A Decision Level Fusion and Signal Analysis Technique for Activity Segmentation and Recognition on Smart Phones

Jian Wu
Department of Computer Science and Engineering
Texas A&M University
jian.wu@tamu.edu

Reese Grimsley
Department of Electrical Engineering
Texas A&M University
reesul@tamu.edu

Ali Akbari
Department of Biomedical Engineering
Texas A&M University
aliakbari@tamu.edu

Roozbeh Jafari
Department of Biomedical Engineering, Computer Science and Engineeing and Electrical and Computer Engineering
Texas A&M University
rjafari@tamu.edu

UbiComp/ISWC '18 Adjunct, October 8–12, 2018, Singapore
 ISBN 978-1-4503-5966-5/18/10...$15.00
DOI 10.1145/3267305.3267525

Abstract

The objective of this work is to recognize modes of locomotion and transportation accurately, with special emphasis on precise detection of transitions between different activities. The recognition of activities of daily living (ADLs), specifically modes of locomotion and transportation, provides an important context for many ubiquitous sensing applications. The precise detection of activity transition time is also important for applications that require immediate response. Many prior signal processing techniques use a fixed-length window for signal segmentation, which leads to poor performance for detecting activity transitions due to the limitation of a single window size. In this paper, we construct weak classifiers based on different window sizes and propose a decision level fusion approach to effectively classify and assign a label for each sample by fusing the decisions from all weak classifiers. Moreover, we propose a set of phone orientation independent features to ensure the system can work with arbitrary phone orientation. Our team, The Drifters, attained an F-score improvement of 1.9%, increasing from 94% to 95.9%, using our proposed method compared to using a single fixed-size window segmentation technique.

Author Keywords

Activity recognition; Modes of locomotion and transportation; Decision fusion; Orientation independent

ACM Classification Keywords
[Human-centered computing]: Ubiquitous and mobile computing.

Introduction
The recognition of activities of daily living (ADLs), especially the accurate recognition of mode of locomotion and transportation and the precise identification of transition time between different activities, has attracted much research attention [6, 7]. The accurate detection of modes of locomotion and transportation provides support for many applications such as healthcare, human-computer interaction and sociology [1, 8, 9]. In addition, the detection of activity transition is of special benefit to applications that require immediate response. For example, the detection of timely activity transition will help monitor the user caroli expenditure more accurately. Signal segmentation and feature extraction are crucial steps towards developing robust recognition systems since the classification algorithms are extensively investigated and can be adopted easily.

The first important step of activity recognition is segmentation. We need to determine the period during which the activity happens before we can extract features or do classification. There are two different approaches. The first approach is to accurately segment the data based on certain pattern observations [12, 13]. For example, dynamic time warping (DTW) is used to do auto-segmentation based on the discriminative pattern of each activity [12]. In another work, surface EMG is used to detect the segment of American Sign Language signs based on the muscle activity energy [13].When the user is performing a sign, the muscle activates and the total energy can be used to detect this period. The second approach is to use a fixed-size window to do segmentation if accurate segmentation is not possible [15, 10]. In this approach, it is very important to select a suitable window size. If the window size is too small, not enough information could be captured to generalize the discrimination of this activity. If the window size is too large, the window may be mixed with other activities, which will affect the classification performance. In practice, people usually select the best window size empirically. In this paper, instead of choosing a single fixed window size, we look at windows of different sizes and treat each of them as a weak classifier. Then a decision level fusion technique is applied to fuse the decisions achieved by each weak classifier and a final decision is made for each sensor sample. In this way, our approach is able to identify the transition time from one mode of locomotion to another more accurately since the large window that covers a mixture of two activities will be treated as an outlier, and the decision will rely on the weak classifiers that cover only one activity.

Another important step of activity recognition is to extract useful features from the raw sensor data that can be used to distinguish different activities. There are usually two different approaches. The first approach is to construct a time series feature vector for each activity which can serve as a discriminative pattern for this activity. Then a pattern matching/recognition algorithm could be applied for classification. The second approach is to extract useful individual features and cascade them together into a feature vector. These features may include time domain statistical features, frequency domain features or other meaningful information extracted from the data. These features can be fit into a discriminative classifier (*e.g.* support vector machine, decision tree or Naive Bayes) to recognize a certain activity. In this paper, the objective is to determine the locomotion or transportation, and no clear discriminative pattern may exist for each mode. For example, for the

modes train and subway, they may have very similar time domain patterns. Thus, the second approach is more appropriate, and is applied by extracting a set of discriminative features from the raw sensor data for classification. One challenge when extracting features is that the sensor orientation is not fixed as the user may put the cell phone at any orientation in his pocket. Therefore, a set of orientation independent features are incorporated to address this challenge.

In the previous work, most of them consider a small dataset that is captured in a lab environment or in a short period [17, 18]. In this paper, the data is captured in real-life in a period of four months. The training data size is about 15 GB. We explore the feasibility of a traditional classification pipeline and show its effectiveness when applied to big data.

The main contributions of our paper include:

- A set of orientation independent features from data collected on a cell phone is proposed to ensure the system works with arbitrary cell phone orientation in the pocket.
- A decision level fusion technique is applied to fuse the decisions achieved from each weak classifier. The weak classifiers are constructed based on different window sizes.
- The traditional machine learning pipeline is evaluated on big data for ubiquitous activity recognition and the lessons learned are described.

The remainder of this paper is organized as follows. The Sussex-Huawei Locomotion-Transportation (SHL) dataset used for this work is briefly introduced followed by the introduction of our proposed approach. We then describe the experimental setup and discuss the experimental results followed by the conclusion of this paper.

SHL Dataset and Task Description

This paper presents the techniques our team, The Drifters, employed for this submission to the Sussex-Huawei Locomotion-Transportation (SHL) recognition challenge at the HASCA Workshop at Ubicomp 2018. The goal of this challenge is to recognize eight modes of locomotion and transportation activities from the inertial sensors data of a smartphone. The activities that have to be recognized are still, walk, run, bike, car, bus, train, subway. The dataset used for this challenge, SHL dataset, comprises 271 hours of training data and 95 hours of test data [4, 3]. The data is recorded by a Huawei Mate 9 smartphone attached to the right front pocket of a single participant over 4 months. The orientation of the smartphone is not necessarily fixed. The participant performed the activities on a daily basis (approximately 5-8 hours per day) with the phone logging the sensors data. The data includes readings from 3-D accelerometer, gyroscope, magnetometer, and ambient pressure sensor as well as linear acceleration, gravity, and orientation. Data is collected from all sensors at the frequency of 100 Hz. All data samples are labeled. For both training and testing dataset, the whole data is segmented with a non-overlapped sliding window of 1-minute length. After segmentation, the order of the frames are randomly permuted, so there is no temporal dependency among the frames. The average F1-score over all of the activity classes is used to evaluate models.

Proposed Approach

Method Overview

Figure 1 shows the diagram of our proposed approach for recognizing modes of locomotion and transportation. We first extract features from the raw sensor data for weak

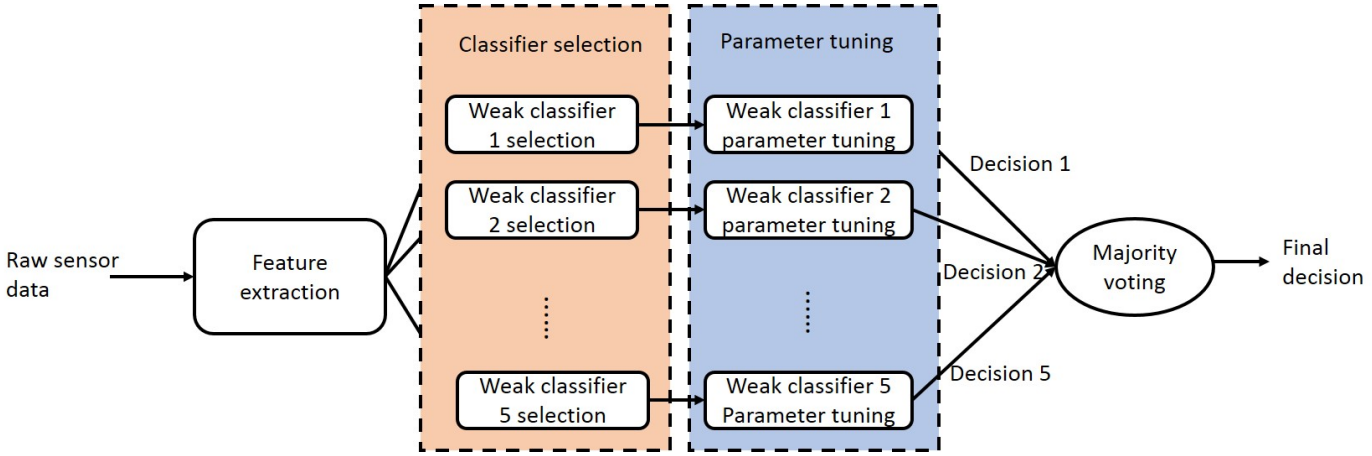

Figure 1: Diagram of the proposed approach

classifier 1 to weak classifier 5. The difference between the five weak classifiers is the window size. We set 10, 15, 20, 30, and 60 second windows for weak classifiers 1, 2, 3, 4, and 5, respectively. After we extract features for each weak classifier, we select the best classifier for each window from four popular traditional classifiers (*i.e.* decision tree, support vector machine (SVM), nearest neighbor and Naive Bayes) by measuring F-score of each classifier for the selected window size. After each weak classifier is determined, the best parameters (*e.g.*, cost and gamma for SVM) are tuned. The determined model along with the best-performing parameters will be used for classification. From each weak classifier, a decision is determined for a certain sample. The five decisions are then fused by majority voting to get the final decision.

Feature extraction

Table 1 lists the features we used in our paper, including the feature name and feature dimension. From the dataset, we have 3-d linear acceleration, 3-d gravity, 3-d gyroscope data, 3-d magnetometer data, 3-d accelerometer data, 4-d orientation data and 1-d pressure data. Before we extract the features, we simply select the modalities that will be used in our approach. We make use of all modalities except magnetometer and accelerometer. It is well known that magnetometer suffers significantly from magnetic interference in the environment. The classification performance will be decreased if magnetometer data is added based on our experiments. Since 3-d accelerometer data is simply the addition of 3-d gravity and 3-d linear acceleration, it is redundant and is not used. Instead, features for gravity and linear acceleration are considered completely separate.

Feature name	*Dimension*
Mean	1
Variance	1
Standard Deviation	1
Root Mean Square	1
Mean Cross Rate	1
Skewness	1
Kurtosis	1
FFT Coefficients	3
Entropy	1
AR coefficients	10
Integration	1
Signal Magnitude Area	1
Band Power Ratio	4
Zero Cross Rate	1

Table 1: Features

As for the features themselves, most of them are well-known features for activity recognition [14]. For the band power ratio, we look at the ratio of power in frequency bins of 0-0.5Hz, 0.5-1Hz, 0-1Hz and 1-2Hz to the total signal power. The auto-regression (AR) coefficients indicate the temporal relationship of the signal within one window. For the FFT coefficients, we use the first three orders.

We extract both cell phone orientation dependent features and orientation independent features. For orientation dependent features, we extracted the features show in Table 1 for each dimension of 3-d linear acceleration, 3-d gravity, 3-d gyroscope, 4-d orientation data and 1-d pressure. For each dimension, we have 28 features and the total orientation dependent feature size is 14*28 = 392. The orientation independent features are introduced in the next section.

Orientation independent features
One of the challenges of recognizing ADLs using cell phones is the sensor orientation displacement. The user could put the cell phone at any orientation in his pocket and the orientation of the cell phone may be changing constantly due to the user's motion. Thus, it is very important to incorporate orientation independent features to ensure the system works with arbitrary sensor orientation.

In this paper, we propose the following orientation independent features: magnitude of 3-d linear acceleration, magnitude of 3-d gravity, magnitude of 3-d gyrocope and axis angle feature. For every sample, the magnitude is calculated as L2-norm (*i.e.* least squares) of 3-d signal. The axis angle *aa* is calculated as in Equation 1. *w* is the last element of an orientation quaternion which is given by [*x*, *y*, *z*, *w*]. Axis angle tells how much total rotation happens with respect to a certain orientation. The total rotation is irrelevant to the phone orientaion.

$$aa = 2 * acos(w) \tag{1}$$

We extract the same features in Table 1 for all these four orientation independent features and it leads to a feature size of 4*28 = 112. Therefore, the total feature size for our approach is 504. All the features are then normalized to the range of [0,1] for before the classification is performed.

Segmentation and Weak Classifier
As discussed in the Introduction, we use a windowing technique for segmentation. However, it is challenging to choose a single, suitable window size. On one hand, if the window size is too small, not enough information might be captured to distinguish a certain activity. On the other hand,

if the window size is too large, it may cover more than one activity, which will increase the misclassification rate. In this paper, we consider five window sizes, each of which is treated as a weaker classifier. The five window sizes we consider are: 10 seconds, 15 seconds, 20 seconds, 30 seconds and 60 seconds. These are all reasonable window sizes for our application.

For each weak classifier, we select the best-performing classifier from four popular traditional classifiers (*i.e.* LibSVM, decision tree (DT), 10 nearest neighbor (10-NN) and Naive Bayes(NB)). An open source machine learning tool, Weka, is used for this task [5]. To select the best classifier for a given window size, the 4 traditional classifiers are trained, and their average F-1 scores from 3-fold cross validation are compared. Table 2 shows the average F-1 scores for different classifiers for a 60-second data window. We can see that LibSVM achieves the best performance and it is chosen as the classifier for 60-second window size weak classifier. From the table, we also observe that Naive Bayes achieves only 68.1% in F-score. This huge difference highlights the necessity of selecting a model. The same test on weak classifiers was used for all window sizes. For each window size, LibSVM achieves the best results, and it is therefore chosen as the classification model for all weak classifiers.

Classifier	*F-1 score*
LibSVM	94.01%
NB	68.1%
10-NN	92.1%
DT	83.1%

Table 2: Average F-1 score for 60-second weak classifier of different classifiers

Once the model is determined, the parameters should be tuned to achieve the best performance. Different parameters will lead to different classification performance and bad selection often leads to poor performance. Table 3 shows the F-1 score of 60 seconds window weak classifier based on different parameter selections for LibSVM. We can see a huge difference when selecting different parameters even for the same classifier. In this paper, we use a grid-search algorithm to determine the best parameters for LibSVM. For all weak classifiers, the radial basis function is used as the kernel. The LibSVM Matlab version is used in this paper for this purpose [2].

Cost	*Gamma*	*F-1 score*
0.5	0.5	19.66%
0.5	0.0078125	90.55%
32	0.5	42.73%
32	0.0078125	94.01%

Table 3: Average F-1 score for 60-second weak classifier of different parameters for LibSVM

Decision Level Fusion

The dataset provides labels sample-by-sample, so this same level of granularity must be provided from the decision level fusion. After we get class label for each sample from 5 weak classifiers, a decision level fusion technique is used to enhance the system performance. There are different decision level fusion techniques: bagging, boosting or rule-based decision making. In this paper, we are dealing with a large dataset and thus far, we do not have trouble with the data size. However, when it comes to fusing the decisions for each sample, the JAVA based software (weka) stops working, as it requires too much RAM for reading files

or training a model. In order to process the data more quickly and efficiently, a simple and efficient rule-based method majority voting is used to generate the class label for each sample. This method applies rules sample-by-sample by comparing the five outputs from the weak classifiers and taking the majority label as the fusion output. The majority voting is implemented by the author and is run on a super computer server which will be discussed later.

Experimental Results

Classifier	*Window size*	*F-1 score*
Weak classifier 1	10 seconds	93.62%
Weak classifier 2	15 seconds	93.75%
Weak classifier 3	20 seconds	93.83%
Weak classifier 4	30 seconds	93.52%
Weak classifier 5	60 seconds	94.01%
Proposed fusion	NA	95.9%

Table 4: Average F-1 score for different weak classifier and proposed decision fusion

Our proposed approach is validated with the training dataset published by the competition committee. This is 271 hours of data collected by a Huawei Mate 9 smartphone placed in the right front pocket of a single participant while he performed eight locomotion and transportation activities including still, walk, run, bike, car, bus, train, subway. 3-fold cross validation is applied to test the performance of our proposed approach. Since the average F-1 score is the metric the competition chooses, we evaluate our approach based on this metric. Table 4 shows the average F-1 score value for different weak classifiers and our proposed decision level fusion. We can see that our proposed approach achieves 1.9% higher F-1 score than the best weak classifier 5.

Computational resources
Since we have 504-d features and it takes a significant amount of time to extract the features for different window sizes; we use the high performance computing server of our university to do this. For each job, we used 8 core 2.4G Broadwell processors and 30G RAM. For classification, a personal laptop with Intel Core i7-6700HQ cpu @ 2.60GHz and 16G RAM is used. A model is trained for each weak classifier and the largest weak classifier model is 170 MB. It takes about two hours to train the largest weak classifier.

Conclusion

We propose a decision-fusion-enhanced, ubiquitous activity recognition system to recognize modes of locomotion and transportation. Instead of using a single window to segment data, we propose five weak classifiers based on different window sizes. A rule-based decision fusion technique, majority voting, is applied to fuse the decisions achieved from each weak classifier. The experimental results show our proposed approach achieve 1.9% improvement in F-1 score comparing to the best weak classifier. The recognition result for the testing dataset will be presented in the summary paper of the challenge [11].

Acknowledgement

This work was supported in part by the National Science Foundation, under grants CNS-1150079 and ECCS-1509063. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

References

[1] Bodor, R., Jackson, B., and Papanikolopoulos, N. Vision-based human tracking and activity recognition. In *Proc. of the 11th Mediterranean Conf. on Control and Automation*, vol. 1, Citeseer (2003).

[2] Chang, C.-C., and Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology 2* (2011), 27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[3] Ciliberto, M., Morales, F. J. O., Gjoreski, H., Roggen, D., Mekki, S., and Valentin, S. High reliability android application for multidevice multimodal mobile data acquisition and annotation. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, ACM (2017), 62.

[4] Gjoreski, H., Ciliberto, M., Wang, L., Morales, F. J. O., Mekki, S., Valentin, S., and Roggen, D. The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access* (2018).

[5] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The weka data mining software: an update. *ACM SIGKDD explorations newsletter 11*, 1 (2009), 10–18.

[6] Mitchell, S., Collin, J., De Luca, C., Burrows, A., and Lipsitz, L. Open-loop and closed-loop postural control mechanisms in parkinson's disease: increased mediolateral activity during quiet standing. *Neuroscience letters 197*, 2 (1995), 133–136.

[7] Pansiot, J., Stoyanov, D., McIlwraith, D., Lo, B. P., and Yang, G.-Z. Ambient and wearable sensor fusion for activity recognition in healthcare monitoring systems. In *4th international workshop on wearable and implantable body sensor networks (BSN 2007)*, Springer (2007), 208–212.

[8] Pirsiavash, H., and Ramanan, D. Detecting activities of daily living in first-person camera views. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, IEEE (2012), 2847–2854.

[9] Poppe, R. A survey on vision-based human action recognition. *Image and vision computing 28*, 6 (2010), 976–990.

[10] Sun, L., Zhang, D., Li, B., Guo, B., and Li, S. Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations. In *Ubiquitous intelligence and computing*. Springer, 2010, 548–562.

[11] Wang, l., Gjoreski, H., Murao, K., Okita, T., and Roggen, D. Summary of the sussex-huawei locomotion-transportation recognition challenge. In *Proceedings of the 6th International Workshop on Human Activity Sensing Corpus and Applications*, HASCA (2018).

[12] Wu, J., and Jafari, R. Orientation independent activity/gesture recognition using wearable motion sensors. *IEEE Internet of Things Journal* (2018).

[13] Wu, J., Sun, L., and Jafari, R. A wearable system for recognizing american sign language in real-time using imu and surface emg sensors. *IEEE J. Biomedical and Health Informatics 20*, 5 (2016), 1281–1290.

[14] Wu, J., Tian, Z., Sun, L., Estevez, L., and Jafari, R. Real-time american sign language recognition using wrist-worn motion and surface emg sensors. In *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on*, IEEE (2015), 1–6.

[15] Yang, J.-Y., Wang, J.-S., and Chen, Y.-P. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern recognition letters 29*, 16 (2008), 2213–2220.