

Allowing Early Inspection of Activity Data from a Highly Distributed Bodynet with a Hierarchical-Clustering-of-Segments Approach

Matthias Kreil
DFKI Kaiserslautern
matthias.kreil@dfki.de

Kristof Van Laerhoven
Technische Universität Darmstadt
kristof@ess.tu-darmstadt.de

Paul Lukowicz
DFKI Kaiserslautern
paul.lukowicz@dfki.de

Abstract—The output delivered by body-wide inertial sensing systems has proven to contain sufficient information to distinguish between a large number of complex physical activities. The bottlenecks in these systems are in particular the parts of such systems that calculate and select features, as the high dimensionality of the raw sensor signals with the large set of possible features tends to increase rapidly. This paper presents a novel method using a hierarchical clustering method on raw trajectory and angular segments from inertial data to detect and analyze the data from such a distributed set of inertial sensors. We illustrate on a public dataset, how this novel way of modeling can be of assistance in the process of designing a fitting activity recognition system. We show that our method is capable of highlighting class-representative modalities in such high-dimensional data and can be applied to pinpoint target classes that might be problematic to classify at an early stage.

I. INTRODUCTION

Activity recognition methods on the basis of wearable inertial sensors have been investigated since at least a decade and a half. It has been discovered early on that combining such sensor data from multiple locations on the body, for instance from both leg- and arm-worn positions, allows these systems to perform well on multiple complex physical activities. Many of these studies have reported that it is especially hard to select and use the right features, as these make a high dimensionality of the input space even larger, leading inevitably to various problems in the following classification steps. This problem, known since decades as the Curse of Dimensionality, has been mostly dealt with by using traditional feature selection methods such as Boosting-based approaches [1], Component Analysis [2], or Wrapper methods [3] to reduce the large input dimensions for these systems.

We propose a method that relies on the hierarchical clustering of discovered patterns of both inertial trajectories and angular data, across all of the body-worn sensors. A benefit of this approach is that this allows thorough human inspection of which data segments are discovered as meaningful for given activities, since the representation of the traces and angular variations can be inspected and visualized at the clustering stage in the system, at which point the cluster centroids can be compared to each other for proximity. Figure 1 depicts an overview for the five phases of our proposed system, in which the raw inertial data is eventually to be classified.

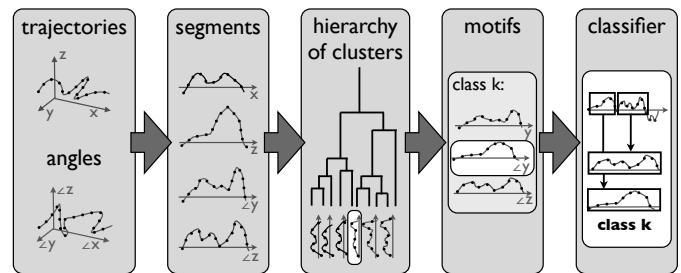


Fig. 1. Overview of the five main phases of the proposed activity recognition system. After hierarchically clustering data segments appearing in the sensors' trajectory and angle time-series, it is possible to use distance measures on the cluster centroids to investigate to which degree the data from different classes will be distinguishable across all sensors and features.

A large amount of related work exists on activity recognition systems using body networks of various inertial sensors. In [4] for instance, three different feature types, signal oriented features, body model features and polynomial features, are evaluated. They are tested in spotting activities using 3 experiments. The result concludes that all 3 feature types have their benefits depending on the classification problem. [5] selects sensors with an information theoretical approach. Accuracy and diversity measures are even used for dynamical reconfiguration of the sensor choice when the availability of sensors changes. In [6] an experiment with 5 accelerometers on different body parts was conducted by 20 subjects. Several signal and frequency domain features are calculated and then tested with a variance of classifiers getting the best overall accuracy rate of 84% with a decision tree classifier. Taking only 2 from 5 accelerometers on thigh and wrist changes the recognition results only slightly for the worse.

Other work such as [2] uses 2 accelerometers mounted on the left and right side of the hip to detect walking tasks. Wavelets were in this work focused on as the selected features and for dimension reduction, ICA- and PCA-based techniques were used. The classification with a multilayer perceptron returned a best recognition result between 83-90%. In [7], a similar experiment setup to this paper's is described with 7 accelerometers and calculating features derived from a body model. The most relevant features for activity recognition are in this work selected by joint boosting.

The contributions of this paper are threefold: 1) We show that it is possible to obtain an early insight into which sensors and features can be expected to perform well in distinguishing the different target classes. 2) We furthermore show that it is also possible to inspect at an early stage which classes are prone to misclassification between each other for particular sensors and features. 3) By comparing the results at the clustering stage with the confusion matrix characterizing the final classification performance, we show that many problems appearing at classification can be identified already at the clustering stage.

The remainder of this paper is structured as follows: In the following section, we introduce the experiment and describe the methodology applied to test our new algorithm. Next, we show the single steps of our algorithm in detail, with special focus on the clustering and classification stages. Then we discuss and illustrate what type of observations can be made in this type of system at the early clustering phase, and conclude the paper in section V.

II. EXPERIMENT AND METHODOLOGY

The system presented in this paper enables inspection of the high-dimensional data at an early stage in the recognition process. To illustrate how such a visualization can be achieved, we will first introduce a publicly available dataset for an inertial body sensor network and describe how we analyze the data in two phases in our system, both at the clustering phase and at the final classification stage.

A. Dataset

We used a dataset containing activities from a car maintenance scenario, which has already been investigated thoroughly in activity recognition [8] [9] and is publicly available. In this dataset’s scenario, workers wearing a body sensor network integrated in a jacket are performing an inspection on cars checking whether they have been assembled correctly on the production line. The sensor data is generated by 7 MTX inertial sensors (3D accelerometer-, gyroscope- and magnetic sensors) fixed on the workers torso, the upper and lower arm, and the hand, for both sides. Nineteen different activities are present in the dataset as annotations, which are shown in Table I. Eight subjects were recorded getting about 12 hours of data in total. Each subject repeated the experiment procedure about 10 times. The data was annotated by observers during the recording.

B. Visual Inspection at the Clustering Level

We will use the aforementioned dataset to investigate two research questions in particular: How can the high-dimensional data be analyzed at the clustering level for its suitability for activity recognition, without resorting to feature selection? And how do the generated clusters correlate to the final activity recognition results?

The central plot to answer these questions can be seen in Figure 2. Each small quadratic subplot corresponds to one feature (x, y, or z axis of a sensor’s trajectory or its dominant

TABLE I
THE 19 DIFFERENT CLASSES CONTAINED IN THE CAR MAINTENANCE DATASET USED FOR EVALUATION OF THIS PAPER’S APPROACH.

class	activity	class	activity
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	mirror	14	check trunk gaps
15	lock check left	16	lock check right
17	check hood gaps	18	open spare wheel box
19	close spare wheel box		

angular changes) corresponding to the different observed body parts (shoulder, elbow, wrist, hand tip). The axes’ indices with the labels 1 to 19 represent the 19 classes of the dataset. The subplots are symmetric divided by the diagonal white line of the cluster precision value of pairs of the same class. The cluster precision value between 2 class pairs show the similarity between the clusters found for these two classes which reflect the difficulty to separate the two classes from each other: Bright colors show high inter-class cluster similarity, which explains the white diagonals of pairs of the same class, and dark colors a low similarity. Knowing this, it is possible to identify channels which tend to mix up different clusters from those who have a good class distinctiveness. It is also possible to see for single channels, which classes’ data tend to be close to each other and might get easily mixed up.

C. A Comparison with the Recognition Results

In order to verify whether it is indeed possible to detect target classes that are hard to distinguish from each other using cluster precision, we use primarily the confusion matrices for the entire system. Obtaining the recognition results requires more time and analysis generally, as this is typically done via n-fold (in our case 8-fold leave-one-user-out) cross-validation. If we would be able to detect classes that are hard to distinguish at the clustering phase, however, it would not only lead to a faster way of inspecting the data: we could also examine which sensors (or body locations) would contribute to such difficulties.

The next section will go in further detail on how the data is clustered hierarchically to produce the measures behind visualizations such as Figure 2.

III. ALGORITHM DESCRIPTION

One problem of the dataset described above is the number of sensors and the dimensionality of the resulting data. It would ease the activity recognition process considerably, if we could concentrate only on the important signals and also only on parts of the labeled data that are significant for the activity, such as the first seconds of data from the hand-mounted sensor when a subject pulls the door handle. Our proposed

algorithm tries to find segments in the signal channels which are significant for the activity to lower the dimensionality in the dataset, using unsupervised learning. Furthermore, we get an early phase vital information on the usefulness of single features or data channels in respect to the distinctiveness between different classes.

A. Data preprocessing and segmentation

The acceleration data from the 7 MTX inertial sensors is used to construct a body model, connecting the single sensor nodes, as proposed in [9], which means that we get 3D trajectories for the back, the shoulder, the elbow, the wrist and the hand tip. Additionally we calculate angles between the arm joints and the hand joint.

To segment the continuous signal, we calculate for each axis of the 3D trajectories the velocity and after low-pass filtering this signal, we cut the signal of the single axes of the trajectory at those points, where the corresponding velocity signal has a zero crossing. Reason for this lies in the fact that before every short activity, the movement direction has to change resulting in a zero crossing of the velocity signal as the appropriate way to segment activity signals. In [7], this form of segmentation was implemented in a similar way, using the variance of the hand position to find fixed but short position of the hand and turning points.

B. Clustering

In total we have 34 so-called channels (which can be interpreted as single-axis features for all sensors), which are eight 3-dimensional trajectories (shoulder, elbow, wrist and hand tip, left and right) and 10 angles (shoulder y z; elbow z; wrist y, z direction, left and right). The goal is to find significant channels and segments for each activity we want to spot. We use the hierarchical clustering built in Matlab with average linkage for each channel on all segments which are labeled for the searched activity class to accomplish that. As distance function between two signals we take Dynamic Time Warping (DTW). The perfect result would be to find for each activity class one cluster with exactly one segment as member. If one cluster contains several segments of one class instance, or if a cluster does not contain segments of several class instances this cluster gathers segments which seem to be not very distinctive for this activity. In the selection process of the best clusters we first remove all cluster which do not cover 20% of all train activity instances. Next we calculate on the training data precision and recall, using the classification routine described later, and sort the clusters according to the highest precision. Another option used is to sort the clusters according to the sum of the smallest DTW distance to all other cluster centers of the other classes, which is related to the cluster analysis described next.

C. Cluster Analysis

To get an idea of the distinctiveness between classes of single channels we calculate the cluster precision on the cluster tree for each clustering. For that we take all clusters of

one channel and one class and search for its cluster centers the cluster center of all other classes, same channel, with the smallest Dynamic Time Warping (DTW) distance. Every found cluster center pair is not considered any more in the next search steps for the nearest cluster centers. The sum of the smallest distances divided by the number of clusters provides the cluster precision result for each class pair. The higher the value the more dissimilar are the cluster center of the two classes on this channel and due to that the distinctiveness of these two classes in this channel is high. Figure 2 shows the cluster precision of all channels. A high color means that the mean distance between the cluster centers of this two classes is short, so they tend to be mixed when classified; dark colors stand for a far distances between the cluster centers so a misclassification is unlikely. For example the Z axis of the hand tip, wrist and also of the elbow have a good distinctiveness, especially in the first classes. Also angles have good cluster precision values. In contrast the shoulder is not very significant, which is logical, as the main activity is usually conducted by arms and hands for activities like opening the door.

D. Classification

After selecting the best clusters like described above, we conduct a basic classification. We calculate the DTW distance between the test data segments and all best cluster centers. All segments of the test data which have a smaller DTW distance to a cluster center than a threshold are counted as found activity. Two thresholds were tested, the first is the DTW distance, where 95% of the cluster members are still nearer to the cluster center, the second is the mean of all DTW distances from cluster members to cluster center. Next we construct a matrix, containing the found segments for the best 200 best cluster centers per selected channels. The channels are sorted by summing up all cluster precision values for each channel and considering the one as best, which has the highest value (see table II). First we select the 2 best channel, then the 3 best channels and so on until all channels are selected. The next step is to take first the result of one cluster center, then two and so on until reaching 200 and also count one activity as found if it was found in one cluster center, then in two and so on until reaching the max number of taken cluster centers. Additionally we tried this for 3 different features to get the best cluster center, which are the best precision on the trainings data with 95 quantile threshold, best precision on the trainings data with mean threshold and sum of the DTW distance of all nearest clusters of all other classes. For each try precision and recall is calculated, leaving us with many results. The best results can be seen in figure 3 and 4. Figure 3 shows 2 confusion matrices with the best results of the sum of precision recall and the best recall. The X axis shows the prediction of classes 1-19 and the Y axis is the actual classes 1-19 and the zero class. Figure 4 shows the precision recall of the best result of the sum of precision recall, precision and recall of all 19 classes.

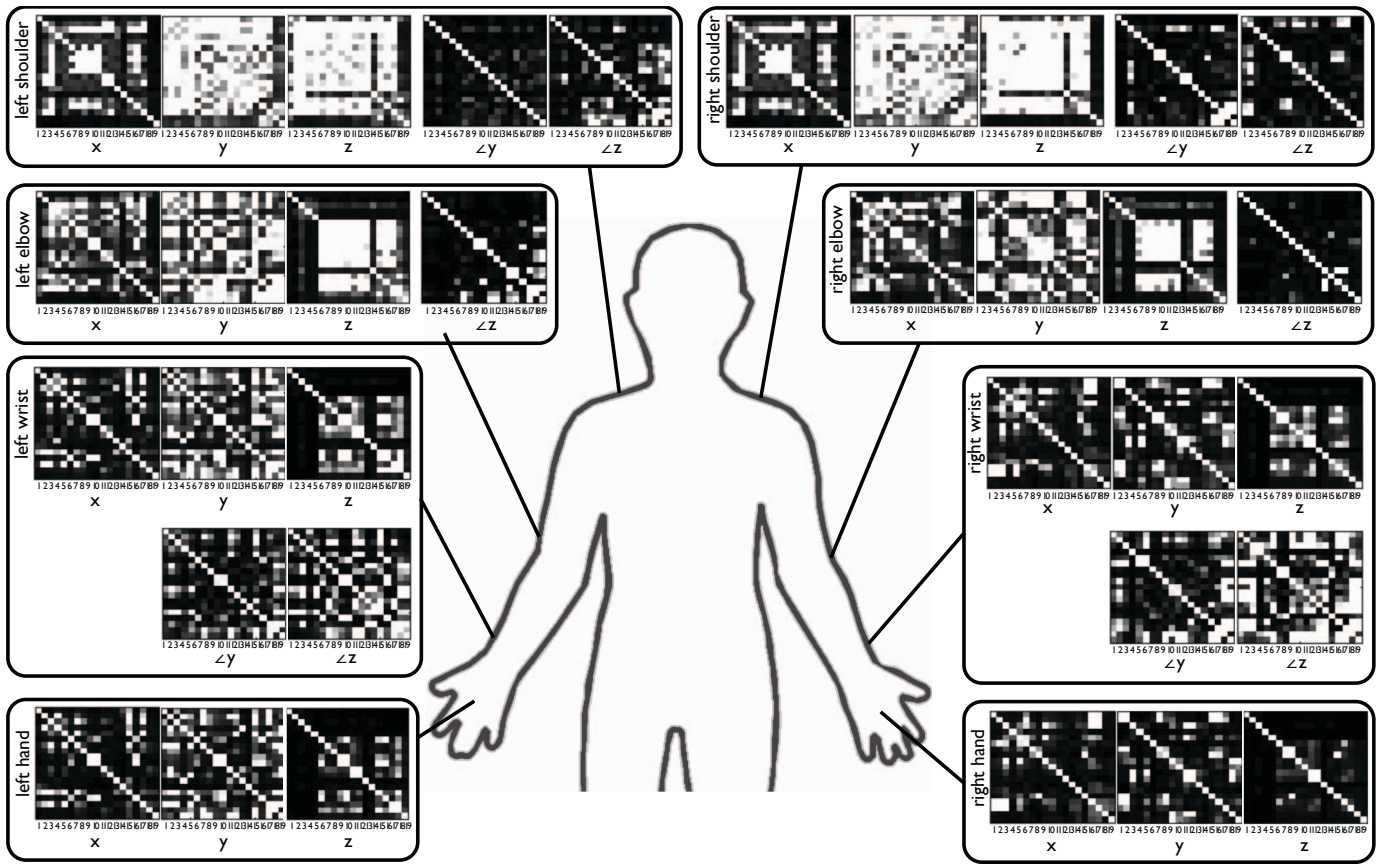


Fig. 2. Visual representation of cluster precision results for all pre-selected trajectories (marked by x,y,z) and angles (marked by a_y, a_z), over all 19 classes; Bright areas stand for high values (i.e., large inter-class similarities when these appear off the diagonal), whereas darker areas stand for low similarity values. Note that this type of inspection allows the pinpointing of several strong and weak sensors and features, as well as a rough overall evaluation of distinctiveness between classes for each of the sensors and features.

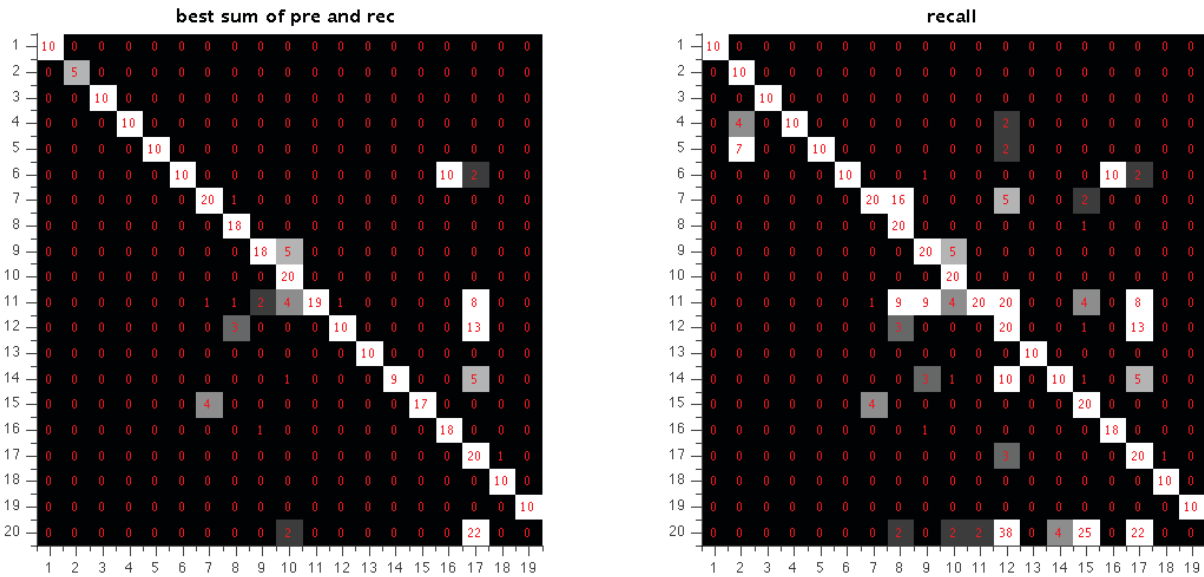


Fig. 3. Two examples of confusion matrices for the best classification results found so far with our approach.

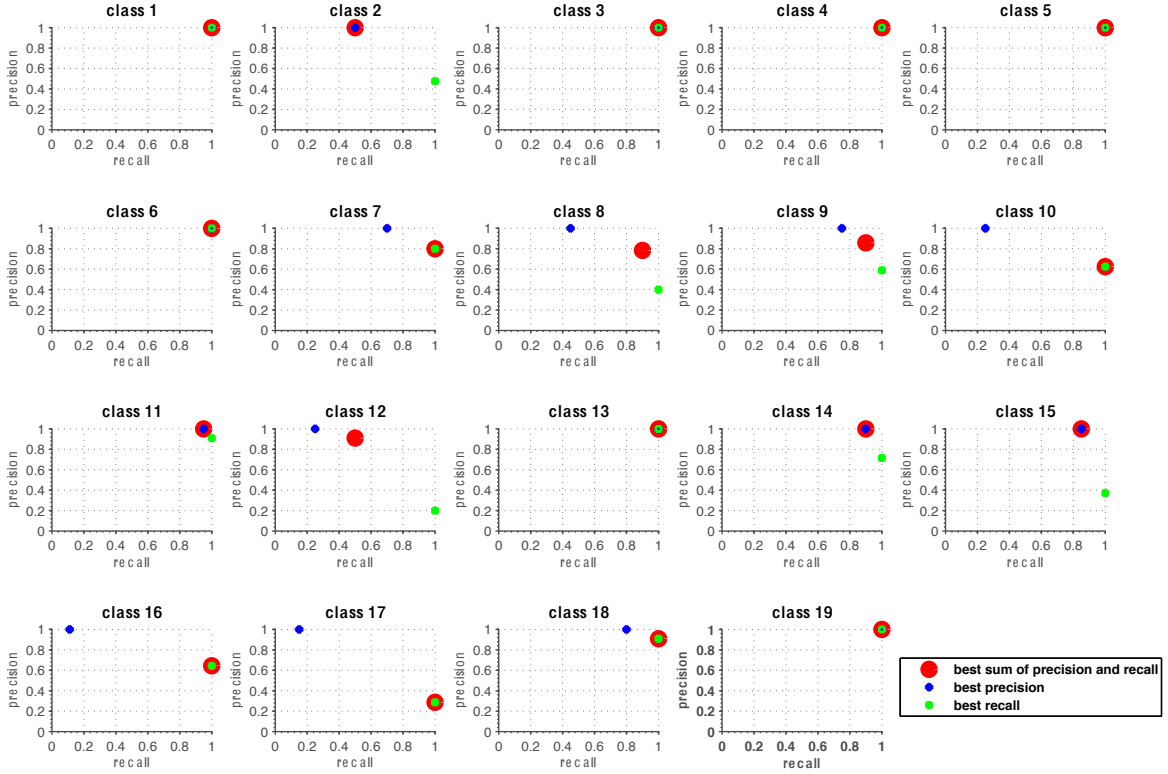


Fig. 4. precision recall per class of best classification result

IV. OBSERVATIONS AND DISCUSSION

As contributions of this paper, we argue that three types of observations can be used at the clustering level of the system:

A. Cluster precision insight of channel performance

To determine which channel has the most distinguishable character between the different classes the sum of all precision recall values returns an order which can be seen in Table II. Especially the z axis of the trajectory and the angles provide the best results. This shows that up and down movements seem to be quite characteristic for the activities of our dataset, which can be explained in the way, that a height variation of, e.g., the hand or the arm to open a door or the trunk is always part of such an action. As worst channels the shoulder in y and z direction is identified. This shows that up and side movements of the torso don't give much information about the conduction of the activities. Only the x axis of the shoulder ranks higher, demonstrating that leaning forward and backward can give hints to some of our activities.

B. Cluster precision insight of class performance

The second information we can extract of the cluster precision evaluation is how easy a class can be mixed up in the classification step with other classes. In Figure III the

class order can be observed. The value derives from summing up all cluster precision values of one class on all channels. Comparing the ordering with the precision recall plot in Figure 4 shows that the results of the cluster precision class order can be connected to the quality of the classification results.

C. Comparison of cluster precision information and classification results

Classes 1, 3, 4, 5 and 6 return very good classification result, which can also be observed in the cluster precision plots in Figure 2), as there can be seen on several cluster precision plots black rows in the area of the small numbered classes. For example the right hand tip z axis shows that these classes are very divergent in their cluster centers to the other classes. The precision recall plots of class 12 are one of the worst results. Inspecting the confusion matrix of the best recall case we see already several classes which are confused with class 12. In most of the channels of the cluster precision there are also several bright spots of class 12, showing that this class can have problems with mixing up with other channels.

In summary, we have shown in this section that three types of observations can be made at the clustering phase: 1) overall performance of single feature-sensor channels, 2) class mixing for the data from these channels, and 3) class-

TABLE II
ORDER OF BEST DISTINCTIVE CHANNELS DUE TO CLUSTER PRECISION

order	clust. prec. sum	channel
1	2277	right handtip z
2	2148	left handtip z
3	1470	left elbow angle z
4	1334	right wrist z
5	1253	right elbow angle z
6	914	right shoulder angle y
7	865	left wrist z
8	780	left shoulder angle y
9	653	right shoulder angle z
10	440	left shoulder angle z
11	351	right handtip y
12	326	right handtip x
13	318	left handtip x
14	278	right elbow z
15	225	right wrist x
16	220	right shoulder x
17	220	left shoulder x
18	214	left wrist x
19	207	right wrist y
20	206	right wrist angle y
21	175	left elbow z
22	173	left wrist angle y
23	168	left handtip y
24	139	right ellbow x
25	119	left wrist angle z
26	97	right wrist angle z
27	96	left wrist y
28	95	left elbow x
29	62	right shoulder z
30	52	right elbow y
31	36	left shoulder z
32	34	left ellbow y
33	29	left shoulder y
34	20	right shoulder y

mixing performance across all channels. This comes on top of the benefit of our approach that raw data sequences are retained as motifs and as such lend themselves very well for visualization (both of cluster centroids or individual sequences in the sensor data stream).

V. CONCLUSIONS AND FUTURE WORK

We presented a five-stage system that evolves around a hierarchical clustering. Central to our approach is a hierarchical clustering where the cluster centers are significant data segments which are used as raw data sequences in the further classification process. We demonstrated in this paper that the clustering gives an instant insight into sensor and feature quality and class inter-mixability, which can be easily represented graphically for visual inspection, prior to any classification evaluation.

TABLE III
ORDER OF BEST DETECTABLE CLASSES DUE TO CLUSTER PRECISION

order	clust. prec. sum	mean	std	class
1	1454	39	57	class 2
2	1328	43	72	class 1
3	1229	23	35	class 5
4	1220	36	53	class 4
5	1131	36	63	class 13
6	845	20	26	class 18
7	822	18	21	class 19
8	794	16	23	class 3
9	788	18	24	class 15
10	755	17	21	class 16
11	721	20	25	class 14
12	691	18	21	class 6
13	686	33	60	class 11
14	628	21	19	class 12
15	626	23	31	class 9
16	615	22	28	class 7
17	564	16	17	class 10
18	553	25	24	class 8
19	543	24	24	class 17

Future work on his approach entails the full evaluation of our approach with respect to classification performance. Since not all options are evaluated at the classification phase, we expect that the performance will increase further from the precision-recall values found in Figure 4. With the techniques discussed in this paper, we furthermore expect to also provide more insight in other datasets containing inertial data from multiple body-worn sensor units.

REFERENCES

- [1] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*, pages 511–518, 2001.
- [2] Mntyjrvi J., Himberg J., and Seppnen T. Recognizing human motion with multiple acceleration sensors. In *Systems, Man, and Cybernetics*, 2001. 2001 IEEE International Conference on Systems, Man, and Cybernetics (SMC2001). Tucson, Arizona, U.S.A., 7.10.-10.10.2001, pp. 2:747-752.
- [3] James Fogarty, Andrew J. Ko, Htet Htet Aung, Elspeth Golden, Karen P. Tang, and Scott E. Hudson. Examining task engagement in sensor-based statistical models of human interruptibility. In *Proc. CHI*, pages 331–340, 2005.
- [4] Ulf Blanke, Bernt Schiele, Matthias Kreil, Paul Lukowicz, Bernhard Sick, and Thiemo Gruber. All for one or one for all? combining heterogeneous features for activity spotting. In *PerCom Workshops*, pages 18–24, 2010.
- [5] Ricardo Chavarriaga, Hesam Sagha, and José del R. Millán. Ensemble creation and reconfiguration for activity recognition: An information theoretic approach. In *SMC*, pages 2761–2766, 2011.
- [6] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, pages 1–17. Springer, 2004.
- [7] Andreas Zinnen, Christian Wojek, and Bernt Schiele. Multi activity recognition based on bodymodel-derived primitives. *Location and Context Awareness*, pages 1–18, 2009.
- [8] Georg Ogris. *Multi-modal on-body sensing of human activities*. PhD thesis, University of Passau, 2009.
- [9] Thomas Stiefmeier. *Real-Time Spotting of Human Activities in Industrial Environments*. PhD thesis, ETH Zurich, 2008.