Detailed Activity Recognition with Smartphones

Conference Paper · January 2018

DOI: 10.1109/EAIT.2018.8470425

Jayita Saha
Dayananda Sagar Institutions

Chandreyee Chowdhury
Jadavpur University

Detailed Activity Recognition with Smartphones

Ishan RoyChowdhury
Jadavpur University
Kolkata, India
Email: ishan.r.chowdhury@gmail.com

Jayita Saha
Jadavpur University
Kolkata, India
Email: gjai.2000@gmail.com

Chandreyee Chowdhury
Jadavpur University
Kolkata, India
Email: chandreyee.chowdhury@gmail.com

***Abstract*—Sensors embedded in smart handheld can be extremely useful in providing information on people's activities and behaviors. Human exact activity recognition through posture identification is increasingly used for medical, surveillance. Existing works mostly uses one or more specific devices (with embedded sensor) for activity recognition and most of the time the detected activities are coarse grained like *sit* or *walk* rather than detailed like *sit on chair* or *brisk walk*. Consequently, in this paper we propose a detailed activity recognition system that uses smartphone accelerometer (available in almost every smartphone) thus does not need any special device to be carried by the user. The system applies inexpensive (in terms of resources consumed) feature extraction and learning mechanism to detect detailed activities, for instance, *slow walk* and *brisk walk*. We introduce a new feature based on jerk to detect both detailed static activities (*sit on chair*) and detailed dynamic activities (*brisk walk*). Implementation of the framework with real devices indicates 95% accuracy with state-of-the-art machine learning techniques while using a minimal set of features.**

***Index Terms*—activity recognition, feature extraction, machine learning, smartphones**

I. Introduction

Activity refers to movement of the body as a whole or different position of the limbs with respect to time upstanding against gravity [1]. Coarse-grained activity (Sit, Stand, Walk etc.) is large sub-component or basic activity whereas fine-grained, that is, detailed activity refers to small distinguishable components which can be composed together to get a coarse-grained activity. Fine grained activity contains finest activities like sit on floor/chair. Identifying detailed activity can be beneficial for many human-centered application domains such as elderly assistance at home, post trauma rehabilitation after a surgery, detection of gestures, motions and fitness etc. Traditionally computer vision-based techniques have widely been used for human daily activity tracking [2]. Though it is helpful to detect any human activity but they mostly require infrastructure support, like it needs to install video cameras in the monitoring areas and depends heavily on lighting conditions. User privacy and cost of cameras are other important issues. The emergence of wearable and ubiquitous technologies is becoming a privileged solution to provide assistive services to humans. Many wearable sensors may sometimes become an obstacle for healthcare services, hamper the movement of a person itself and generate incorrect result. In literature existing works [3], [4] recognize activity in a reasonably accurate fashion at the cost of user convenience as they consider many wearable sensors that are worn on several parts of the body.

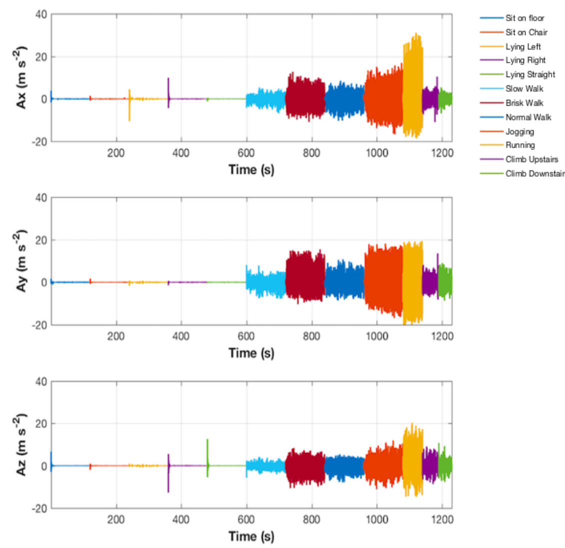

Fig. 1. Raw accelerometer data collected for 1200s from Smartphone.

Smartphones, equipped with a rich set of sensors, together with their ubiquity, unobtrusiveness, ease of use, communication channels becomes a suitable platform for activity recognition. Inertial sensors of smartphones may sense activity related data conveniently as most users carry smartphones almost always. Few works consider several inertial sensors (accelerometer, gyroscope, magnetometer) that may not be present in many smartphones [5], [6]. Activity identification using minimal sensors and with minimal feature in an efficient manner is a challenging problem. Activity recognition framework consists of phases like data collection, data preprocessing, feature extraction, learning and classification. However, it is difficult to extract features from the skewed data collected from accelerometers thus making it difficult to distinguish two static and dynamic activities as shown in Fig. I. Discover meaningful representation of data and formulate the relation of raw data with the expected knowledge for decision making is the objective of feature extraction. Most of the existing works handle activity recognition using both time domain and frequency domain feature [5], [7]. Few works consider only time domain features [6] while frequency domain features are considered in few works [8], [10] in literature. Some authors considered methods like Principal Component Analysis (PCA) [8], Linear Discriminant Analysis [9]. However, these works [3], [5] mostly focus on coarse grained activity recognition through these time and frequency domain features. Only a few works could be found on detailed activity [11], [12] using several inertial sensors that may not be present in any smartphone configuration, thus the system is not ubiquitous. Consequently, our main contribution in this work is to propose a framework for detailed activity recognition to identify five static (sit on chair/floor etc.) and seven dynamic activities (normal/slow/brisk walk etc.), explore the combination of lightweight feature extraction techniques, and identify the machine learning algorithm that best solves the problem.

The rest of the work is organized as follows: Proposed framework is detailed in Section II. Section III summarizes experimental results. Finally Section IV presents conclusions and directions for future work.

II. Fine grained Activity recognition System

Activity recognition problem can be defined in terms of the following. $A' = \{a_1, a_2, a_3, \ldots a_n\}$ represents set of static and dynamic detailed activities to be recognized by the system. Feature space $F = \{f_1, f_2, f_3, \ldots f_m\}$ contains the features extracted from preprocessed dataset $DS = \{ds_1, ds_2, ds_3, \ldots ds_x\}$. Given the classification model $(F \times DS \rightarrow A')$, the problem is to find, F_{min} such that $((F_{min} \times DS') \rightarrow A')$ where F_{min} are minimal feature set for identifying static and dynamic detailed activities and DS' is a test set. In this context we propose a new feature based on jerk in addition to generic features and insert it to F and check for its validity w.r.t. activity dataset.

Twelve detailed activities including static activities (*Sit on floor*(1)/ *chair*(2), *Lying left*(3)/ *right*(4)/ *straight*(5)) and dynamic activities (*Slow*(6)/ *Brisk*(7)/ *Normal*(8) *walk*, *Jogging*(9), *Running*(10), *Climbing upstairs*(11)/ *downstairs*(12)) are considered for this work. The raw data trace shown in Fig.I reveals that static and dynamic activities grossly show different patterns but it is difficult to distinguish between two static/dynamic activities.

The accelerometer readings are collected from smartphones kept in Right pant PockeT (RPT) that is front pocket. Noise or abnormal spikes occur in the collected sensory data, due to certain change of position or fall of the device. Accelerometer signal noise and outliers like low frequency acceleration (gravity) which depends on the orientation of the smartphone sensor with respect to ground data and high frequency signal component are eliminated using filtering techniques. Continuous filtered data are segmented into small slice through windowing or segmentation for extracting features. An orientation insensitive dimension is added in order to achieve usage behavior independent recognition along with existing three dimensions (A_x, A_y, A_z) of accelerometer readings. This is defined as Signal Vector Magnitude ($SVMag$) [6],

$$SVMag = \sqrt{(A_x^2 + A_y^2 + A_z^2)} \tag{1}$$

Feature vectors F_is are extracted on the set of segments s of the preprocessed dataset, Dt, by applying $function$. Extracted features constitute feature space.

$$F_i = function\,(s_i)\, \forall s_i \in Dt \tag{2}$$

Initially generic time domain features including mean, standard deviation, variance are extracted from preprocessed accelerometer data on 4 dimensions $(A_x, A_y, A_z, SVMag)$. Mean is used to measure central tendency. Only Mean is not sufficient to get accurate reflection of several activities on the skewed data, when data set contains significant deviation of mean accelerometer readings, as can be obsersverd in dynamic activities like jogging. So, variance and standard deviation are used in association with mean to summarize continuous data. The combination of these three features is known as feature subset 1. These features when applied to the dataset, Support Vector Machine, that is, SVM ($quadratic$) are found to be somewhat effective for static detailed activities as there is relatively less fluctuations in the acceleration signal but not for dynamic detailed activities. Min and Max are applied to define minimum and maximum value on each window respectively as the acceleration is expected to be restricted to a certain range for each class of activity. The overall accuracy would be increased using subspace K Nearest Neighbor (KNN) ensemble when the min and max features are added with feature subset 1. Entropy, measures the disorder in data set. This can be calculated in terms of non-normalized Shannon Entropy for each window. Calculation of entropy , in addition to feature subset 2 is found to improve the accuracy for all the three categories of walk, jogging and running, however no improvement in climbing stairs and static classes.

As the learning algorithms fails to learn about few detailed dynamic and static activities from the generic features, here we have introduced a new feature. The rate of change in acceleration (Jerk) is not same for several dynamic activities *(normal walk and brisk walk or climbing upstairs or climbing downstairs)*. Consequently, the new feature is computed by finding Jerk in the signal, then finding the peaks of this Jerk signal and then finally computing the summation of these peaks. Jerk is the rate of change in acceleration with respect to time or the second derivative of velocity or the third derivative of position. The maximum amplitudes of waveform are calculated to define the peaks of signal. We can get large amount of information like value of peaks, number of peaks and the space between two maximum peaks, which are distinct for different class of activities. Summation of all peaks of each individual activity for same amount of time interval helps to differentiate between two activities which have similar number of peaks. The Sampling Frequency of the signal being 50 Hz , implies that while recording data we got one observation every 0.02 seconds, hence δt = 0.02s. First, the rate of change of acceleration is recorded. The peaks in this signal are then computed and since the number of peaks vary from activity to activity, the length of the peaks array is computed . The summation of the peaks is computed to create a greater distinction between two dynamic activities.

Algorithm 1 summarizes the proposed algorithm for detailed activity framework. The experimental results for validating proposed activity recognition system are detailed in the next section.

Algorithm 1 Fine Grain Human Activity Recognition

procedure FINEGRAINHAR(A_x, A_y, A_z)
  $D_i = A_x, A_y, A_z$ ▷ Extract accelerometer data
  $D_i' = filter\,(D_i)$
  $SVMag = \sqrt{(A_x^2 + A_y^2 + A_z^2)}$ ▷ $(A_x, A_y, A_z) \in D_i^{'}$
  $Group\{f_i\} \in F$ into F_j : $f_i^{'}$ s are m consecutive samples/sec and $(F_j \cap F_k) \neq \phi$
  $\forall F_i$, compute:
    $mean = 1/n * \sum_{i=1}^{n} F_i$
    $variance = 1/(n-1) * \sum_{i=1}^{n} |F_i - \mu|$
    $standarddeviation = \sqrt{1/(n-1) * \sum_{i=1}^{n} |F_i - \mu|}$ ▷ where μ is the mean of F_i
    min = minimum element in F_i
    max = maximum element in F_i
    $entropy = -\sum_{i=1}^{n} F_i^2 \; log(F_i^2)$
    $feat$ = JerkPeaksSum(F,m,n)
  $F' = mean, variance, standard-deviation, min, max, entropy, feat$
  $clf = Classify(F', labels)$
  $pred = clf.predict(F_{test})$
  $result = accuracy\,(labels_{test}, pred)$
  return $result$

procedure JERKPEAKSSUM(F, m, n) ▷ $m = |F|, n = |F_i|$
  $\delta t = 0.02$
  $\forall F_i$, compute:
  for j = 1 to n-1 **do**
    $A_i[j] = (F_i[j+1] - F_i[j])/(\delta t)$
  $\forall A_i, peaks = findpeaks(A_i)$
  $len = peak.length$
  $feat_i = \sum_{k=1}^{len} peaks(k)$
  return $feat$

III. PERFORMANCE EVALUATION

In this section the performance of the proposed framework is evaluated for real data collected from 2 smart handheld devices for 5 users. An android application is developed to collect the embedded tri-axial accelerometer reading with 50 samples per sec. Butterworth and median filter [13] removes the low-frequency acceleration (gravity) and noise from accelerometer signal. Preprocessed data are grouped into overlapping 2.56 seconds windows and features are extracted. The classifiers that were used to test the proposed framework are : i.Trees(Complex Tree) ii. SVM (Linear) iii. SVM(Quadratic) iv.SVM(Cubic) v. KNN (Fine Knn, No of Neighbors = 1) vi. Ensemble(Boosted trees) vii. Ensemble(Bagged Trees). Five fold cross validation is used for avoiding over fitting.The comparison for activity wise classification accuracies are summarized in Fig.2 for different feature subsets. The highest

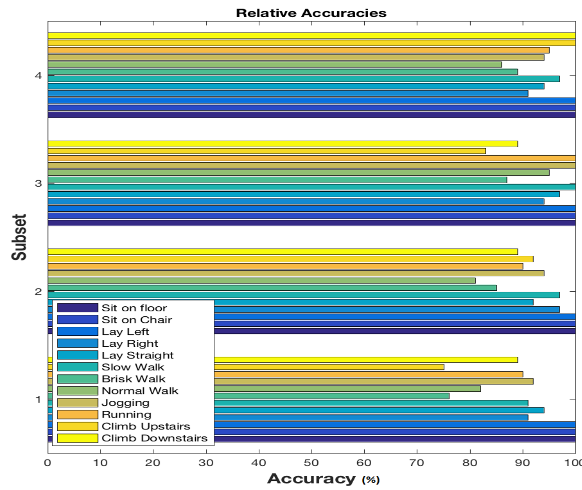

Fig. 2. Classification accuracy of different activities for different feature subset

overall accuracy is obtained for feature subset 4 and its performance for every classifier is summarized in Table I.

Initially we apply feature subset 1 on segmented data, which contains mean, standard deviation and variance. The best overall accuracy is 90.9% and it performs well for static activities with very little misclassification error for classes lying right and lying straight. The feature subset is able to perform well on static activities as there is relatively less fluctuations in the acceleration signal while performing the static activities and for a particular activity the data points are very close to their mean and the variance is very low which is enough to differentiate between the static classes of activities but in the dynamic classes the acceleration values of two activities are overlapping and are more spread out and feature subset 1 fails in learning to differentiate between them. The results for feature subset 2 (addition of min and max features) are even better for static classes. There is improvement in the accuracy for running, jogging, slow and brisk walk, but the accuracy dips a little for normal walk. However, the misclassification for dynamic classes is large as min and max features are not good enough to learn the fine differences in the dynamic classes (often overlapping range).

There is no such improvement in static activity when entropy is appended to form feature subset 3, however, significant relative improvement can be observed in dynamic activity accuracies. Brisk walk accuracy increases, however, no notable improvement of accuracy can be found for climbing upstairs/downstairs. The features introduced previously fail to differentiate the more detailed distinctions in dynamic activities hence the introduction of new feature. The relative performance of feature subset 4 is summarized in Table 1. The decision tree classifier has lesser accuracy as they are non robust and prone to overfitting. The KNN classifier, on the other hand, fails as being an instance based learning, it is highly dependent upon its neighbor. The SVM classifier

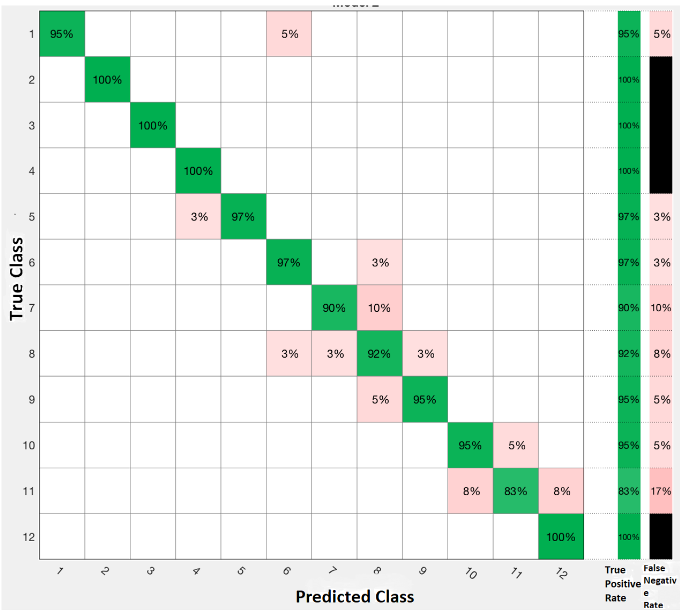

Fig. 3. Confusion matrix for average recognition accuracy for all activities with feature subset 4 subject to SVM(Cubic)

performs poorly with linear kernel, however, its accuracy goes up with increase in polynomial degree of the kernel, in fact, SVM (Cubic) has the best accuracy and its accuracy for each class is summarized in the confusion matrix in Fig.3. The higher degree polynomial kernels allow a more flexible decision boundary as they create a greater non linear separation between the closely related classes of activities, however further increase in polynomial degree would result in overfitting. The prediction accuracy for walk classes increase significantly with the introduction of the proposed feature, so as for running and jogging. Climbing upstairs and downstairs have tolerable misclassification error. Our proposed framework performs best for SVM(cubic) classifier with feature subset 4. The implementation of the framework made use of the findpeaks function present in MatlabR2016b version .

TABLE I
A COMPARATIVE STUDY OF SEVERAL CLASSIFIERS

Classifier	Accuracy
Tree($ComplexTree$)	91.7%
SVM($LinearSVM$)	89.5%
SVM($QuadraticSVM$)	93.1%
SVM($CubicSVM$)	95%
KNN($FineKnn, Noofneighbours = 1$)	90.9%
Ensemble($BoostedTrees$)	92.8%
Ensemble($baggedTrees$)	94.2%

IV. CONCLUSION

This work proposes a detailed activity recognition framework using smartphone accelerometer. The framework can be used to identify generic detailed activities with reasonable accuracy. The proposed framework uses time domain features along with proposed feature and feeds it to classifiers to predict activity. Here we have considered time domain features mean, standard deviation, variance, min, max and entropy for feature extraction and seven classifiers to analyze meaningful patterns. We have proposed here a feature extraction technique based on jerk, peak and summation of peak of accelerometer signal to identify better pattern. With this feature, detailed activity recognition can be done with 95% accuracy. The use of only one sensor, accelerometer (that is available in any smartphone configuration) makes the system cost-effective and ubiquitous. We plan to add more activities to our experiments like lying on floor/bed, stand on floor/elevator etc. The number of participants are to be increased as well to collect an extensive set of data with different device and usage behaviour.
Acknowledgements: The project is supported by *Mobile Computing and Innovative Applications* funded by UGC UPE II of Jadavpur University, India.

REFERENCES

[1] G. Sebestyen, I. Stoica and A. Hangan, "Human activity recognition and monitoring for elderly people," in IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), 2016, pp. 341–347.

[2] T. Banerjee, J. M. Keller, M. Skubic and E. Stone, "Day or Night Activity Recognition From Video Using Fuzzy Clustering Techniques," IEEE Transactions on Fuzzy Systems, 22(3), 2014, pp. 483–493.

[3] T.Liu, Y.Inoue, and K.Shibata, "Development of a wearable sensor system for quantitative gait analysis," Measurement, 42(7), 2009, pp.978-988.

[4] Y.Zhang, S.Markovic, I.Sapir, R.C.Wagenaar, and T.D.Little,"Continuous functional activity monitoring based on wearable tri-axial accelerometer and gyroscope", in 5th International Conference on Pervasive Computing Technologies for Healthcare(PervasiveHealth), 2011, pp. 370-373.

[5] D.N.Tran and D.D.Phan,"Human activities recognition in android smartphone using support vector machine," in 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS),2016 , pp. 64-68.

[6] M. Shoaib, H. Scholten and P. J. M. Havinga, "Towards Physical Activity Recognition Using Smartphone Sensors," in IEEE 10th International Conference on Ubiquitous Intelligence and Computing, 2013, pp. 80–87.

[7] J. Wannenburg and R. Malekian, "Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing," IEEE Transactions on Systems, Man, and Cybernetics: Systems, 99, 2016, pp. 1–8.

[8] Z. He and L. Jin,"Activity recognition from acceleration data based on discrete consine transform and svm," in IEEE International Conference on Systems, Man and Cybernetics, 2009, pp. 5041-5044.

[9] Y-P. Chen, J-Y. Yang, S-N. Liou, G-Y.Lee, and J-S. Wang, "Online classifier construction algorithm for human activity detection using a tri-axial accelerometer," Applied Mathematics and Computation, 205(2), 2008, pp. 849-860.

[10] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data", in PERVASIVE 2004, pp. 1-17.

[11] D. De, P. Bharti, S. K. Das and S. Chellappan, "Multimodal Wearable Sensing for Fine-Grained Activity Recognition in Healthcare," IEEE Internet Computing, 19(5), pp. 26–35, 2015.

[12] C.Pham, N.N.Diep, and T.M.Phuong, "A wearable sensor based approach to real-time fall detection and fine-grained activity recognition", Mobile Multimedia, 9(1 and 2), pp.15-26, 2013.

[13] W. z. Wang, Y. w. Guo, B. y. Huang, G. r. Zhao, B. q. Liu and L. Wang, "Analysis of filtering methods for 3D acceleration signals in body sensor network," in International Symposium on Bioelectronics and Bioinformations 2011, pp. 263–266.