Poster: Improving Sensor-based Activity Recognition Using Motion Capture as Additional Information

Paula Lago
Kyushu Institute of Technology
Tobata-ku, Kitakyushu-shi,
Fukuoka, Japan
paula@mns.kyutech.ac.jp

Shingo Takeda
Kyushu Institute of Technology
Tobata-ku, Kitakyushu-shi,
Fukuoka, Japan
takeda@sozolab.jp

Tsuyoshi Okita
Kyushu Institute of Technology
Tobata-ku, Kitakyushu-shi,
Fukuoka, Japan
tsuyoshi@mns.kyutech.ac.jp

Sozo Inoue
Kyushu Institute of Technology
Tobata-ku, Kitakyushu-shi,
Fukuoka, Japan
sozo@mns.kyutech.ac.jp

UbiComp/ISWC'18 Adjunct, October 8–12, 2018, Singapore, Singapore
ACM 978-1-4503-5966-5/18/10.
https://doi.org/10.1145/3267305.3267596

Abstract
We propose a new method for human activity recognition using a single accelerometer sensor and additional sensors for training. The performance of inertial sensors for complex activities drops considerably compared with simple activities due to inter-class similarities. In such cases deploying more sensors may improve the performance. But such strategy is often not feasible in reality due to costs or privacy concerns among others. In this context, we propose a new method to use additional sensors only in training phase. We introduce the idea of mapping the test data to a codebook created from the additional sensor information. Using the Berkeley MHAD dataset our preliminary results show this worked positively; improving in 10.0% both the average F1-score and the average accuracy. Notably, the improvement for the *stand, sit and sit to stand* activities was higher, typical activities for which the inertial sensor is less informative when using the wrist-worn accelerometer.

Author Keywords
activity recognition; inertial sensors; motion capture; clustering; additional information

ACM Classification Keywords
H.1.2 [User/Machine Systems]: Human information processing

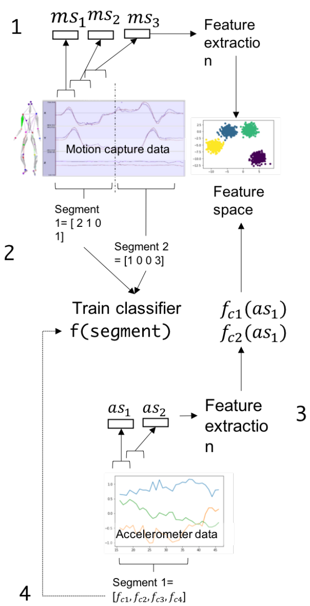

Figure 1: An overview of the proposed approach for activity recognition. Clusters learned from motion capture data are used as feature space. Accelerometer data is mapped into this space for classification

Introduction

Inertial sensor-based activity recognition accuracy for physical activities like walking and sitting is over 80%, however, accuracy for complex activities like daily life activities or nurse care activities is significantly lower [3, 6].

Current proposals to increase the accuracy rely on using additional data by combining multiple sensing modalities and/or sensors. For instance, results on the Berkeley MHAD Dataset show that activity recognition accuracy using a single accelerometer is 82.19% or 93.0% [2] depending on the positioning. In contrast, using 2 accelerometers the accuracy improves to 96.81% [2] and using the motion capture data recognition accuracy is reported as 97.58% [8]. The combination of four sensing modalities achieves 100.0% accuracy [4]. However, in real life setting using more than one sensing device or modality is restrictive due to costs, privacy concerns or battery usage trade-off. Also, the sensor positioning may not be the best for all activities.

The contribution of this paper is that ***additional data (motion capture) is only used to train the classifier*** for accelerometer based activity recognition.

Proposal Description

An overview of the proposal is shown in Figure 1. To use additional information for training, we learn clusters of motion capture data to create a codebook that is then used to recognize the activities. We then train multiple classifiers to identify the clusters from the accelerometer data. A similar idea is pursuit by other researchers [7, 1] but they use the same data for training and testing.

1. Unsupervised learning of motion capture data
We use motion capture data only during training phase to obtain a feature space with low inter-class similarity. We do this by creating clusters of movements. For this, we obtain subsequences $(ms_1, ms_2...ms_k)$ from each segment using a sliding window approach with windows of length n and overlap $p < n$. We represent each subsequence using statistical features. We cluster all movements to create a feature space of dimension ℓ, the number of clusters.

2. Supervised learning of activities
To learn the model for activity recognition we use tumbling windows of length m, $m > n$ for motion capture data segmentation. Each window is represented with a bag-of-words vector of its subsequence clusters. With these vectors, a classifier $f(segment)$ is trained for activity recognition. Notice that the classifier is trained using the motion capture data.

3. Mapping of accelerometer data to cluster
The next step is to represent the accelerometer data in the cluster space. For this, we segment accelerometer data into subsequences $as_1, as_2...as_k$ using sliding windows of length n and overlap p, same as in the motion data. The window lengths are given in seconds, to ensure that the subsequences represent the same movements of those of motion capture data. We then learn a function for each cluster $(f_{c1}, f_{c2}, ...f_{cl})$. Function $f_{ci}(as_j)$ predicts the probability of the subsequence as_j belongs to cluster i.

4. Recognizing Activity in Test Phase
For classification, we segment accelerometer data using tumbling windows of length m, $m > n$ in the same way it is done for the motion capture data. Each segment is divided into its corresponding subsequences. Each subsequence's probability of belonging to every cluster i is obtained using f_{ci}. A probability vector of length ℓ is obtained by applying each function to the subsequence. We then use the normalized sum vector for all the subsequences of one segment as input for $f(segment)$, the classification model, to predict

the activity.

Experimental Evaluation

We evaluated our proposal using the Berkeley Multimodal Activity Dataset (MHAD) [4]. In this section we detail the data, implementation and report our results together with a discussion of such results.

Data Description

The MHAD dataset contains 11 actions performed by 12 subjects with 5 recording per action and subject. Each recording contains data of an optical motion capture system (with 48 markers) and 6 accelerometers. We use all the marker data but only use the h1 accelerometer (wrist).

We used 1.25 second segments and sliding windows of 0.5 seconds with 0.25 second overlap to create the subsequences. We obtained a total of 16973 subsequences and 654 segments from the data. We divided the data into train set (13604 subsequences and 522 segments) and test set (3368 subsequences and 132 segments). Record 4 of each subject and action was chosen for the test set, so there are 12 examples of each class in the test data set.

Implementation

We implemented using python and the sklearn library [1]. We used 7 features per axis (mean, variance, kurtosis, skewness, time-weighted variance, standard deviation and median) to represent the subsequences in both motion capture and accelerometer data.

We used the affinity propagation algorithm for clustering using euclidean distance and the square root of the maximum all-pair distance as preference value. From the 16973 subsequences, $\ell = 19$ clusters were found by the algorithm. Segements are 19 column-vectors x, where $x_i, i \in [0, 18]$ represents the number of subsequences belonging to cluster i. We used SVM classifiers for activity recognition ($f(segment)$) and for cluster learning from accelerometer data (f_{ci}).

[1] http://scikit-learn.org/stable/

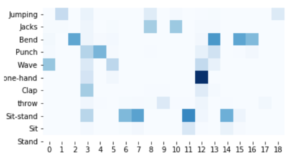

Figure 2: Correlation between cluster and activity labels

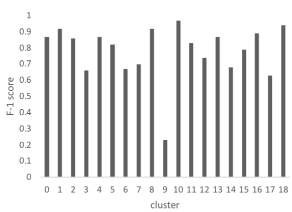

Figure 3: F-1 Score for cluster prediction from accelerometer data

To compare our approach with a traditional approach, we represent each segment of the accelerometer data with its statistical features and trained a SVM with linear kernel classifier with it.

We evaluate both classifiers in the test subset.

Results

The correlations between each cluster (x-axis) and the activities (y-axis) are shown in Figure 2. Each segment is composed of 4 subsequences. Each activity is represented by a different combination of clusters (mainly four).

Cluster prediction average accuracy from accelerometer data was 0.79 (Figure 3). All cluster except cluster 9 have relatively good prediction scores.

We used the F1-score (Figure 4) and the accuracy score (Figure 5) to evaluate activity classification. The average F1-score was improved from 0.81 to 0.83 and the accuracy was improved from 0.81 to 0.84 in the test dataset. In both cases, the activities with the lowest scores in the traditional approach (*sit*, *stand* and *sit-stand*) have the biggest improvement in both scores with our approach. In the case of the traditional approach, these activities do not have enough information from the wrist accelerometer data to be correctly classified. This is because most of the movement occurs in the lower limbs of the body and wrist acceleration data is similar for all classes. However, the cluster representation contains information from the lower limbs and therefore is able to better classify those activities.

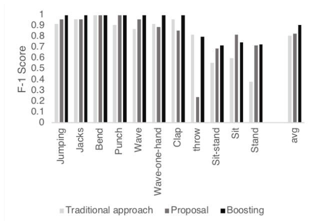

Figure 4: F-1 score for each activity in the traditional, proposed and boosting approach

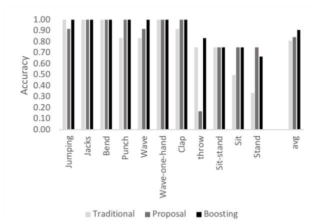

Figure 5: Accuracy score for each activity in the traditional, proposed and boosting approach

The *throw* activity had a significantly lower accuracy and F-1 score with our approach compared to the traditional approach. One reason for this result may be that the *throw* activity is mainly represented by cluster 9 (see Fig. 2) but the prediction accuracy for this cluster is low (see Fig. 3)

Because the activity had already a good recognition accuracy with the traditional approach, we implemented the SAMME algorithm [5] for boosting using traditional approach as first classifier and our approach as the second classifier. Both the average accuracy and average f-1 score improved to 91% and the *throw* recognition was also improved to 83% (See boosting bar in Figures 4 and 5).

Conclusions

We have proposed an approach to use motion capture data for training an accelerometer-based activity recognition system. Our method uses the motion capture data as additional data for the accelerometer-based activity recognition to improve the performance. Our procedure does not use mocap data in test phase, allowing users to deploy this approach in settings where only one accelerometer is installed in test phase while mocap data is available in training phase.

Our results show that the proposed approach is better than the traditional approach, specially when trained in combination with the accelerometer data using boosting. This means that the additional information did help to create a feature space where the activities are better discriminated. Our method improves both the F1-score and the accuracy of activity recognition, specially in those classes in which the wrist data is less informative (*stand, sit, sit to stand*) and the traditional approach has low-accuracy. The average accuracy for these 3 activity classes improved from 52% to 72% and the average f-1 score improved from 51% to 73%.

As future work we will evaluate the method in complex activities and in nurse care activities.

REFERENCES

1. S. Changpinyo, W. L. Chao, and F. Sha. 2017. Predicting Visual Exemplars of Unseen Classes for Zero-Shot Learning. In *2017 IEEE ICCV*. 3496–3505.
2. C. Chen, R. Jafari, and N. Kehtarnavaz. 2015. Improving human action recognition using fusion of depth camera and inertial sensors. *IEEE Trans. on Human-Machine Sys.* 45, 1 (2015), 51–61.
3. S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook. 2012. Simple and Complex Activity Recognition through Smart Phones. In *2012 8th Int. Conf. on Int. Env.* 214–221.
4. F. Ofli et al. 2013. Berkeley MHAD: A comprehensive Multimodal Human Action Database. In *2013 IEEE Workshop on Applications of Computer Vision*. 53–60.
5. Trevor Hastie, Saharon Rosset, Ji Zhu, and Hui Zou. 2009. Multi-class adaboost. *Statistics and its Interface* 2, 3 (2009), 349–360.
6. S. Inoue et al. 2015. Mobile Activity Recognition for a Whole Day: Recognizing Real Nursing Activities with Big Dataset. In *P. (UbiComp 2015)*. ACM, New York, NY, USA, 1269–1280.
7. K. Shirahama, L. Köping, and M. Grzegorzek. 2016. Codebook Approach for Sensor-based Human Activity Recognition. In *Proceedings of (UbiComp ’16)*. ACM, New York, NY, USA, 197–200.
8. S. Vantigodi and V. Radhakrishnan. 2014. Action recognition from motion capture data using Meta-Cognitive RBF Network classifier. In *2014 IEEE ISSNIP*. 1–6.