displacement of the hands, the activities of the *woodshop* dataset contain less movement. *Body model* features such as the *hand height*, *the distance of two hands*, or *bending* can help to distinguish activities such as *drilling holes into a wall* and *drilling holes into the backpart*. However, many activities, e.g., *hang up box* and *mark holes in wall* with

	1 Sawing	2 drill holes (fix side part)	3 screw (fix side part)	4 drill holes (join side part)	5 screw (join side part)	6 drill back part	7 screw back part	8 hammering sidearms	9 hammering back part	10 drill holes in wall	11 screw into wall	12 sanding	13 painting	14 mark holes in wall	15 hang up box	16 mark holes (fix side part)	17 mark holes (join side part)	18 mark positions back part	19 mark template	20 mark holes in template	21 cut template	22 glueing	Average EER
Poly+Signal+Bodymodel	.81	**.86**	.64	.71	.68	.71	.71	.36	**.50**	.78	.71	.43	.37	**.93**	.52	.61	.38	.00	**.57**	**.57**	**.43**	**.64**	**.59**
Poly+Signal	.84	.78	.59	.69	**.69**	.69	.69	.44	.00	**.87**	.69	.50	.41	.87	**.80**	**.63**	.45	.00	.25	.25	.37	.59	**.55**
Signal+Bodymodel	.94	.69	.66	.66	.59	**.75**	**.75**	.37	.25	.74	**.87**	**.56**	.43	.87	.75	.59	.39	.07	.37	.00	.00	.53	**.54**
Signal	**.97**	.72	**.69**	**.72**	.34	.69	.69	**.56**	.00	.81	**.87**	**.56**	**.52**	.88	.64	.59	**.49**	.00	.37	.00	.00	.54	**.53**
Poly	.75	.84	.50	.59	.44	.69	.50	.31	.31	**.87**	.81	**.56**	.43	.83	.60	.55	.34	**.19**	.37	.25	.00	.56	**.51**
Bodymodel	.75	.72	.59	.59	.38	.25	.69	.00	.00	.88	.75	.38	.28	.56	.34	.59	.41	.00	.00	.25	.00	.50	**.41**

Table I. Results for the *woodshop* dataset using different feature types (*signal oriented*, *body model* and *polynomial* features and their combinations. The rows are sorted according to the achieved Equal Error Rate (EER). The bold numbers indicate the maximum EER per activity.

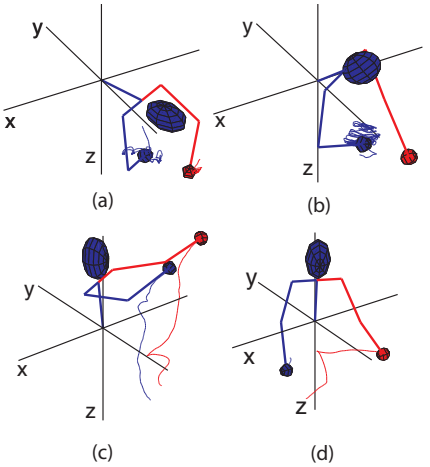

Figure 6. Examples for activities from the *woodshop* (a, b) and the *car quality inspection* (c, d) datasets: (a) Marking holes in template, (b) sawing, (c) opening the hood, (d) opening the left door. The stick figure shows the end posture while the colored thin lines show the trajectories of the hands.

similar posture and motion demand more specific features. This indicates that the design of *body model* features is not yet complete and a more complete set of features is required to enable good performance for a wider range of scenarios.

Table II shows the selected coefficients when using *polynomial* features. Looking at the normalized coefficients, the zero-order coefficient (i.e., the mean) has the highest fraction (49%). From the first to the forth coefficient the selection is almost equally distributed with about 9%. Beyond the forth coefficient the fraction drops to 2–5%. Unnormalized coefficients are almost completely ignored. Only for the first and the second we can observe a slight usage of 7% in total. Different instances of a particular activity have not necessarily the same duration. Hence, the unnormalized coefficients can be different for the same activity as they strongly depend on the activities' length. The rather uniform distribution of the normalized coefficients with a higher degree indicates that subtle changes in the signal, represented by the respective higher degree, can be important to distinguish different activities. It also reveals that the posture, which is described by the mean of the signal, proves to be a good (initial) cue for several activities in this dataset.

Coefficient	0	1	2	3	4	5	6	7	8
Normalised	46%	9%	9%	5%	10%	5%	2%	3%	4%
Unnormalised	0%	5%	2%	0%	0%	0%	0%	0%	0%

Table II. Distribution of coefficients for the *polynomial* features (*woodshop* dataset). For the approximation a degree of 8 was used.

C. Car quality inspection

The overall results are given in Table III. Again, the best results can be achieved by combining different feature types derived from the *signal* and the *body model* (89% EER). Combining all feature types, the EER is 88%. While in the former case a higher EER is achieved for 10 activities, the EER differs less than 1% for 7 activities. For 3 activities, the combination of all features performs better.

In contrast to the woodshop dataset, the *body model* features alone perform better (87%) on this dataset and outperform *polynomial* (83%) and *signal* features (84%) individually and in combination (86%). Here again, combining *polynomial* and *signal oriented* features performs better than both feature types individually.

Coefficient	0	1	2	3	4	5	6	7	8
Normalised	48%	13%	5%	4%	2%	2%	1%	3%	1%
Unnormalised	0%	15%	7%	2%	1%	0%	0%	0%	0%

Table IV. Distribution of coefficients for the *polynomial* features (*car quality inspection* dataset). For the approximation a degree of 8 was used.

Table IV shows the distribution of selected coefficients. As in the woodshop dataset, the fraction for the zero-order coefficient is the highest with nearly 50%. Again, the posture can be a good cue for discriminating the activities in this dataset. This is followed by the first (normalized and unnormalized) coefficient with a total of 28%. The fraction of the remaining coefficients of higher degree drops to 0–5%. Compared to the woodshop dataset, this indicates that the slope of the polynomial, which can be interpreted as direction of movement/signal change, seems to have a stronger impact in the *car quality inspection* scenario than coefficients of higher degrees. This corresponds to the intuition that activities in this scenario are coarser and do not profit from a detailed signal description by higher degrees.

VIII. Lessons Learned

Across a broad range of recognition problems as represented by the experiments investigated in this paper the following can be said: *Overall* (averaged overall all data sets and activities), none of the three feature types emerges as a

	1 Open hood	2 Close hood	3 open trunk	4 check trunk	5 close trunk	6 fuel lid	7 open left door	8 close left door	9 open right door	10 close right door	11 open two doors	12 close two doors	13 check mirror	14 check trunk gaps	15 check lock left	16 check lock right	17 check hood gaps	18 open swi	19 close swi	20 writing	Average EER
Signal+Bodymodel	.99	.97	.88	.96	.87	.87	**.79**	**.84**	.83	.78	**.94**	**.81**	.95	**.97**	**.90**	.85	**.85**	.95	**.92**	**.91**	**.89**
Poly+Signal+Bodymodel	.98	**1.0**	.86	**.97**	.77	.84	.78	.80	**.86**	.81	.93	.79	.96	.95	.86	.85	.83	.95	.90	.88	**.88**
Bodymodel	.99	.94	**.92**	**.97**	**.90**	**.89**	.77	.81	.85	.77	.90	.70	.94	.96	.84	.81	.83	.90	.84	.84	**.87**
Poly+Signal	.99	.96	.75	.96	.77	.86	.76	.81	.79	**.83**	.91	.72	**.97**	**.97**	.84	.83	.77	**.97**	**.92**	.88	**.86**
Signal	.99	.97	.82	.88	.64	.79	.69	.81	.74	.78	**.94**	.71	.95	.94	.84	.85	.74	.96	**.92**	.88	**.84**
Poly	**1.0**	.97	.79	.91	.65	.86	.66	.76	.78	.79	.82	.73	.94	.96	.78	**.88**	.74	.92	.85	.90	**.83**

Table III. Results for the *car quality inspection* dataset using different feature types (*signal oriented*, *body model* and *polynomial* features and their combinations. The rows are sorted according to the achieved equal error rate (EER). The bold numbers indicate the maximum EER per activity.

clear "winner" or "looser". The combination of feature types consistently leads to better performance on all datasets. This indicates that the features types provide complementary information, which can be leveraged by algorithms containing feature selection mechanisms.

While *overall* there is no clear ranking, there are significant variations at the level of individual activities and even data sets. Thus, body model features clearly outperform the polynomial features on the car quality inspection data set (EER 87% to 83%), while on the workshop data set the polynomial features do far better (EER 51% to 41%). Looking at the variations in feature performance at a more detailed level, the following interesting observations emerge:

As expected the body model features perform best in the car inspection data set for which they have been hand crafted. Polynomial features have performed particularly well for classes that are very difficult to recognize (*marking* activities and *cut template*). Best performance on those classes always involves polynomial features with the difference being very significant (e.g. 57% vs. 25% on *marking holes in template*). On the car inspection data set, the combination of polynomial and signal features is very close to body model alone. This is significant because hand crafting features for an application involves a lot of effort from a human, while the signal and polynomial ones are "generic" and can be automatically generated.

Altogether, it can be said that while the polynomial features can not replace signal oriented or body model features, they can provide an added value. While we investigated polynomial features per dimension in this work, we plan to continue our work on potentially more expressive polynomials of trajectories in a 3D space and using its representation for improved detection of potential segments of interests in the segmentation step.

ACKNOWLEDGMENT

This work was funded by the German Research Foundation (DFG) within the project "Methods for Activity Spotting With On-Body Sensors" and the graduate training group "Topology of Technology".

REFERENCES

[1] A. Zinnen, C. Wojek, and B. Schiele, "Multi activity recognition based on bodymodel-derived primitives," in *LoCA09*.

[2] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *Pervasive04*.

[3] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," in *IJCAI05*.

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

[5] T. Huynh and B. Schiele, "Analyzing features for activity recognition," in *EUSAI05*.

[6] D. Minnen, T. Starner, I. Essa, and C. Isbell, "Discovering characteristic actions from on-body sensor data," in *ISWC06*.

[7] D. Lemire, "A better alternative to piecewise linear time series segmentation," in *SDM05*.

[8] H. J. L. M. Vullings, M. H. G. Verhaegen, and H. B. Verbruggen, "ECG segmentation using time-warping," in *IDA-97, Reasoning about Data*, London, U.K.

[9] O. Amft, H. Junker, and G. Troster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," in *ISWC*, Osaka, Japan, 2005.

[10] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *ICDM*, 2001.

[11] E. Keoh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," in *Data Mining in Time Series Databases*, ser. Machine Perception and Artificial Intelligence, no. 57, 2004.

[12] A. Zinnen, U. Blanke, and B. Schiele, "An analysis of sensor-oriented vs. model-based activity recognition," in *ISWC09*.

[13] E. Fuchs, T. Gruber, J. Nitschke, and B. Sick, "On-line segmentation of time series based on polynomial least-squares approximations," *PAMI '10*.

[14] E. Fuchs, T. Grube, J. Nitschke and B. Sick, "On-line motif detection in time series with SwiftMotif," *Pattern Recognition*, vol. 42, no. 11, 2009.

[15] D. Fisch, T. Gruber, and B. Sick, "SwiftRule: Mining comprehensible classification rules for time series analysis," *TKDE'10*.

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

[17] D. Minnen, T. Westeyn, D. Ashbrook, P. Presti, and T. Starner, "Recognizing soldier activities in the field," in *IFMBE07*.

[18] N. Ravi, N. Dandekar, P. Mysore, and M. Littman, "Activity recognition from accelerometer data," in *AAAI05*.

[19] A. Torralba, K. Murphy, and W. Freeman, "Sharing visual features for multiclass and multiview object detection," in *CVPR04*.

[20] "Xsens." [Online]. Available: http://www.xsens.com

[21] G. Ogris, T. Stiefmeier, P. Lukowicz, and G. Tröster, "Using a complex multi-modal on-body sensor system for activity spotting." in *ISWC08*.

[22] T. Stiefmeier, G. Ogris, H. Junker, P. Lukowicz, and G. Tröster., "Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario," in *ISWC06*.

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