Analyzing Features for Activity Recognition

Tâm Huynh and Bernt Schiele

Multimodal Interactive Systems, TU Darmstadt, Germany
{tam, schiele}@informatik.tu-darmstadt.de

Abstract

Human activity is one of the most important ingredients of context information. In wearable computing scenarios, activities such as walking, standing and sitting can be inferred from data provided by body-worn acceleration sensors. In such settings, most approaches use a single set of features, regardless of which activity to be recognized. In this paper we show that recognition rates can be improved by careful selection of individual features for each activity. We present a systematic analysis of features computed from a real-world data set and show how the choice of feature and the window length over which the feature is computed affects the recognition rates for different activities. Finally, we give a recommendation of suitable features and window lengths for a set of common activities.

1. Introduction

Context awareness is a central issue in ubiquitous and wearable computing. The opportunity to perceive the world from a user's perspective is a key benefit of wearable systems compared to stationary, desktop-centered computers. While context information can consist of any information describing the situation of the user, for many applications the current activity and location of the user are considered to be highly important. In this paper we focus on selecting features for activity recognition using wearable sensors.

Activities such as walking, standing, sitting and jogging naturally lend themselves to recognition using acceleration sensors, since these activities are clearly defined by the motion and relative positions of the user's body parts. Being small and cheap, acceleration sensors can easily be integrated into accessories such as mobile phones, cameras or wrist watches that the user carries around. 2D- and 3D-acceleration data has been successfully used for activity recognition by various groups [1, 3, 5, 4, 6, 8, 9]. (For a more comprehensive overview, see, e.g. [1].)

Popular features computed from the acceleration signal are mean [1, 3, 2, 4, 9], variance or standard deviation [3, 2, 6, 9], energy [1, 9], entropy [1], correlation between axes [1, 9] or discrete FFT coefficients [4]. Energy and entropy are usually derived from the latter. [5] uses peaks in raw data; [8] uses powers of wavelet coefficients. The window length over which the features are computed is usually fixed, e.g. 6.7 sec in [1], 1 sec in [3], ~2 sec in [8], 8 sec in [4] and 5.12 sec in [9].

Comparing the different approaches to activity recognition, we observed that a common approach is to decide on a fixed set of features and a fixed window length and use this combination for the whole set of activities to be recognized. Even though the resulting recognition rates can be generally high, they might be improved by selecting features and window lengths for each activity separately.

In this paper we propose to use a simple measure of cluster precision to rank individual features according to how suitable they are for recognition of a given activity. We then show that the ranking obtained from this cluster analysis directly translates to recognition results, therefore validating the proposed cluster analysis.

The rest of the paper is organized as follows. Section 2 describes the data set we used for our work. In section 3, we describe the features that we computed and the cluster analysis we applied on them. Furthermore, we show that there is a direct correspondence between the cluster precision and the recognition rate for a given activity. In section 4, we report on our recognition results and discuss the impact that different features and window lengths have on the recognition rates. Finally, in section 5 we summarize our results and draw some conclusions.

2. Data Set

For our experiments, we used data recorded by Intel Research, Seattle [7]. The subset we used consists of roughly 200 minutes of sensor data recorded by two subjects who are not affiliated with the researchers. The subjects were given a script containing the activities to perform, namely walking, standing, jogging, skipping, hopping and riding bus. They recorded these activities in everyday life situations without supervision of a researcher. Later the data was annotated with the help of recorded video and audio data.

The data was recorded using an integrated sensor board developed by Intel Research that was attached to the shoulder strap of a backpack the subjects were carrying. The board contains sensors for 3D-acceleration, audio, temperature, IR/visible/high-frequency light, humidity and barometric pressure, as well as a digital compass.

3. Cluster Analysis

Clustering is a method to uncover structure in a set of samples by grouping them according to a distance metric. Our rationale for this study was that if the groups produced by the clustering algorithm were homogeneous in terms of the activity their members were labeled with, this would be a strong indicator that a recognition algorithm would be able to differentiate between the different activities. Thus, in this section we propose a simple measure of cluster homogeneity and use it to rank different features according to the quality of the resulting clustering. Then we show that this measure is indicative of recognition performance by feeding the features into a simple classifier. We use a simple classifier because the purpose of the classifier is not to yield high recognition rates, but only to show that the results of the clustering can be used to decide on features for recognition.

In this paper we were interested in evaluating the performance of individual features for activity recognition. For this reason we confined ourselves to one-dimensional features, both for the clustering and the subsequent recognition. Using the knowledge of the suitability of single features, one can later easily combine them to form higher-dimensional features and/or use them in a more elaborate classifier scheme such as the popular AdaBoost or SVM frameworks.

3.1. Features

In this study we were mainly interested in features derived from the accelerometer, as these have been successfully used for the activities we are considering (see introduction). In order to be able to study the effect of different windows lengths, all acceleration features were computed on windows of 128, 256, 512, 1024 and 2048 samples. At a sampling rate of 512 Hz, the windows correspond to 0.25, 0.5, 1, 2 and 4 seconds respectively. The windows were shifted over the data in steps of 0.25 seconds. For each window, we computed the magnitude of the mean, variance, energy, spectral entropy, as well as the discrete FFT coefficients. The FFT coefficients were grouped in six exponential bands, and another 19 features were obtained by pairwise addition of coefficients 1&2, 2&3,…, and 19&20. In addition to that, we computed three features representing the pairwise correlation of the acceleration in x-, y- and z-direction. Apart from the acceleration features, we included the variance of the digital compass and the visible light sensor.

3.2. Clustering

The feature computation yielded a set of about 50,000 samples for each feature. We performed k-means clustering on these sets, using a five-fold cross validation as follows: Each set was randomized and divided into five equally sized partitions. K-means was applied to four of the five partitions, the fifth being left for testing. Testing was done by assigning each sample in the test partition to the nearest cluster centroid produced by k-means. The result was a set of 1000 clusters for each of the five passes, each set covering about 10,000 samples. We also used other numbers of clusters (e.g., 100), but the results vary relatively little.

Ideally, each cluster would contain samples of only one activity. This would indicate that the data of the given feature was clearly separable and thus well-suited as an input for classification. In the worst case, the fraction of samples of an activity in each cluster would be equal to the a priori probability of the activity. This would imply that the feature was not discriminative for the given set of activities and thus unlikely to be suited for recognition.

In order to measure the distribution of samples for different activities in the clusters, we first computed for each cluster *i* and activity *j* the fraction

$$p_{i,j} = \frac{|C_{i,j}|}{\sum_j |C_{i,j}|} \quad (1)$$

where $C_{i,j}$ is the set of samples in cluster *i* labeled with activity *j*. We then formed a weighted sum of these fractions to obtain a cluster precision p_j for each activity *j*:

$$p_j = \frac{\sum_i p_{i,j} |C_{i,j}|}{\sum_i |C_{i,j}|} \quad (2)$$

Thus, if an activity has a cluster precision close to one, this indicates that there are many clusters mainly consisting of samples for this activity. We weight each fraction by the number of samples it represents in order to prevent smaller clusters from dominating the result.

Figure 10 shows the average cluster precisions over the five passes of cross validation for all features that we computed. Each plot holds data for one activity, and each line in a plot represents one window length. (Note that the lines connecting the different values are only drawn for better readability, as the function is only defined for discrete values.) The horizontal line marks the a priori probability of the activity.

3.3. Analysis

The precision plots in Figure 10 show a clear difference between the stationary activities ‘standing’ and ‘riding bus’ (which consisted mainly of sitting in the bus) and the moderate to high intensity activities, namely ‘walking’, ‘jogging’, ‘hopping’ and ‘skipping’. The variance in the cluster precision of different features is much higher for the activities with moderate to high intensity levels. Not surprisingly, for these activities the FFT features are clearly better than most of the other features. However, there is much variation between the cluster precision of the different FFT coefficients. E.g., for ‘skipping’, the cluster precision between FFT coefficients 13&14 and 15&16 drops from 0.9 by almost 80% to 0.12. Similar differences in precision can be observed for ‘hopping’ and ‘jogging’. Furthermore, there is no FFT coefficient that outperforms the others for all activities. The coefficients 1&2 are among the top five features for ‘walking’, ‘hopping’ and ‘riding bus’. Coefficients 2&3 have the highest precision of all features for ‘skipping’, ‘walking’ and ‘riding bus’. Coefficients 3&4 attain high precision for ‘jogging’, ‘hopping’ and ‘riding bus’. For all other coefficients, no clear statement across multiple activities can be made. Instead, one has to take a close look at each activity to see which coefficients are best. For ‘Standing’, coefficients 12 to 16 and 7 to 8 perform best, for jogging coefficients 3&4, and for hopping coefficients 7&8. For ‘walking’ and ‘riding bus’ variance does remarkably well, being in third and fourth place, respectively. For ‘walking’, ‘riding bus’ and ‘hopping’, the third exponential FFT band might serve as a compromise to the FFT coefficients, since it ranks among the first five features for these activities.

Comparing the different window lengths to each other, we observe that for ‘walking’, ‘jogging’ and ‘riding bus’, the 1 second window attains the highest precisions. For ‘skipping and hopping’, the 2 and 4 second windows score best, while the 0.25 and 0.5 second windows attain relatively low precision for all features of these two activities. For ‘standing’, the short windows of 0.5 and 0.25 seconds achieve high precision for a range of FFT coefficients. The longer windows of 2 and 4 seconds are not suited for ‘standing’ – the precision for these window lengths is quite low. In contrast to this, ‘jogging’ has some peaks with more than 80% precision for 2 and 4 second windows. The 0.25 and 0.5 second windows work not very well for jogging, except for the FFT coefficients 1&2.

When looking at features and window lengths combined, the following are the best combinations for each activity: ‘Hopping’: (FFT coeff. 7&8/ 4.0 sec); ’Skipping’: (FFT coeff. 2&3/ 2.0 sec); ’Jogging’: (FFT coeff. 3&4/ 1.0 sec); ’Riding Bus’: (FFT coeff. 2&3/ 1.0 sec); ’Walking’: (FFT coeff. 2&3/ 1 sec); ’Standing’: (FFT coeff. 12&13/0.5 sec.).

An important result of this analysis is therefore that there are features and window lengths which perform quite well for different activities, but in order to achieve best performance one should choose features separately for each activity.

4. Recognition

If the order imposed by the cluster precision values translates to recognition performance, the cluster analysis of the previous section could serve as a valid method to select suitable

features for recognition. To find out, we built a simple classifier by dividing the feature samples into training and test sets in the same fashion as we had done for the clustering, then applied k-means clustering on the training set and labeled the centroid of each training cluster i with the dominating activity $\bar{j}$ in the cluster, i.e.

$$\bar{j} = \arg\max_{j}(p_{i,j}) \quad (3)$$

Each sample of the test set was then either classified according to the label of the nearest centroid i, if $p_{i,\bar{j}} > t$, or as 'unknown' otherwise. Varying the threshold t between 0 and 1 allowed us to plot precision versus recall for a given activity and feature.

To test our hypothesis, we picked the activity 'walking' and used the three features with the highest precision values as input for the classifier. The results are shown in Figure 1, Figure 2 and Figure 3. Note that the orders imposed on the curves by the equal error rates (the intersections of the curves with the diagonal line) and by the precision values plotted in Figure 10 are the same for all three plots. This indicates that for a given feature, we can compare the cluster precision of different window lengths to estimate how well recognition rates for a particular window length will be.

In order to be useful, the results of the cluster analysis must not only generalize across different window sizes, but also across different features. In the next section, we validate this by comparing the recognition rates of different features to each other.

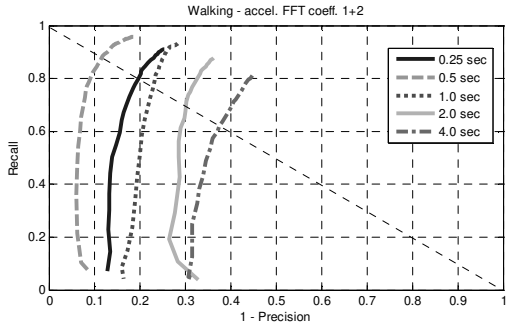

Figure 1: Recognition results for the activity 'walking' using the FFT coefficients 1&2 computed over different window sizes.

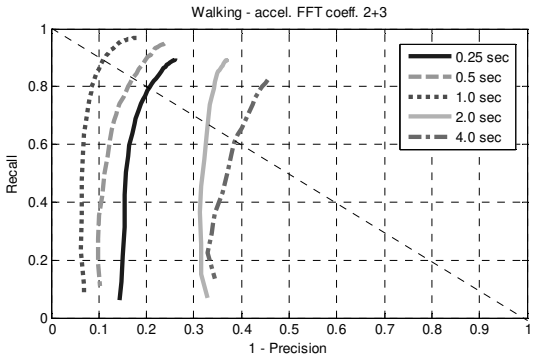

Figure 2: Recognition results for the activity 'walking' using FFT coefficients 2&3 as feature.

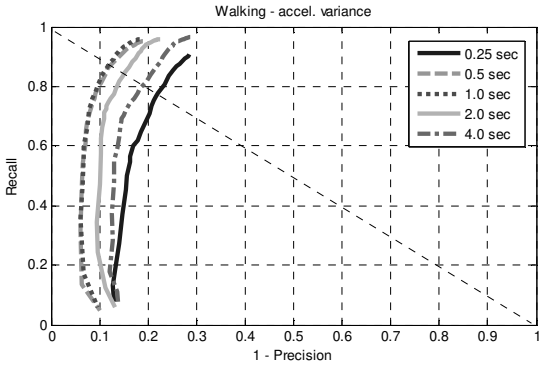

Figure 3: Recognition results for the activity 'walking' using the variance of the acceleration signal as feature.

4.1. Recognition Results

Figures 4 to 9 [TH1]show recognition results for one activity each, using the best combinations of feature and window length in terms of cluster precision for each activity. (In the legend for each plot, the features are sorted by cluster precision in descending order.) In most cases these are FFT coefficients. Recognition for 'jogging' and 'walking' performs particularly well, with equal error rates up to about 90%. Note that many curves are very steep, indicating that by lowering the threshold of the classifier, higher recall can be obtained without sacrificing precision.

Our main goal, however, was less to attain high recognition rates than to investigate to what extent the results of the cluster analysis generalize to the recognition results. One can see that for 'walking', 'jogging' and 'hopping', the order is preserved, i.e. features with higher cluster precision also have better recognition rates. For 'standing' and 'riding bus' there are only very subtle differences in the equal error rates, just like in the precision values. For 'skipping', the order is preserved except for one feature (FFT coefficients 2+3). The reason for this might be that the differences in cluster precision for these features are very small. Also, the samples for 'skipping' constitute only about 1.5% of the total number of samples, which might introduce artifacts.

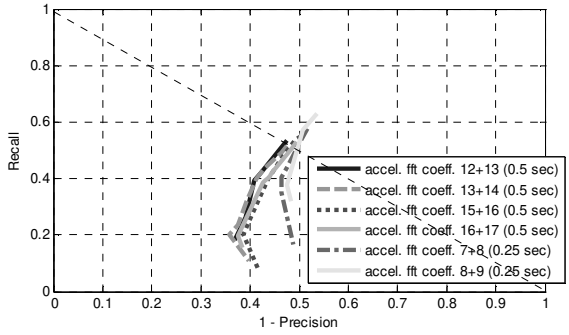

Figure 4: Recognition Results for 'Standing', using the combinations of (feature; window length) with the highest cluster precision for this activity.

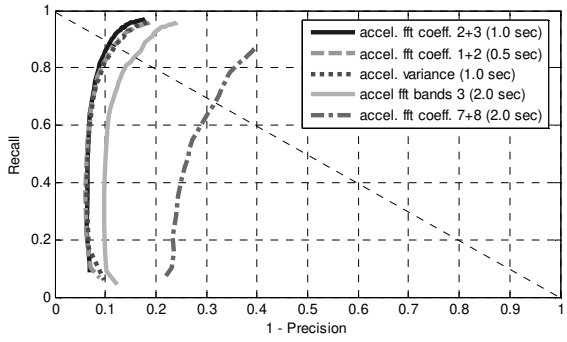

Figure 5: Recognition Results for 'Walking', using the combinations of (feature; window length) with the highest cluster precision for this activity.

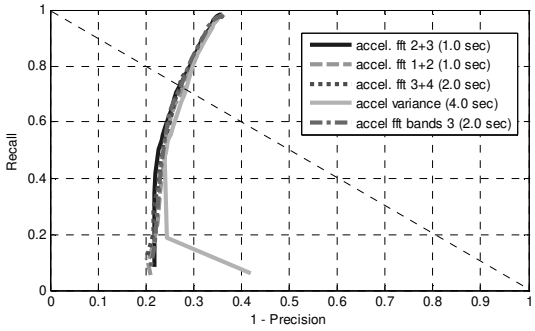

Figure 6: Recognition Results for 'Riding Bus', using the combinations of (feature; window length) with the highest cluster precision for this activity.

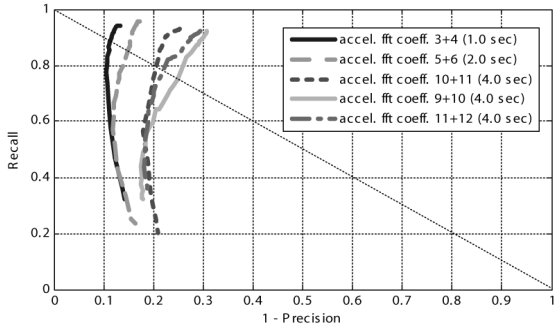

Figure 7: Recognition Results for 'Jogging', using the combinations of (feature; window length) with the highest cluster precision for this activity.

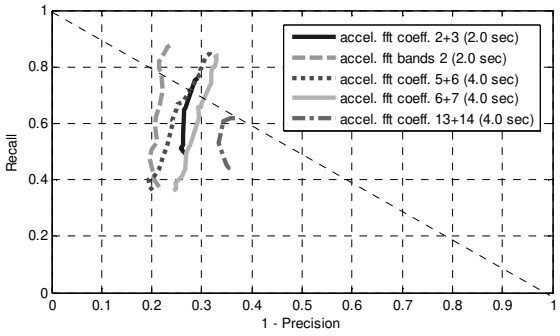

Figure 8: Recognition Results for 'Skipping', using the combinations of (feature; window length) with the highest cluster precision for this activity.

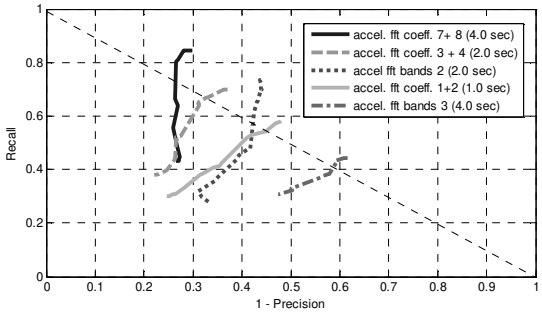

Figure 9: Recognition Results for 'Hopping', using the combinations of (feature; window length) with the highest cluster precision for this activity.

5. Summary and Conclusion

In this work we have shown that by clustering features and comparing them to each other in terms of cluster precision, one can obtain detailed information about how well a particular feature is suited for activity recognition. Our proposed measure of cluster precision turned out to be a good indicator for the recognition performance of a feature. We gave a detailed comparison of the cluster precisions of a range of features and showed that the ranking obtained from the cluster analysis is reflected in the recognition rates of the different features.

Overall, our results indicate that in contrast to an assumption that is sometimes implicitly made, there is neither a single feature nor a single window length that will perform best across all activities. By looking at the different features, we found that the FFT features always rank among the features with the highest cluster precision. However, the FFT coefficients that attain the highest precision are different for each activity, and recognition can be improved by selecting features for each activity separately. Generally, the highest peaks for the FFT coefficients can be found between the first and the tenth coefficient. Our recognition results also indicate that combining different FFT coefficients to bands of exponentially increasing size might be a compromise to using individual or paired coefficients. For the non-FFT features, we found that variance has consistently high precision values, except for the activity 'standing', where spectral entropy has the highest values. Surprisingly, the often-used mean of the acceleration signal has lower precision values than variance throughout the set of activities, except when used with a window length of 0.25 seconds for 'jogging' and 'skipping'.

In terms of window lengths, we found that on average, features with window lengths of one and two seconds attain slightly higher precision values than those with other window lengths. However, there are significant differences across the different activities, and as for the features, selecting different window lengths for different activities leads to better recognition rates. E.g., the 1 second window has the highest average precision values for the activities 'jogging' and 'walking'; the 2 and 4 second windows attain high values for 'skipping' and 'hopping', and the 0.25 and 0.5 second windows reach relatively high precision for the activity 'standing'.

Besides extending the approach taken in this paper to a larger set of activities, we plan to apply more elaborate classifier schemes such as the AdaBoost or SVM frameworks.

References

[1] Bao, L., & Intille, S. (2004, April). Activity recognition from user-annotated acceleration data. In *Proc. pervasive* (p. 1-17). Vienna, Austria: Springer-Verlag Heidelberg: Lecture Notes in Computer Science.

[2] Heinz, E., Kunze, K.-S., Sulistyo, S., Junker, H., Lukowicz, P., & Troester, G. (2003, November). Experimental evaluation of variations of primary features used for accelerometric context recognition. In *Proc. eusai, lncs* (Vol. 2875, p. 252 - 263). Eindhoven, The Netherlands.

[3] Kern, N., Schiele, B., & Schmidt, A. (2003, November). Multi–sensor activity context detection for wearable computing. In *Proc. eusai, lncs* (Vol. 2875, p. 220-232). Eindhoven, The Netherlands.

[4] Krause, A., Siewiorek, D., Smailagic, A., & Farringdon, J. (2003, October). Unsupervised, dynamic identification of physiological and activity context in wearable computing. In *Proc. iswc* (p. 88-97).

[5] Laerhoven, K. van, & Gellersen, H.-W. (2004, Nov). Spine vs. porcupine: a study in distributed wearable activity recognition. In *Proc. iswc.* Washington DC, USA.

[6] Lee, S.-W., & Mase, K. (2002). Activity and location recognition using wearable sensors. *IEEE Pervasive*, *1*(3), 24-32.

[7] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, Blake Hannaford (2005, July). A Hybrid Discriminative/Generative Approach for Modeling Human Activities. To appear in *Proc. Int. Joint Conf. on Articial Intelligence (IJCAI-05)*, Edinburgh, Scotland, 2005.

[8] Mantyjarvi, J., Himberg, J., & Seppanen, T. (2001). Recognizing human motion with multiple acceleration sensors. 747–752.

[9] Nishkam Ravi et al. (2005, July). Activity recognition from accelerometer data. To appear in *Proceedings of the Seventeenth Innovative Applications of Artificial Intelligence Conference,* 11-18. Menlo Park, Calif.: AAAI Press.

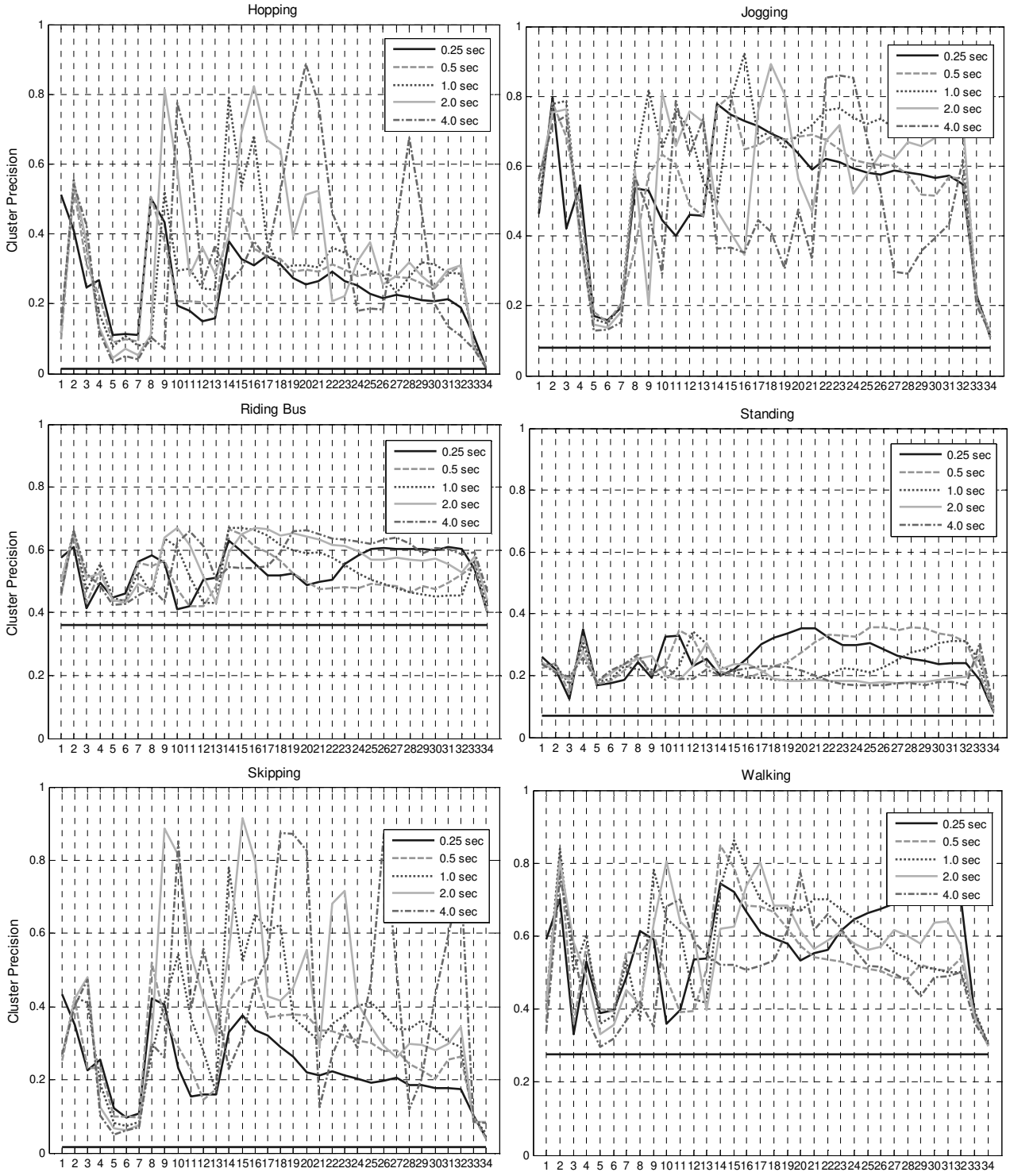

Figure 10: Cluster Precision for different activities. The numbers on the x-axis represent the following features: acceleration mean (1), variance(2), energy(3), spectral entropy(4), pairwise correlation of xy-(5), yz-(6) and xz-axes(7), exponential FFT bands 1, …,6(8-13), FFT coefficients 1&2, 2&3, …, 19&20 (14-32), digital compass variance(33) and visible light variance(34) The horizontal line marks the a priori probability of the activity.