Daily Routine Recognition through Activity Spotting

Ulf Blanke and Bernt Schiele

Computer Science Department, TU Darmstadt, Germany
blanke@cs.tu-darmstadt.de,schiele@cs.tu-darmstadt.de

Abstract. This paper explores the possibility of using low-level activity spotting for daily routine recognition. Using occurrence statistics of low-level activities and simple classifiers based on their statistics allows to train a discriminative classifier for daily routine activities such as working and commuting. Using a recently published data set we find that the number of required low-level activities is surprisingly low, thus, enabling efficient algorithms for daily routine recognition through low-level activity spotting. More specifically we employ the JointBoosting-framework using low-level activity spotters as weak classiers. By using certain low-level activities as support, we achieve an overall recall rate of over 90% and precision rate of over 88%. Tuning down the weak classifiers using only 2.61% of the original data still yields recall and precision rates of 80% and 83%.

Key words: Activity Recognition, Wearable Computing, Human Routines

1 Introduction

Human activity is an important ingredient for context-aware systems. Its recognition has gained a lot of interest recently for diverse domains spanning from industrial applications to modeling of human behaviour in medical care. The different requirements of these domains have led to a multitude of approaches targeting recognition on different scales and complexities ranging from gesture recognition (happening within seconds) to complex daily routines (lasting often for hours and consisting of multiple activities).

Many state-of-the-art activity recognition approaches use and model all sensor data generated during the user's activity. For example a Hidden Markov Model can be used to model the sensor data of a particular activity. However, modeling and recognizing highly variable and in particular long-lasting activities such as daily routines (e.g., lunch or commuting activities) using such holistic approaches might be suboptimal and even infeasible in the presence of limited training data. In contrast, *activity spotting* aims to identify those activities that are most distinctive for a particular long-term activity. Identifying those distinctive parts of an activity and spotting them for recognition has at least two major advantages. First, the computational requirements can be reduced significantly and second, even limited amounts of training data can suffice to obtain good recognition performance.

Another predominant aspect of most recent activity recognition approaches is that the recognition of complex (high-level) activities follows a bottom up approach. The classification of low-level activities is used as a basis to model and infer more complex activities. This work takes rather the perspective of a top down approach. Having the annotation of high level routines, the aim is to identify low level activities, which have a high descriptive power to distinguish the routines. For this we use a feature selection algorithm to reveal the low level activities which matter most for classifying a routine.

The main contribution of this paper is to investigate the feasibility of spotting distinctive low-level activities to recognize complex activities, in our case daily routine activities. We achieve this by an automatic discriminative analysis of the routines and spotting the discriminative part of the routine's data. Experimental results show that far less data is needed than one might intuitively expect and that only a small subset of lowlevel activities suffice to support the classification of daily routines. As large portions of the data are not used and therefore irrelevant for activity classification, the approach is highly efficient. As a result the approach lends itself to computationally inexpensive embedded activity recognition.

The paper is organized as follows. First we situate our work within related work (section 2). After presenting our approach in detail (section 3), we introduce a publicly available dataset (section 4), on which the evaluation is done (section 5). We complete this paper with a conclusion (section 6).

2 Related Work

In human activity recognition, many different methods have been proposed focusing on different types, scales and complexities of activities. However, surprisingly little literature exists on long-term and high-level activity recognition and therefore remains an open research challenge. The following summarizes the most related work even though we are not aware of any prior work to use low-level activity spotting to recognize complex or high-level activities.

2.1 Layered Activity Recognition

A popular approach to activity recognition is a layered inference of complex activities based on prior classification of simple activities as subcomponents. In [7] the authors show that Hidden Markov Models (HMM) have the ability to capture different levels of abstraction with respect to time granularity. Constructing a cascade of HMMs, low-level video and audio data is used for a first HMM and its output is fed into a second set of HMMs. The evaluation is done on office activities, like *attending a presentation* or *making a phone call*.

In the work of Dong Zhang et al. [11] a two-level classifier is used on auditory and visual data to recognize human interactions patterns in meetings. In the first layer individual actions are recognized such as *speaking*, *writing*, and *idle*. The output of these models is then used as input for the second layer, which models group interaction such as *presentation*, *discussion*, *monologue*, and *white-board*.

Clarkson and Pentland [1] propose a method to discover *events* and *scenes* on ambulatory audio and video data. The authors find that clustering with regular HMMs only separates specific events (e.g, cashier beeps, supermarket music, walking through aisles), but does not capture the fact, that these events occur together in a scene. Using a hierarchy of HMMs, capturing first the lowlevel events, scenes such as being at the supermarket, at a busy street or in the video store can be discovered.

In [5], an approach is used to separate classes such as walking, jogging, or driving a car on sample level discriminatively. The posterior probabilities of the classification are then fed into an HMM, which captures the temporal context and hereby improving the results of the discriminative inference.

2.2 Activity Spotting

Following the paradigm of activity spotting, the aim is to find subcomponents of activities that allow to distinguish different classes. The authors of [6] approach the discovery of activities by identifying motifs in multivariate time series. In their context, motifs mean reoccurring subsequences, which have a high intra-motif similarity and can be distinguished from other subsequences. They use unsupervised techniques to identify these motifs. To evaluate their approach, they record a set of different dumbbell exercises.

In [12], gesture recognition on continuous streams is approached by identifying segments of interest. These simple segments are then combined as subparts of a certain gesture.

2.3 Longterm Activity Recognition

Research on long-term activity recognition is still in its infancy. Besides the difficulty of the task, also practical reasons have hindered progress: recording of long-term data is a non trivial task and annotating the data tends to be cumbersome and time consuming. In [4], a probabilistic model (originated from text-document analysis) is used to discover daily routines. By discretizing low level sensor data into a 'word'-type of representation, routines can be interpreted as 'documents'. Using topic models, the authors can discover daily routines such as going for lunch or commuting.

Others investigate the recognition of activities on a larger timescale such as shopping or doing housework without having low level activities as a prerequisite and using standard algorithms to infer labeled activities on larger timescales [3]. Van Laerhoven et al. [9] use a model of the user's rhythms to improve the recognition on a dataset of 27 days and compare their approach to the exclusive use of sensor data, such as motion, light and temperature. The model is constructed by a daily probability distribution of the user's annotated activities.

3 Activity Spotting for Daily Routine Recognition

The main goal of this work is to use low-level activity spotting to recognize high-level activities such as daily routines. Given a particular set of such high-level activities we therefore need to identify parts of the sensor data that enable reliable discrimination of the activities. For this we employ a boosting framework using low-level activity spotters as weak classifiers.

In this paper we use the occurrence statistics of low-level sensor data as the basic representation of sensor data. This representation closely follows the representation previously used for daily routine modeling and recognition [4]. We first perform unsupervised clustering of the low-level sensor data. For a given time-window we then compute the occurrence statistics of the cluster centers. This is the basic representation of the sensor and is used as input for the boosting framework. Figure 1 illustrates the steps in more detail. After extracting the features (1), distances to the cluster centers are computed (2) and the occurrence statistics are calculated (3). The result of boosting is a set of scores for the activity classes (4).

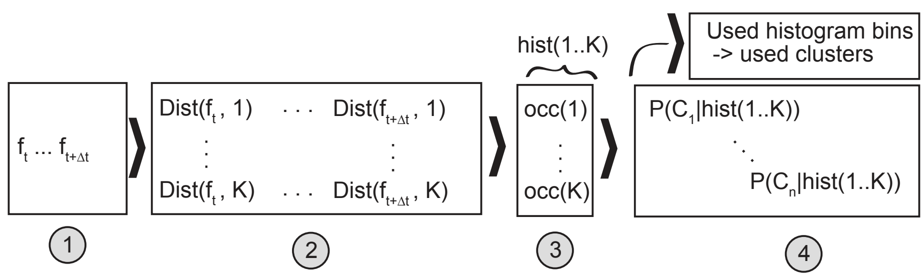

Fig. 1. (1) Feature extraction over a sliding window of raw sensor data. (2) All data points are (softly) assigned to cluster centers. (3) Occurrence statistics of the cluster centers are computed for the sliding window and stored in a histogram. (4) Each histogram is classified using JointBoosting resulting in scores, i.e., posterior probabilities for each activity. Note that boosting selects the most discriminative cluster centers (= histogram bins) for classification, thereby reducing the required sensor data substantially.

The following describes the major components in more detail. The next section describes the feature calculation (according to step (2) and (3) from Fig. 1). Then we introduce the feature selection method based on JointBoosting and describe how it is being used to automatically discover discriminative low-level activities.

3.1 Extracting low level events

Given activity data in form of acceleration data, we first cluster its features using K-means clustering. The obtained cluster centers are then used to assign each data sample the n nearest cluster centers. In the experiments below we use $n = 1$ for hard assignments as well as $n > 1$ for soft assignments. In this work we set $n = 3$. For soft assignment we use the standard softmax function:

$$w_i = \frac{\exp\{-\frac{d_i}{\sigma}\}}{\sum_{j=1}^{K} \exp\{-\frac{d_j}{\sigma}\}} \tag{1}$$

with d_i being the distance to the cluster and σ the standard deviation of all distances.

In the second step we calculate the occurrence statistics, by summing the soft assignment over a certain window length. This way we include the duration of a specific activity. Alternatively, to get the distribution of hard assigned clusters, we count the occurrence of each assigned cluster within the window. As time has proven to be a powerful cue for daily routine recognition, we added a time-of-day timestamp to the feature vector.

We use occurrence statistics and timestamps as input for the JointBoosting. v_0 corresponds to the timestamp and all other entries v_i correspond to one bin of the occurrence statistics, containing the sum of the weighted assignments of cluster center i within a window. Hence, the size of the feature dimension $|v|$ corresponds to the number of centers K plus one dimension for the timestamp. These are then labeled with the highlevel daily routines.

3.2 JointBoosting

In general, boosting [2] is used to combine a pool of weak classifiers to form a single strong classifier. The mechanism of boosting is interesting in many ways. Besides the classification improvement compared to individual weak classifiers, boosting can be used to find the most discriminant features [10]. In this work we employ regression stumps of the following form as weak classifiers:

$$h_m(v) = a \cdot \delta(v^f > \theta) + b \cdot \delta(v^f \leq \theta) \tag{2}$$

given a feature vector $\{v^f : f = 1, .., N_{features}\}$. θ is the optimal threshold being automatically found, so that h_m is positive if $v^f > \theta$ or negative if $v^f \leq \theta$. The Kronecker-δ results in 1 or 0, depending on the condition being true or false. Weak classifiers are combined additively to a strong predictor as defined:

$$H(v) = \sum_{m=1}^{M} h_m(v) \tag{3}$$

The number of weak classifiers M is also referred to as *rounds*.
The regression stump parameters

$$a = \frac{\sum_i w_i z_i \delta({v_i}^f > \theta)}{\sum_i w_i \delta({v_i}^f > \theta)} \tag{4}$$

$$b = \frac{\sum_i w_i z_i \delta(v_i{}^f \leq \theta)}{\sum_i w_i \delta(v_i{}^f \leq \theta)} \tag{5}$$

are computed from the weighted square error of the training data and can be seen as a weak classifier voting for or against a class. w_i denotes the weight for each training sample. In each round these weights are updated and increased for samples which are misclassified and decreased for samples which are correctly classified. This makes the training focus on harder training samples in future rounds. $z_i = \{+1, -1\}$ denotes the binary class membership label. Intuitively, a is the confidence in judging a sample positively, if the feature is greater than θ. b is the confidence this sample not being part of the class.

To analyze how well a weak classifier separates one class from the remaining classes, we take the difference of the regression stump parameters a and b:

$$vote_{confidence} = sign(a) \cdot |(a - b)| \tag{6}$$

The sign of a indicates wether this classifier votes for (when positive) or against (when negative) a specific class.

In [8], Boosting is extended by the ability to share features across different classes. The basic idea is to not only to separate between two classes but to find subsets of classes (for each weak classifier) that are best separated. This increases the computational cost during training and typically greedy a search is used to reduce training times [8]. During testing however, this extension allows to reduce the computational costs significantly as weak classifiers are shared across classes.

Besides the computational advantage, JointBoosting also allows the analysis which low-level activities can be used jointly to discriminate between multiple high-level activities. How the parameters of the JointBoosting can be interpreted to allow such an analysis is described in section 5.2.

4 Dataset

To test our approach of selecting the most significant low level subcomponents of an activity, we evaluate it on a realistic and representative dataset of activities and routines of daily living. We briefly describe the dataset, and refer to [4] for further details on the recording.

Routine	Duration
Dinner	217.5 min
Commuting	289.0 min
Lunch	391.3 min
Office Work	2814.7 min

Table 1. Daily routines observed over seven days

Two 3-axis-acceleration sensors, sampling at a rate of 100Hz are worn, one at the wrist and one in the right pocket. As features, mean and variance are calculated over a window of 0.4s, resulting in approximately 2.5Hz. The dataset contains 7 days of continuous data leaving out the sleeping phases. Table 1 shows the four annotated daily routines. These contain different types of low-level activities. The routine *dinner* for instance groups low-level activities such as *preparing food*, *having dinner (eating)*, and *washing dishes*. In total the data set contains 34 labeled low-level activities of which a subset of 24 activities occurred during the routines. A complete list of the low level activities occurring during these routines is provided in Table 2. The total duration of each activity is given as well as the mean duration for each instant of one activity.

Activity	Average duration	Occurrences	Total
sitting / desk activities	49.41 min	54	3016.0 min
unlabeled	1.35 min	239	931.3 min
having dinner	17.62 min	6	125.3 min
walking freely	2.86 min	38	124.2 min
driving car	10.37 min	10	120.3 min
having lunch	10.95 min	7	75.1 min
discussing at whiteboard	12.80 min	5	62.7 min
attending a presentation	48.9 min	1	48.9 min
driving bike	11.82 min	4	46.3 min
walking while carrying something	1.43 min	10	23.1 min
walking	2.71 min	7	23.0 min
picking up mensa food	3.30 min	7	22.6 min
sitting / having a coffee	5.56 min	4	21.8 min
queuing in line	2.89 min	7	19.8 min
using the toilet	1.95 min	2	16.7 min
washing dishes	3.37 min	3	12.8 min
standing / having a coffee	6.7 min	1	6.7 min
preparing food	4.6 min	1	4.6 min
washing hands	0.32 min	3	2.2 min
running	1.0 min	1	1.0 min
wiping the whiteboard	0.8 min	1	0.8 min

Table 2. Low level activities occuring during the routines

5 Experimental Results

This section reports on experimental results for the dataset introduced in section 4. We begin with a quantitative analysis of the algorithm's performance and compare the results with [4] (section 5.1). Then we discuss in more detail the classifiers obtained with JointBoosting (section 5.2). In particular we give an interpretation which part of the low-level activities are characteristic and therefore chosen by JointBoosting to recognize daily routines.

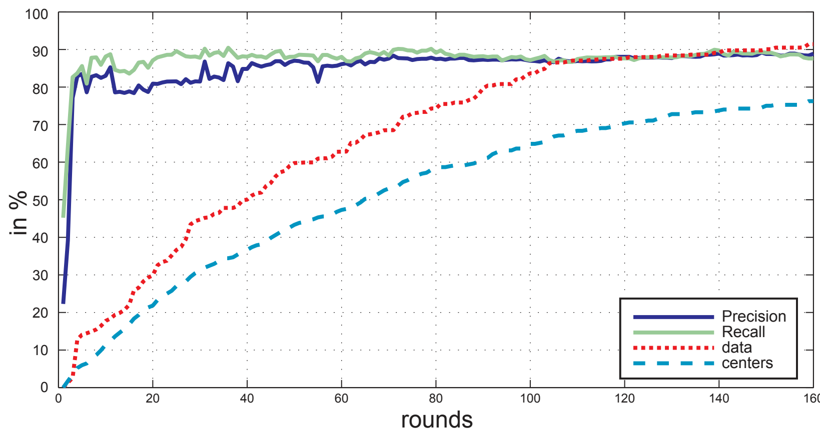

Fig. 2. Overall recognition results, amount of used data using different numbers of JointBoosting rounds and soft assignments of K-means clusters. *Centers* denotes the percentage of selected features of all given features (= K-mean-centers) and *data*, the percentage of their assignments in the data.

5.1 Quantitative Results

To make our results comparable to published results on the same dataset we aimed to follow the original parameter settings as closely as possible [4]. Therefore we used the following parameters: for K-means clustering we set K to 60 and we calculated histograms over a window of 30 min with an overlap of 5 min. As mentioned previously we also added a dimension for the time-stamp resulting in a feature vector with dimension 61. We evaluated the approach on a seven-fold crossvalidation, training a classifier on 6 days and leaving one day out for testing.

Overall Performance The overall recognition results for different numbers of rounds are given in Fig. 2. The highest precision and recall rates are 87.81% and 90.17% at 80 rounds. Using more rounds results in a slight decrease in recall and a small increase in precision. Table 3 shows several results in more detail.

Using 4 rounds, the classifier yields a precision of 82.85% and a recall of 83.67% by using 12.78% of the given data. Applying boosting with 10 rounds increases the recall by about 5%, but has no noticeable effect on precision. Increasing the rounds to 80 improves the precision about 5% and recall about 6.5%. Here more then half of the clusters (57.93%) and nearly 3/4 of the data (74.30%) are used.

Table 3 also shows a comparison of our results to the probabilistic approach given in [4]. It can be seen that our discriminant approach yields clearly better results. What is quite surprising though is that already using but four boosting rounds (i.e. four low-level activity spotter) allows to achieve a clear gain in overall performance. Table 4 gives a more detailed comparison by not only showing the overall performance but also showing the performance for each routine separately. Here we compare the results of [4] with the results obtained with 4 and 80 boosting rounds. Using 80 rounds always outperforms the approach of [4] with one exception: precision for *commuting*. Again it is surprising that using 4 boosting rounds outperforms the approach of [4] for three out of four daily routine activities and only obtains lower precision and recall for *commuting*. This clearly shows the applicability of the proposed approach of low-level activity spotting for daily routine recognition.

Soft Assigments to K-means centers						
Rounds	4	10	20	80	160	Huyhn et al [4]
Precision	82.85%	82.98%	80.91%	87.81%	**88.87**%	76.90%
Recall	83.67%	88.12%	87.09%	**90.17**%	87.16%	65.80%
Used Centers	**5.20**%	11.39%	15.70%	57.93%	76.25%	-
Used Data	12.78%	17.74%	21.81%	74.30%	91.41%	-

Hard Assigments to K-means centers					
Rounds	4	10	20	80	160
Precision	72.71%	77.34%	82.19%	86.40%	**88.37**%
Recall	82.67%	82.39%	87.09%	**90.32**%	89.18%
Used Centers	**5.19**%	12.14%	22.26%	50.82%	56.13%
Used Data	**2.11**%	4.94%	14.27%	45.42%	49.75%

Table 3. Overall recognition results for different numbers (4, 10, 20, 80 and 160) of rounds of the JointBoosting algorithm using soft- (top) and hard-assigments (bottom) to the K-means cluster centers.

	4 rounds		80 rounds		Huyhn et al [4]	
Routine	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
Dinner	84.31%	**100.0%**	**85.27%**	90.48%	56.90%	40.20%
Commuting	70.04%	60.27%	81.77%	**82.36%**	**83.50%**	71.10%
Lunch	78.85%	81.79%	**84.56%**	**90.04%**	73.80%	70.20%
Office Work	97.86%	92.61%	**98.12%**	**93.63%**	93.40%	81.80%

Table 4. Results per Routine using soft assignment of K-means cluster centers.

Fig. 3 illustrates the classification on one day of the seven day dataset using the same number of boosting rounds as in Table 3. It can be seen that the borders of two routines are seldomly precise. For instance, using 10 boosting rounds the *lunch* routine is predicted before it actually happens (which is reflected in a lower precision). However, the transitions from one routine to another happens smoothly, that is, there is often no exact start and end point. As the obtained classifiers also show the same smooth transitions between the routines we think that the results are indeed sufficient for many applications.

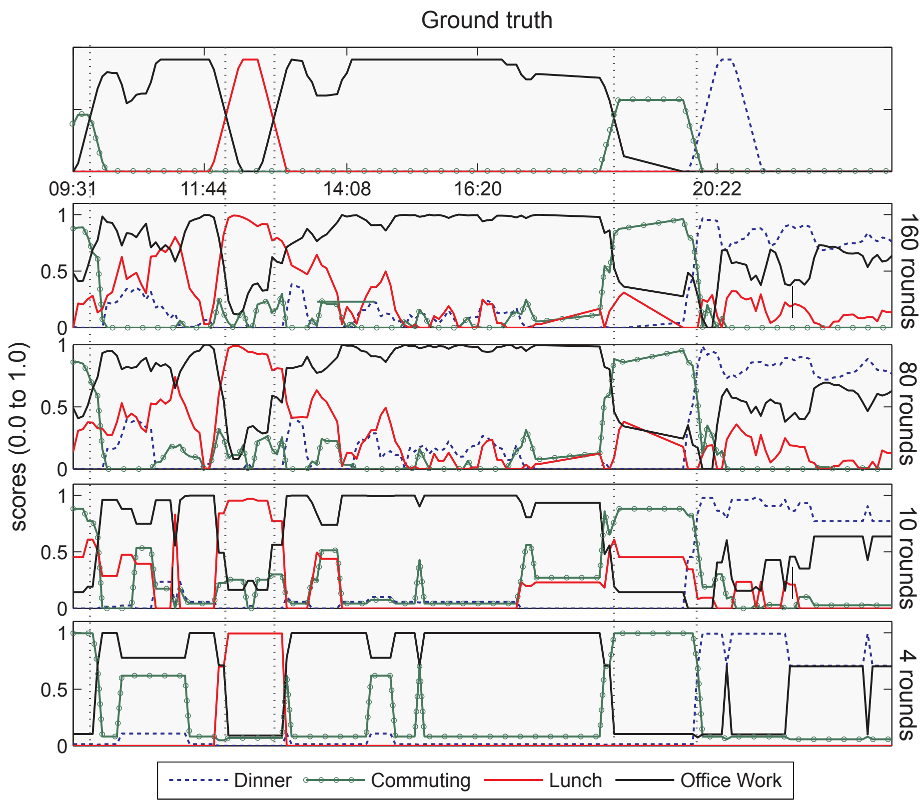

Fig. 3. Top: Ground truth for one day. Below the classification by the JointBoosting algorithm using 160, 80, 10 and 4 rounds.

So far we reported results for soft assignments of the clusters. As hard assignments have the potential to reduce the amount of data that needs to be considered for classification we also evaluated hard assignments of the K-means centers(, i.e., each sample is assigned to exactly one cluster). The overall results using 1 to 160 rounds are depicted in Fig 4.

We note the maximum at 40 rounds with 87% precision and 93% recall using 30% of the data and 37% of the centers. Table 3 shows the results in more detail. At 4 rounds we obtain a precision of 72.71% and a recall of 82.67%. This is a decline of 15% of precision and about 11% recall compared to the maximum.

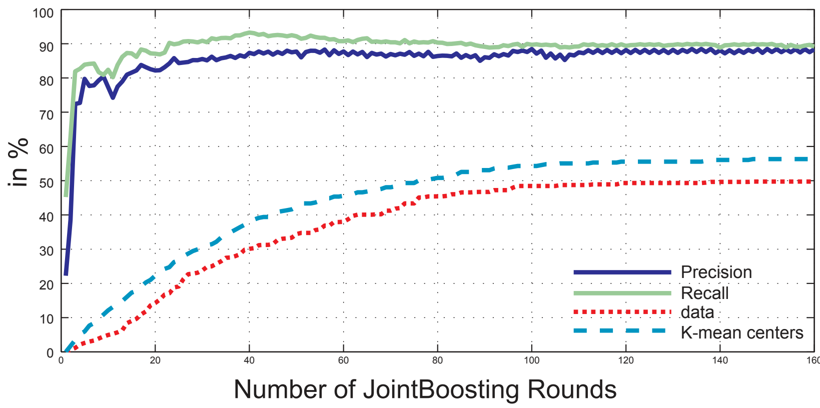

Fig. 4. Overall recognition results and amount of used data using different numbers of JointBoosting rounds and hard assignments of K-means clusters. *K-mean centers* denotes the percentage of selected features of all given features (= K-mean-centers) and *data*, the percentage of their assignments in the data.

However, the usage of data with 2.11% is minimal and considerably lower than for soft assignments. Adding one round, already improves the result to a precision of 79.75% and a recall of 83.40% using less than 1% more data (2.6%). This is due to the fact that only the major vote, that is the winning K-means centers can be observed even if some others have almost the same distance. Preserving the information of distances to a few centers using a soft assignment seems to be beneficial for lower number of classifiers.

We also have started to experiment with smaller window sizes. E.g. using a window length of 15 min results in a reduced precision (for 4 rounds) of 71.06% but with only a small decrease in recall (80.07%). Such a decrease in performance is in agreement with previous results [4]. As this is an important and interesting direction to pursue we plan to investigate this further.

5.2 JointBoosting Classifier Discussion and Visualization

Moving from activity recognition into the paradigm of activity spotting our goal is to analyze which and how much data needs to be observed to reliably recognize daily routines such as the ones used above. Applying JointBoosting, we get a selection of the most important features to discriminate the given classes. As JointBoosting selects the most important (= discriminative) occurences of a specific cluster, we can identify which parts of low-level data is important for the daily routine classification task.

To identify the amount of data used for classification, we count how often these clusters occur during the total length of the data. For soft assignments we look if the selected clusters appear in the top 3 assignments. While soft-

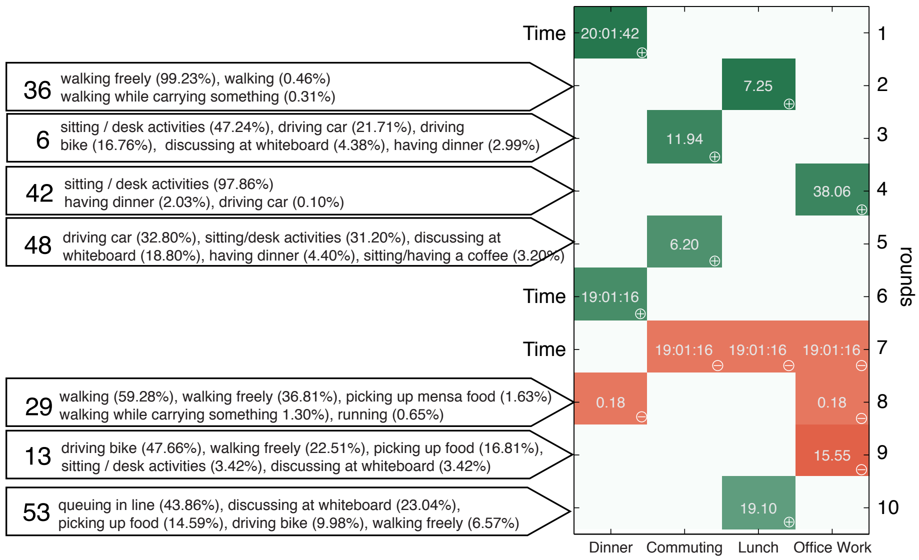

Fig. 5. The used features (occurrences of K-means centers and time) per round and per class, starting from the top with the first round. For each K-means center the label distribution of low-level activities are denoted. The threshold of the weak classifier is given in the colored boxes. Green (/plus) marked classifiers indicate a positive voting for the routine, red (/minus) a negative voting.

assignment increases the amount of data used it also has the potential to reduce the number of weak classifiers required for good performance.

For a better understanding which data is selected we use the provided low-level activity annotation and assign to each cluster a set of low-level activity labels. In the following we use the distribution of the top five assigned labels to each cluster center.

Fig. 5 visualizes the first ten weak classifiers chosen by boosting. The color indicates the confidence of the weak classifier, predicting the sample to be from a specific class. The corresponding thresholds are given inside the boxes. Positive weights are colored green, respectively marked with a plus-symbol, negative weights are colored red and marked with a minus-symbol. Intuitively the reader can interprete the color as a voting for a specific routine when colored green or against it when colored red. The intensity of the color represents the absolute difference between the regression stump parameters a and b, given in equation 6 (Section 3.2).

Starting with round one from the top (one weak classifier), it separates *dinner* from the rest using time as feature. In the second round, the weak classifier separates *lunch* from the rest by observing the occurence of cluster 36 (= feature 36). This cluster corresponds mostly to *walking freely* which turns out to be more discriminant than other low-level activities such as *having lunch*. This can be easily explained by the fact that *having lunch* is too similar to other low-level activities such as *having dinner* or *sitting/desk activities* and is therefore not chosen by boosting. The third classifier separates *commuting* from the other classes using cluster 6. This cluster is dominated by activities occurring during commuting: *sitting/desk activities*, *driving car* and *driving bike*. In the forth round cluster 42 is chosen to classify routine *office work* which mostly contains *sitting/desk activities*. Note that the threshold shows a comparatively high value, which means the activities of this cluster have to appear often.

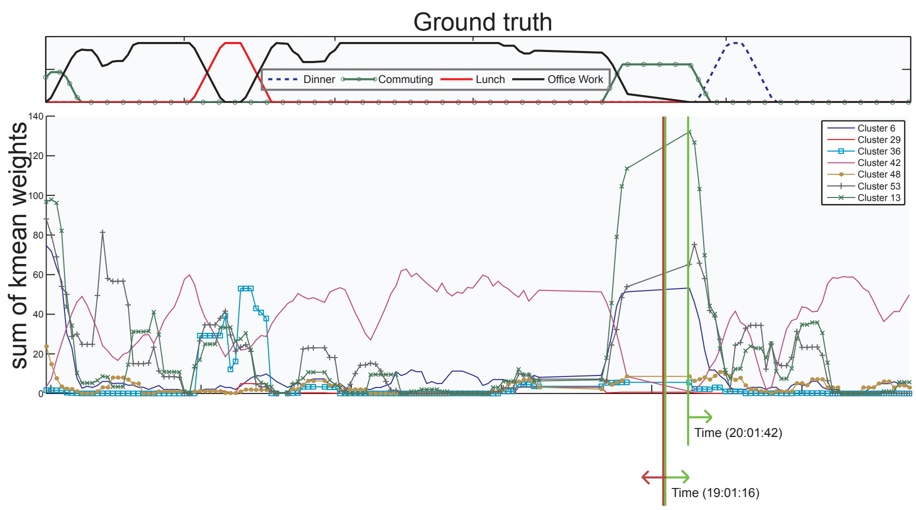

Fig. 6. The top feature occurrences during one day. (Top) The groundtruth for one day. (Bottom) The sum of weights of the soft assigned kmean-centers for each sample. The ranking of the selected features is as follows: *20:01:42 (time), 36, 6, 42, 48, 19:01:16 (time), 29, 13, 53.*

In the eighth round cluster 29 is used to separate *dinner* and *office work* from *commuting* and *lunch*. More specifically the classifier votes against *dinner* and *office work*, if the given data sample is bigger than the treshold. The threshold of 0.18 indicates that the occurrence of this activity is fairly low, but enough to discriminate between classes. For each feature, i.e. the cluster center the top five low level-activities (Table 2) are given to get an intuition why the cluster is used. It can be seen that cluster 29 shares the labels *walking, walking freely, picking up mensa food* etc. Due to the fact that these activities only happen during *lunch* and *commuting* it is a good cue that *dinner* and *office work* are not happening.

An interesting observation is that the first 4 rounds separate each routine without sharing any features. Time turns out to be sufficient as feature to separate *dinner* from the others. After 8:00pm the confidence that *dinner* is the current routine is very high. On the other hand, looking at round 7, *commuting*, *lunch* and *office work* is less likely to happen after 7pm and added as another weak classifier.

We did not compare JointBoosting with Adaboost quantitatively. However, looking at the first 10 rounds depicted in Fig. 5, note that fewer features are shared as expected. Given this dataset, jointBoosting ranks those features first, which are discriminative enough for each class individually. This leads to the assumption that jointBoosting might yield similar results compared to AdaBoost for a low number of weak classifiers. However we observe at higher rounds that more features are shared. E.g., in rounds 10 to 20 seven features are jointly used, which is not possible with AdaBoost. JointBoosting therefore leads to a computationally more efficient solution.

Low Level Activities Figure 6 illustrates the occurrences over time of the selected features in the tenth round. We encourage the reader to use Fig. 5 as reference to lookup the low level activities per cluster. Given the routine *office work* one can observe that the occurrence of cluster 42, which is selected as a weak classifier for this routine, is most of the time above the threshold of 38.06 to vote for this routine. One can see that cluster 53 and 42, which classify *lunch* and *office work* happens also during the dinner routine. However none of those routines happen after 19:01:16, whereas dinner is likely to happen after 20:01:42 and the combination of each of these classifiers leads to a higher posterior probability of the routine *dinner*. Cluster 53 is used by a weak classifier voting for the *lunch* routine. It is mainly assigned to the low level activity *queuing in line*. Although this cluster has a high occurrence during the *office work* in the morning before lunch, its confidence is weaker than the confidence of cluster representing sitting at the desk, yielding a higher score for *office work than* lunch.

Observing the assigned labels to the used clusters, it can be noticed that, e.g, for the routine *lunch* it is not the activity *having lunch* discriminant. But activities which surround the actual activity of this routine, *which are walking (freely), queuing in line and picking up cafeteria food*. The explanation is simple - as *having lunch* is virtually the same as *having dinner* or similar to *sitting at the desk* and can be easily confused. Whereas walking or standing in line at a certain time can be very discriminative.

5.3 Discussion

The previous sections showed that discriminative classification and spotting of low-level activities yields promising results when applied to complex and highly variable daily routines. The proposed approach achieves good overall results with recall and precision above 80% in general to recognize the presented daily routines. Using a small number of classifiers (that is a small number of JointBoosting rounds) already achieves good performance. It is important to note that by using small numbers of classifiers, it is possible to reduce the required data substantially. Using only 2.6% of the data (i.e., 5 rounds and hard assigment to K-mean centers), it is possible to distinguish between different routines with a recall of 79.75% and a precision of 83.40%. We can determine a tradeoff between used data and performance, which favors using less data, as performance decrease is marginal.

6 Conclusion

This paper investigates the possibilities of *activity spotting* to recognize complex activities, like daily routines consisting of a series of low level activities. A top-down perspective is taken using a feature selection algorithm and the annotation of those higher-level routines to automatically spot discriminant low level data, respectively activities.

Our findings show that the approach of *activity spotting* leads to viable results. It could be seen that a surprisingly small amount of data is discriminant enough to differentiate routines such as *lunch* or *commuting* and yield to results that are better than those of current research. Our algorithm discovers successfully meaningful low level activities, which can be intuitively connected to the high level routines.

Applying this approach we are able to filter insignificant data, which does not support recognition. Once the training is done offline, the most discriminant spots in the data are obtained, which the classifier has to process. This allows to reduce the computational costs in the classification task, while the distance to only a few K-means centers needs to be calculated. Thresholding the distance filters the unimportant data. Additionally, we reduce the amount of data which has to be stored. Implementing such classification on embedded systems will greatly benefit from this reduced complexity.

Acknowledgments. This work has been supported by the DFG Graduate School Topology of Technology. The authors of this paper gratefully acknowledges Christian Wojek for providing them with a JointBoosting implementation.

References

1. B. Clarkson and A. Pentland. Unsupervised clustering of ambulatory audio and video. In *ICASSP*, 1999.
2. J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann. Statist*, 28(2):337–407, 2000.
3. T. Huynh, U. Blanke, and B. Schiele. Scalable recognition of daily activities with wearable sensors. In *3rd International Symposium on Location- and Context-Awareness (LoCA).* Springer, 2007.
4. T. Huynh, M. Fritz, and B. Schiele. Discovery of activity patterns using topic models. In *Proceedings of the 10th International Conference on Ubiquitous Computing (Ubicomp)*, 2008.

5. J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.
6. D. Minnen, T. Starner, I. Essa, and C. Isbell. Discovering characteristic actions from on-body sensor data. In *Proceedings of the 10th IEEE International Symposium on Wearable Computers (ISWC)*, 2006.
7. N. Oliver, E. Horvitz, and A. Garg. Layered representations for human activity recognition. In *In Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, 2002.
8. A. Torralba, K. Murphy, and W. Freeman. Sharing visual features for multiclass and multiview object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
9. K. Van Laerhoven, D. Kilian, and B. Schiele. Using rhythm awareness in long-term activity recognition. In *Proceedings of the 12th IEEE International Symposium on Wearable Computers (ISWC)*, 2008.
10. P. Viola, P. Viola, M. Jones, and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition (CVPR)*, 2001.
11. D. Zhang, D. Gatica-Perez, S. Bengio, I. McCowan, and G. Lathoud. Modelling individual and group actions in meetings: a two-layer hmm framework. In *Computer Vision and Pattern Recognition Workshop*, 2004.
12. 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.