A Feature Extraction Method for Realtime Human Activity Recognition on Cell Phones

Article

4 authors, including:

Sheikh Iqbal Ahamed

Marquette University

135 PUBLICATIONS **1,053** CITATIONS

SEE PROFILE

Roger O Smith

University of Wisconsin - Milwaukee

62 PUBLICATIONS **435** CITATIONS

SEE PROFILE

A Feature Extraction Method for Realtime Human Activity Recognition on Cell Phones

Mridul Khan[1], Sheikh Iqbal Ahamed[2], Miftahur Rahman[1], Roger O. Smith[3]

[1]Department of Electrical Engineering and Computer Science, North South University, Dhaka, Bangladesh
[2]Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, USA
[3]Department of Occupational Science & Technology, University of Wisconsin-Milwaukee, Milwaukee, USA
Email: mridul.khan@gmail.com, iq@mscs.mu.edu, mrahman@northsouth.edu, smithro@uwm.edu

Abstract—**In this paper we contribute a novel linear-time method for extracting features from acceleration sensor signals in order to identify human activities. We benchmark this method using a standard acceleration-based activity recognition dataset called SCUT-NAA. The results show that the described method performs best when the training and testing data are from the same person. In this context, a linear kernel based support vector machine (SVM) classifier and a radial basis function (RBF) based one produced similar levels of accuracy. Finally we demonstrate an application of the proposed method for realtime activity recognition on a cell phone with a single triaxial accelerometer. This feature extraction method can be used for realtime activity recognition on resource constrained devices.**

Keywords-accelerometer; activity recognition; context-aware systems; machine learning; sensor signal processing

I. Introduction

The recent trend of embedding a large variety of sensors in consumer electronics has made many pervasive computing applications more practical than ever before. Smart phones and gaming console controllers now often have sensors to observe acceleration, location, orientation, ambient lighting, sound, imagery etc. [1, 2]. Among these sensors, accelerometers are now ubiquitous due to their inclusion in most mid-range and high-end cell phones. These accelerometers measure the amount of acceleration being felt by the device along all three dimensions. Machine learning can be used to infer human activities from accelerometer signals since similar types of motion cause similar acceleration sequences.

Activity recognition has applications in healthcare and context-aware pervasive computing systems among others. It has been used to assess physical activity [3] and aid cardiac rehabilitation [4]. Cell phone based activity recognition systems [5] are an active area of research since they can lead to new types of context-aware mobile applications. Normally recognition is carried out in three steps. First small time segments or windows of the sensor signal are taken. Then some features that describe the general characteristics of each window are extracted. Finally a classification algorithm is used to infer the activity. Of course the classification algorithm has to be trained beforehand using a set of samples representing the activities that have to be recognized.

The feature extraction step is possibly the most important part of the activity recognition problem since classification can be handled by any existing machine learning algorithm if the features are robust. In general frequency domain features have been found to perform best [6]. However oftentimes extracting these require too much computation to be feasible in realtime systems [7]. The feature extraction scheme that we devised is computationally efficient but less tolerant of person to person variations. We combined modified versions of techniques previously used in this domain with quantitative description methods used in electroencephalography (EEG) signal analysis. Our intended use case is activity recognition on cell phones. Important characteristics of that scenario are – minimal processing capability, only one 3D accelerometer, device is carried in a mostly static orientation in the user's pocket or purse, and that the system can be trained and used by the same person, namely the owner of the phone. Performance on the standard dataset and the prototype cell phone application proves that our method is applicable for the targeted use case. As a whole this work makes the following contributions:

- A novel linear-time feature extraction scheme that uses various disparate methods to identify human activities is presented.

- Accuracy of the proposed method is shown using various classification methods on a standard accelerometer-based dataset and realtime data on a cell phone.

- Prototype application demonstrates that activities can be detected on modern cellphones in realtime without help from any external sensing or computing device.

The next section provides an overview of the related work. After that the method is presented. Then we describe how the benchmarking was performed. Finally the results are analyzed and conclusions drawn.

II. Related Work

A review of the recent research done on activity recognition using ambient and body mounted sensors is available in [9]. The paper provides a holistic view of the field by summarizing a large number of publications. Covered classification methods include: Hidden Markov Model, Neural Network, Fuzzy Logic and Support Vector Machine. Performance in previous research projects is presented for each algorithm. Then a critical analysis is carried out about the recognition metrics and results. The findings are also summarized in tabular format. This makes it easy to compare the relative performance of the approaches.

Various feature extraction methods were explored in [6]. They compared wavelet features with time-frequency domain features and found that frequency based features performed better. A K-Nearest Neighbor classifier was used. After leave one subject out cross-validation, the best inter-subject classification accuracy was achieved using resource intensive FFT features. Our work avoids such resource hungry schemes.

Support Vector Machine was used by He and Jin [7] to classify the acceleration signal's Autoregressive Model's coefficients. They explained that although frequency domain features outperform time domain features, popular frequency domain features like FFT require too much computation to be useful in real time applications. They segmented the raw accelerometer data into small windows with 50% overlapping and then computed the 4th order Autoregressive Model's Coefficients for each window. They used Support Vector Machine for classification and assessed the classifier using fivefold cross validation. They showed that depending on the activity, their computationally efficient method either outperformed or performed the same as traditional time and frequency domain based methods. An extension [8] to this work uses wavelet transform for feature extraction. The efficiently computed features we used are completely different. We also compared multiple classification algorithms.

Walking is important for cardiac rehabilitation. Six wavelet decomposition based measures were explored in [4] to identify walking amidst other activities. The research aimed to help chronic disease patients undergoing cardiac rehab. In a clinical setting, they collected data for: biking, rowing, walking on stairs, arms exercise, walking on treadmill and ball work. They down-sampled the three dimensional accelerometer readings to 25Hz and found that this was enough when working with human activities. They concluded that the good accuracy and low computational complexity of their method makes it a good choice for real time clinical applications but did not explain the complete work flow.

Cell phone accelerometers were used for activity recognition in [5]. The activities they selected are: walking, jogging, ascending stairs, descending stairs, sitting and standing. They used android phones for data collection. Readings were taken at 20Hz. Then time domain features were extracted from 10 second windows. Finally they used J48 decision tree, logistic regression and neural network for classification. Tenfold cross validation was performed. They found that to detect walking, logistic regression performed better. Jogging and ascending stairs were well detected by a neural network. J48 could best detect descending stairs, siting and standing. Activity recognition was carried out offline. In our work we demonstrate a cell phone application that does this in realtime.

Han et al. [10] presented a group of tilt invariant features and then used Hidden Markov Model with some post-processing for activity recognition. The novel features they identified are: Short Time Mean, Short Time Variance, Short Time Fourier Transform and Complex Autoregressive Coefficients. The results show that their method is robust to tilt and reasonably accurate. Our work uses a slightly modified version of the tilt invariance calculations they proposed. We also use a simpler classification method.

TABLE I. COMPARISON OF THIS PAPER WITH PREVIOUS RESEARCH

Paper	Cell Phone Based System	Focused on Feature Extraction	Compares Multiple Classifiers	Realtime Recognition
Preece et al. [6]	no	yes	no	no
He and Jin [7]	no	yes	no	no
Biddargaddi et al. [4]	no	no	no	no
Kwapisz et al. [5]	yes	no	yes	no
Han et al. [10]	no	yes	no	no
Xue and Jin [11]	no	no	yes	no
(This Paper)	yes	yes	yes	yes

A dataset called SCUT-NAA was created by Xue and Jin [11] for 3D acceleration based activity detection research. They used standalone accelerometers. But, like phones, these were simply put in a pocket or attached to the waist belt. For each activity Fast Fourier Transform, Discrete Cosine Transform, Time Domain and Autoregressive feature based recognition accuracy is given. We used their dataset to benchmark our method.

Sensor fusion can be used to further improve the results of our work and for complex activity detection. The work done by Zappi et al. [18] should be studied for this.

III. ALGORITHMS AND METHODOLOGIES

First we extract two tilt invariant signals from the three acceleration signals. From each of these two signals we compute total energy and two quantitative parameters. This gives us a total of six features. Using these six features a Support Vector Machine (SVM) could recognize activities that produced an arbitrary three-second window of the acceleration signal. To build the prototype of the cell phone application we applied the same feature extraction method on a sliding five-second window and used K-Nearest Neighbor (KNN) algorithm with K=3 for classification.

A. Tilt Invariance

A set of tilt invariance calculations proposed in [10] was used in this work. At time t let the column vector representing the raw accelerometer reading be,

$$\vec{a}_0(t) = [a_{0x}(t), a_{0y}(t), a_{0z}(t)]' \qquad (1)$$

Compute the bias vector by averaging $\vec{a}_0(t)$ over N samples,

$$\vec{b}(t) = \frac{1}{N}\sum_{t=1}^{N}\vec{a}_0(t) \qquad (2)$$

The tilt compensated acceleration vector is,

$$\vec{a}_1(t) = \begin{bmatrix} \cos\theta_2 & -\sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_2 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_2 & \sin\theta_1\cos\theta_2 & \cos\theta_1\cos\theta_2 \end{bmatrix} \times \vec{a}_0(t) \quad (3)$$

Here the tilt angles θ_1 and θ_2 are computed using the following equations,

$$\theta_1 = \arctan\left(\frac{b_y}{b_z}\right) \quad (4)$$

$$\theta_2 = \arctan\left(\frac{b_x}{b_y\sin\theta_1 + b_z\cos\theta_1}\right) \quad (5)$$

Han et al. computed the bias again and then removed it. Since the gravity component of $\vec{a}_1(t)$ is along the z-axis, the same effect can be achieved by correcting only that axis. This gives us the tilt compensated and gravity free instantaneous acceleration,

$$\vec{a}_2(t) = \left[a_{1x}(t), a_{1y}(t), \left(a_{1z}(t) - \frac{1}{N}\sum_{t=1}^{N} a_{1z}(t) \right) \right]' \quad (6)$$

To cancel the effect of accelerometer orientation, Han et al. considered the x and y components as a joint complex number. We found that taking the magnitude of the horizontal acceleration is sufficient. We extract features from two signals: the vertical acceleration, $a_v(t) = a_{2z}(t)$ and the magnitude of the horizontal acceleration, $a_h(t) = \sqrt{a_{2x}(t)^2 + a_{2y}(t)^2}$

B. Accelerometer Energy

Since we are dealing with human activities, measuring the amount of physical activity is important. Time integrals of the absolute values of accelerometer readings were summed up to assess physical activity in [3]. The same method was used in [12] and named "accelerometer energy". For activities like jumping most of the energy will be vertical while for many others most of it will be horizontal. So we extracted vertical and horizontal energy separately. When the window size is T, these are: $e_v = \int_{t=t_0}^{t_0+T} |a_v| dt$ and $e_h = \int_{t=t_0}^{t_0+T} |a_h| dt$. Since the sampling rate is known these integrals can be estimated by taking weighted sums. It should be noted that accelerometer energy is dependent on physiological factors like body weight. So, for most activities, it will vary from person to person.

C. Hjorth Mobility and Complexity

Hjorth asserted the need for quantitative methods in the description of EEG traces because the physical system generating the signals cannot be associated with the sine function concept used for frequency domain analysis. Three parameters were derived in [13] to describe EEG patterns in terms of amplitude, time scale and complexity. The parameters called Mobility and Complexity were used in [14] for epilepsy diagnosis. These are applicable to any stochastic signal but, to the best of our knowledge, have never been used for accelerometer based activity recognition.

Hjorth Mobility is the square root of the ratio between the variances of the first derivative and the amplitude. It describes the curve shape by measuring the relative average slope.

Hjorth Complexity is the ratio between the mobility of the first derivative of the signal and the mobility of the signal itself. It measures the frequency domain irregularity.

Both of these can be computed in linear time [15] using the first order difference sequence of the time series. They give us some frequency domain information without incurring a significant computational load.
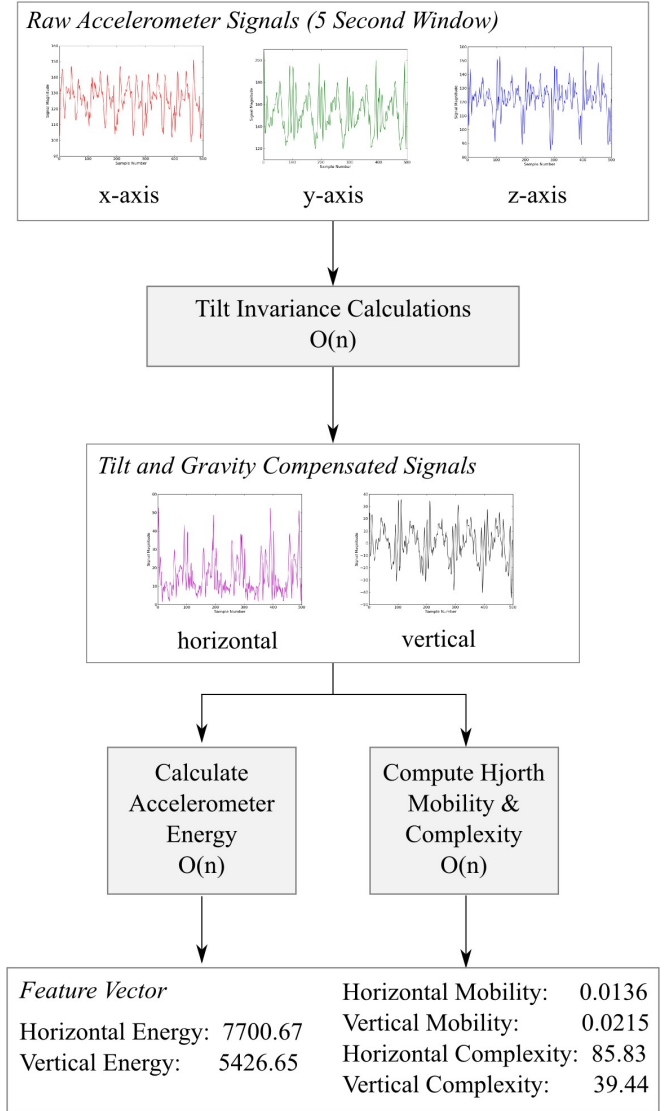

Figure 1. Proposed feature extraction scheme applied to a five seconds long accelerometer trace of a person walking.

D. Support Vector Machine

Support Vector Machines (SVM) are widely used for data classification. An SVM maps training vectors to a higher dimensional space and builds a model separating the classes. Class labels of testing vectors are predicted using this model. Basic parameters for tuning an SVM are the cost parameter and choice of kernel. For this work we used C-SVC type SVM available in LIBSVM [16]. We set cost parameter, C=10 and tried the four commonly used kernels – linear, polynomial, radial basis function (RBF) and sigmoid.

E. Prototype of Cell Phone Application

Python for Series 60 [17] can be used to quickly build prototypes of mobile applications for the Symbian platform. We built pure Python implementations of the proposed feature-extraction scheme and the well known K-Nearest Neighbor (KNN) algorithm. During development, the application was tested on a Nokia X6 and a Nokia N79. Both of these are Symbian Series 60 devices and have accelerometers. The sampling rate of the accelerometer depends on the CPU load of the phone. On our test devices it stayed around 37Hz when the device was idle. When other applications were being run the sampling rate would sometimes become as low as 3Hz. We resampled the signal at 25Hz using linear interpolation after passing it through a median filter for noise removal.

IV. STANDARD DATA

To measure the relative performance of our method we benchmarked it using a standard dataset called SCUT-NAA [11]. It contains 1278 samples from 44 subjects. They used an ADXL330 accelerometer. The signal was sampled at 100 Hz. The accelerometer was placed in the subject's shirt pocket, trouser pocket or on the waist belt.

They did not collect data for all ten activities from every subject. Furthermore some activities like walking, walking slowly and step walking are too similar to require separate treatment. We worked with six activities – climbing upstairs, climbing downstairs, jumping, relaxing, running and walking.

SCUT-NAA had data for these six activities from all of the 44 subjects. We selected random 5 second windows from the acceleration traces. For each person we took 20 windows per activity. The six previously described features were extracted from each window. We did not use conventional sliding window feature extraction. Our method could classify the windows in isolation.

We prepared two types of datasets out of the extracted feature vectors. In the first type we kept data from every person separate. Data from multiple subjects were mixed in the second case. The data was linearly scaled to a range of -1 to 1. For this data, five-fold cross validation was used to calculate accuracy.
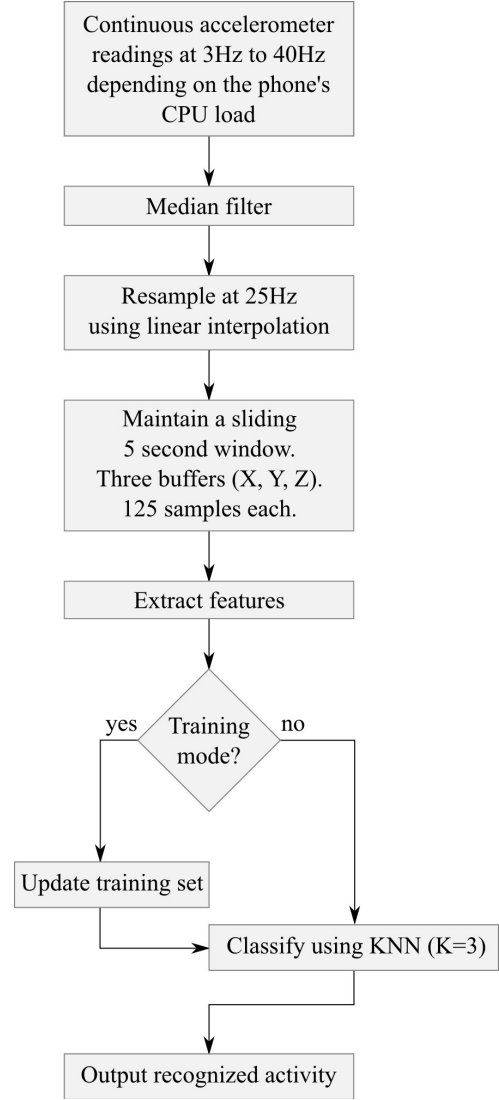

Figure 2. Basic workflow of our cell phone application.

V. EXPERIMENTAL SETUP

The results produced when we used the SCUT-NAA dataset to benchmark our feature extraction method were good. So we built a cell phone application to test its performance in realtime.

This prototype cell phone application recognizes activities from the continuous stream of acceleration data provided by the phone and displays the name of the recognized activity. We documented its performance empirically. For this we worked with the same six activities that we selected from SCUT-NAA. We selected two cell phone positions – in trouser pocket and held in hand.

The experiments were done in a naturalistic setting by a 24 year old male. The software was trained with three samples for each of the six activities. Then the realtime output of the

system was observed randomly while performing an activity. We found that the system produced spurious results when transitioning from one activity to another. So the observations were made once the signal window was known to have slid past the transition region. This happens about five seconds after the beginning of an activity. Ten such observations were used to calculate accuracy.

VI. ANALYSIS OF RESULTS

A. Single Subject Data

With the SCUT-NAA dataset, recognition accuracy was highest when the SVM was trained and tested using data from the same person. Table II shows the high level of mean accuracy achieved this way. The linear and RBF kernels performed best. The box plots in Fig. 3 and Fig. 4 show the person wise breakdown of accuracies for these two kernels. We see that, although mean recognition accuracy was high, a few outliers exist. For these outliers the recognition accuracy is abysmal and sometimes too low even for a random condition. We think that this low accuracy is caused by errors in data collection or by some unique physiological characteristics of the person from whom the data was collected. It is clear that, in the single subject case, accelerometer position does not significantly affect recognition accuracy.

TABLE II. MEAN SINGLE SUBJECT ACTIVITY RECOGNITION ACCURACY (%)

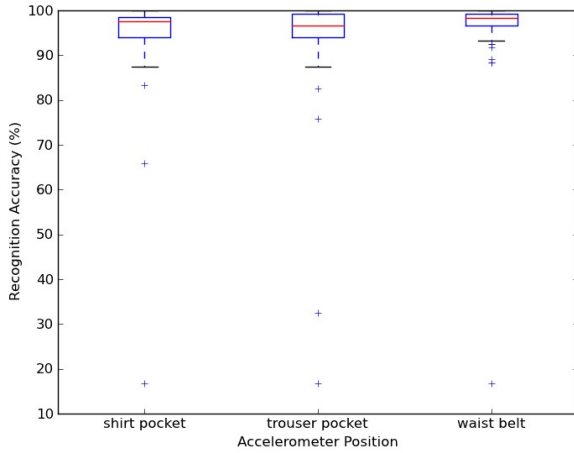| Kernel	Accelerometer Position		
	Shirt Pocket	Trouser Pocket	Waist Belt
Linear	93.43	92.58	95.33
Polynomial	88.64	84	89.97
RBF	92.65	91.89	94.38
Sigmoid	90.19	89.56	92.08

Figure 3. Box plot of activity recognition accuracy using linear kernel.
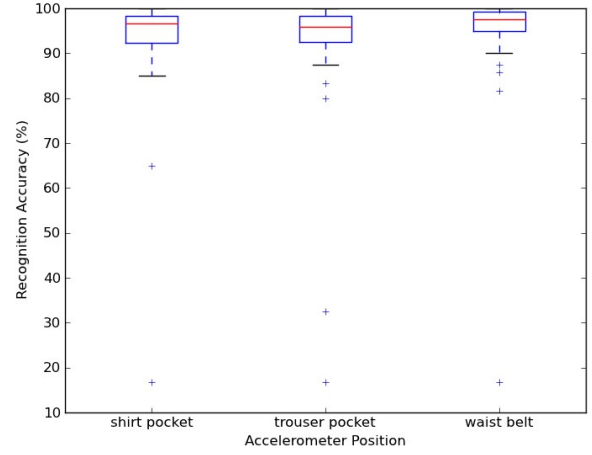

Figure 4. Box plot of activity recognition accuracy using RBF kernel.

TABLE III. MULTIPLE SUBJECT ACTIVITY RECOGNITION ACCURACY (%)

Number of Subjects	Accelerometer Position		
	Shirt Pocket	Trouser Pocket	Waist Belt
5	88	79.33	83.83
20	80	16.67	79.87
35	76.45	16.67	77.1
44	16.67	16.67	16.73

B. Multiple Subject Data

Activity recognition accuracy dropped significantly when the selected subset of the SCUT-NAA dataset contained input from more than one person. Table III shows the change in mean activity recognition accuracy as the number of people in the dataset is increased. In this case the position of the accelerometer is significant. Also there are sudden drops in recognition accuracy when certain subjects are included. Problems in the standard dataset is most likely the cause of this. The "trouser pocket" position is most easily affected by person to person variations. This is probably because this position is most sensitive to limb movement.

C. Realtime Cell Phone Data

The cell phone application showed good accuracy for most activities. Figures 5 and 6 show the empirically documented realtime accuracies for the two selected phone positions. Carrying the phone in the trouser pocket and holding it in hand produced similar results. In both cases stair climbing was recognized poorly since much of the data that can be used to differentiate it from other high intensity activities like running and jumping is lost during the tilt invariance calculations. Activities for which most of the motion is horizontal is recognized best. This is why relaxing, walking and running is detected with very high levels of accuracy.
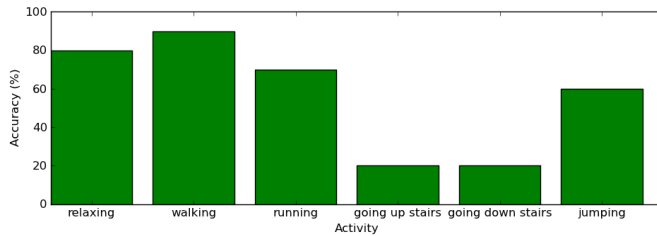
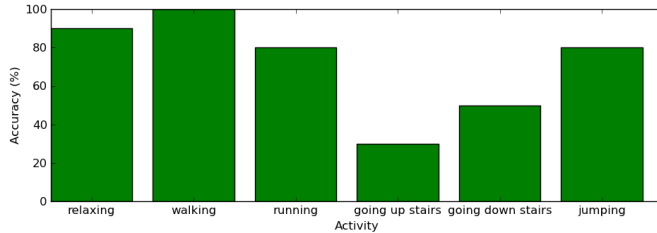Figure 5. Realtime recognition accuracy with the phone in trouser pocket.

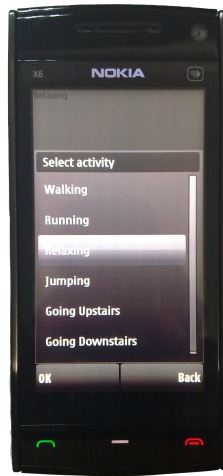Figure 6. Realtime recognition accuracy with the phone held in hand.

Figure 7. Prototype cell phone based realtime activity recognition application running on a Nokia X6. The training menu is open with the currently detected activity visible in the background.

VII. Conclusions

We found that the contributed feature extraction scheme is robust when the training and testing data is from the same person. The activity recognition accuracy decreases when the data is from multiple subjects. We were able to use the proposed method for realtime activity detection on a generic cell phone. The linear-time computation and low dimensionality makes our method suitable for realtime activity recognition on resource constrained devices like cell phones.

Future work might deal with the recognition of more complex activities. Many human activities, like cooking, studying and watching television, do not induce significantly unique acceleration traces. For these, ambient light and sound based methods can be explored by utilizing the other sensors included in modern cellphones. Phones that come with digital gyroscopes can work as crude inertial navigation systems. By combining all these, high level patterns in physical mobility can be detected. In short sensor fusion is the way forward.

References

[1] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury and A. T. Campbell, "A Survey of Mobile Phone Sensing," IEEE Communications Magazine, Vol. 48, No. 9, pp. 140-150, September 2010.

[2] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya and V. Vasudevan, "uWave: Accelerometer-Based Personalized Gesture Recognition and Its Applications," in Proc. IEEE Int. Conf. Pervasive Computing and Communication (PerCom), March 2009.

[3] C. V. C. Bouten, K. T. M. Koekkoek, M. Verduin, R. Kodde and J. D. Janssen, "A Triaxial Accelerometer and Portable Data Processing Unit for the Assessment of Daily Physical Activity," IEEE Transactions on Biomedical Engineering, Vol. 44, No. 3, March 1997.

[4] N. Biddargaddi, A. Sarela, L. Klingbeil and M. Karunanithi, "Detecting Walking Activity in Cardiac Rehabilitation by Using Accelerometer," in Proceedings of International Conference on Intelligent Sensors, Sensor Networks and Information, 2007, pp. 555-560.

[5] J. R. Kwapisz, G. M. Weiss and S. A. Moore, "Activity Recognition Using Cell Phone Accelerometers," in Proceedings of ACM SensorKDD 2010.

[6] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney and D. Howard, "A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities from Accelerometer Data," IEEE Transactions on Biomedical Engineering, Vol. 56, No. 3, March 2009.

[7] Z. He and L. W. Jin, "Activity Recognition from Acceleration Data Using AR Model Representation and SVM," in Proceedings of IEEE International Conference on Machine Learning and Cybernetics 2008.

[8] Z. He, "Activity Recognition from Accelerometer Signals Based on Wavelet-AR Model," In Proceedings of IEEE International Conference on Progress in Informatics and Computing 2010.

[9] B. Companjen, "Classification Methods for Activity Recognition," in Proceedings of Twente Student Conference on IT 2009.

[10] C. W. Han, S. J. Kang and N. S. Kim, "Implementation of HMM-Based Human Activity Recognition Using Single Triaxial Accelerometer," IEICE Transactions Fundamentals, Vol. E93-A, No. 7, July 2010.

[11] Y. Xue and L. Jin, "A Naturalistic 3D Acceleration-Based Activity Dataset & Benchmark Evaluations," In Proceedings of IEEE Systems, Man and Cybernetics 2010.

[12] Y. Fujiki, "iPhone as a Physical Activity Measurement Platform," in Proceedings of ACM CHI 2010.

[13] B. Hjorth, "EEG Analysis Based on Time Domain Properties," Electroencephalography and Clinical Neurophysiology, Vol. 29, pp. 306–310, January 1970.

[14] F. S. Bao, J. M. Gao, J. Hu, D. Y. C. Lie, Y. Zhang and K. J. Oommen, "Automated Epilepsy Diagnosis Using Interictal Scalp EEG," In Proceedings of IEEE Engineering in Medicine and Biology Society 2009.

[15] F. S. Bao and C. R. Zhang, "PyEEG: A Python module for EEG Feature Extraction," SciPy 2010, UT Austin, June 2010.

[16] C. Chang and C. Lin, "LIBSVM : A Library for Support Vector Machines," 2001. Software available at:
 http://www.csie.ntu.edu.tw/~cjlin/libsvm

[17] Nokia Corporation, "Python for Series 60," 2010. Software available at: https://garage.maemo.org/projects/pys60/

[18] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Troster, "Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness," in IEEE International Conference on Intelligent Sensors, Sensor Networks and Information, ISSNIP, 2007, pp. 281–286.