Multisensor Data Fusion for Human Activity Recognition

Sarthak Ahuja
B.Tech Computer Science
Indraprastha Institute of Information Technology
Okhla, Delhi 110020
Email: sarthak12088@iiitd.ac.in

Anchita Goel
B.Tech Computer Science
Indraprastha Institute of Information Technology
Okhla, Delhi 110020
Email: anchita12019@iiitd.ac.in

***Abstract*—Human activity recognition is a well known area of research in pervasive computing, which involves detecting activity of an individual by using various types of sensors. This finds great utility in the context of human-centric problems in the real world not only for purposes of tracking ones daily activities but also in monitoring activities of others - like the elderly, patrol officers, etc for purposes of health-care and security. With the growth of interest in AI, such a system can provide useful information to make the agent much more intelligent and aware about the user, thus giving a more personalized experience. Several technologies have been used to get estimates of a person's activity like sensors found in smartphones(accelerometer, gyroscope, magnetometer etc.), egocentric cameras, other wearable sensors to measure vital signs like heart rate, respiration rate and skin temperature (apart from the same data provided by smartphones), worn on different parts of the body like chest, wrist, ankles, environment sensors to measure humidity, audio level, temperature etc. The activities that can be recognised are daily activities like walking, lying down, sitting, standing, running, travelling. However, to the best of our knowledge we have come across no work where a fusion of these sensors and egocentric cameras has been put to use. In this paper we explore the suggested fusion of sensors and share the results obtained. Our fusion approach shows significant improvement over using both the chosen sensors independently.**

Keywords*—*accelerometer, gyrometer, sensor, smart watch, egocentric, fusion, human activity.

I. Motivation

As of today the state-of-the-art architecture for Human activity detection is found in two different domains. One is wearables where wearable sensors like accelerometers, gyrometers mounted on smartphones and smartwatches are used and the other where egocentric cameras are used. The use of egocentric cameras on human activity detection has just come up recently and has been used to detect only a limited set of activities. We here perform data fusion between these two domains in an attempt to improve the performance of the existing architectures.

II. Literature Survey

We performed a comprehensive literature survey as a part of this project to identify the classifier popularly used and the activities that they have successfully able to detect, which has been summarised and tabulated in table 2. We clearly observe that among popular approaches, RDF and Decision trees have proven to be successful in classifying accelerometer and gyrometer data while SVMs have been the best choice for optical flow related features.

III. Data

A. Collected Data Set

1) Data Collection Procedure: We used the following devices for data collection:

1) OnePlus One Android Smartphone
2) LG G Watch R (W110)
3) GoPro Hero3+

We developed separate applications for the phone and the smartwatch which take as input the name of the person, the activity being recorded. The output is a file containing the data for the three axes for both sensors on each device. We collected data from 25 different people. Each person was made to perform seven activities: sitting, standing, walking, jumping, running, climbing up the stairs and going down the stairs. Each activity was performed for about a minute.

2) Data Set Pre-Processing: Each activity was recorded separately. But before actually starting with any activity there is some noise at the beginning and the end because of adjusting the application input. Hence, we performed visual editing of the data after plotting each activity for each person on a graph and choosing the relevant time period.

3) Annotation: We talk about ELAN which we use to annotate and clean our videos. We add either the name of the activity or 'DontCare' to remove parts which contain moving objects or pure darkness, low gradient.

4) Data Set Highlights:

1) For purposes of data collection we covered the entire 23 acre campus of IIIT-Delhi.
2) Does not include moving objects.
3) Equal distribution of gender, height, weight.
4) Good distribution over illumination conditions.
5) Wider variety of activities.
6) Preference to indoor conditions due to high gradient.

B. Public Data Sets

We primarily used three databases:

1) Huji/Chetan + Allen: Egocentric Videos

Research	Devices Used	Activities recognised	Database	Comments
Foerster et al.,1999[1]	External accelerometer sensors	Sitting, Standing, Walking, Climbing, Cycling	50 min recordings for 24 participants	Database not available online. Separate the DC and the AC components of the accelerometer signals.
Parkka et al., 2006[2]	Accelerometer, Temperature Sensors, Compass, IR Light Reflectance, Piezo Sensor, Microphone	Biking, Sitting, Standing, Running, Rowing	31 hours of annotated 35 channel data from sixteen participants	Database not available online. Classifiers include Artificial Neural Network, Custom Decision Trees, Automatically Generated Decision Trees
Maurer et al., 2006[3]	eWatch with Accelerometer, Light and Temperature Sensor, Microphone	Sitting, Standing, Walking, Climbing Up, Walking Down, Running	50 minutes data each from six participants	Database not available. Classifiers include Decision Trees (C4.5 algorithm),k-Nearest Neighbor (k-NN), Naive-Bayes and the Bayes Net classifier
Tapia at al., 2007[4]	3-D accelerometers, wireless Heart Rate monitor	Lying down, Sitting, Standing, Cycling, Running, Walking, Ascend stairs, Descend stairs, Lifting weights	21 participants performing 30 different gymansium activities	Along with activity recognition, also identified the intensity of the activity (in rpm). Classifiers include Decision Trees (C4.5 algorithm) and Naive Bayes Classifier
Poleg et al., 2014[5]	Egocentric Camera	Moving in car,bus, Sitting, Standing	29 videos - a few from youtube and rest self-recorded	Database available online. Used One-vs-One model for SVM classification.

TABLE I. LITERATURE SURVEY CARRIED OUT AS A PART OF THE PROJECT

2) UCI Dataset: Smart Phone
3) MHEALTH Dataset: Smart Watch

IV. APPROACH

A. Classification Problem

We begin by dividing the classification problem primarily into two stages: the first one predicts the person's state of motion, whether the person is static or in motion. The second stage involves further classification of the classes in the first stage. When a person is static, his activity can be further classified into either sitting or standing. When a person is in motion, his activity can be classified into the following categories:

1) Jumping
2) Walking
3) Running
4) Stairs Up
5) Stairs Down

Performing the classification in such a manner optimises both run time and memory as compared to directly performing a seven class classification problem. It also allows us to have a variety of different classifiers in the two stages with customized parameters to give us better results. This has also been depicted in figures ahead.

B. Procedure

The following summarises the strategy adopted to collect data, pre process it and use it for classification.

1) We segregated the tasks of solving the entire problem using data from each available sensor individually. We performed the initial classification for individual sensors on publicly available data sets so as to obtain expected results for each sensor.
2) For each sensor, we performed classification on the available data set on line and the one we collected during the course of this project.
3) We tried a varied range of classifiers and grid searches to obtain the best parameters for specific classifiers.

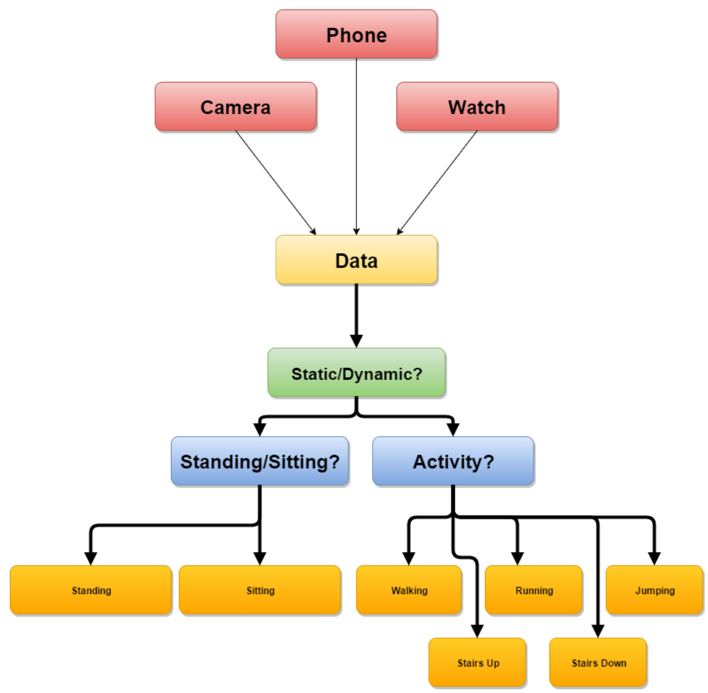

Fig. 1. Classification of Activities

V. CLASSIFICATION USING SMART PHONE

A. Data Collection

The smart phone was put in the back pocket at the hip. This ensures that we record activity of the legs carefully to capture the signature movements for any activity.

B. Process

After obtaining data for accelerometer and gyrometer along the three axes from the smart phone we performed visual pre processing of the data. Since, there was a lot of noise in the data, we performed one dimensional fourth order median filtering on the collected dataset. The dataset was divided into windows and over each window the features were extracted. Keeping a window size of 40 and an overlap of 50% between the windows, the total dataset comprises of around 9000

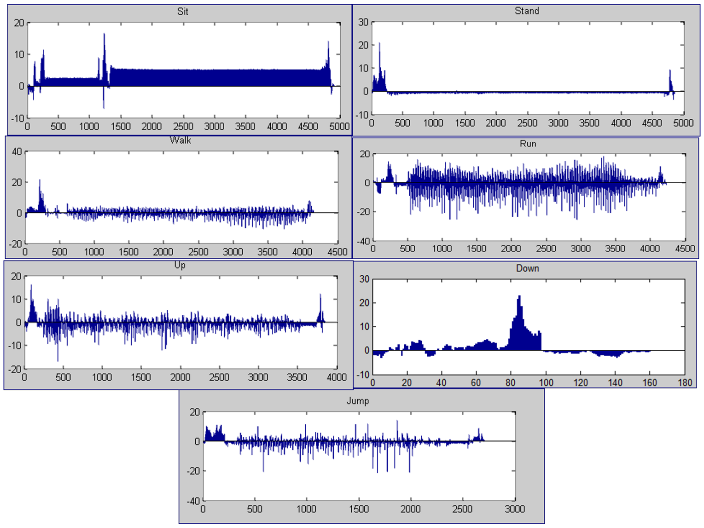

Fig. 2. Plotted accelerometer raw data for different activities

samples.
For the UCI database, the features were already extracted and filtered using several noise filters and Butterworth Filter. The features were extracted from the raw data. The features colected are enlisted below.
A visualization of the acclerometer data collected from smart phone on various activities in our dataset is depicted in figure 2.

C. Features

The following features were calculated from the raw data for both the sensors:

1) Time Domain: Mean, Standard Deviation, Maximum, Minimum, Signal Magnitude Area, Inter-quartile range, Entropy, Auto-regresion coefficients with Burg order equal to 4, Correlation Coefficient between each pair of the three axes
2) Frequency Domain: Mean, Standard Deviation, Maximum, Minimum, Signal Magnitude Area, Inter-quartile range, Entropy, Skewness, Kurtosis

There were a total of 148 features (features were obtained for all the three axes of a sensor) obtained for each sensor.

D. Results

We used the following methods for classification: Random Decision Forests, Support Vector Machines, Naive Bayes Classifier. The results obtained from testing the data on the above classifiers on the two stages are shown in Table II.

VI. Classification using Smart Watch

A. Data Collection

The smart watch was worn at the wrist by the participants. This ensured that we captured the movement of the arms carefully to capture the signature movements for any activity.

TABLE II. Phone Results

Classifier	Accuracy(per)
Stage1	
RDF	99.95
SVM(Polynomial)	97.62
NB	99.37
Stage2(a)	
RDF	98.87
SVM(Polynomial)	98.59
NB	63.94
Stage(b)	
RDF	35.83
SVM(Polynomial)	26
NB	39.05

B. Process

The raw data from the smart watch was pre-processed in the same way as done for the smart phone data. The filters and the method of feature extraction were the same as done for the smart phone data. Keeping a window size of 40 and an overlap of 50% between the windows, the total data set comprises of around 3000 samples.

C. Features

For the smart watch data set, the same features that were extracted for smart phone were extracted. Since the data in the publicly available data set for smart watch was also in raw form, we extracted the same features for this data set too. We obtained a total of 148 features for both the data sets.

D. Results

We used the following methods for classification: Random Decision Forests, Support Vector Machines, Naive Bayes Classifier. The results obtained from testing the data on the above classifiers on the two stages are shown in Table III.

TABLE III. Watch Results

Classifier	Accuracy(per)
Stage1	
RDF	83.85
SVM(Polynomial)	87.85
NB	92.89
Stage2(a)	
RDF	72.67
SVM(Polynomial)	63.33
NB	71.33
Stage(b)	
RDF	29.92
SVM(Polynomial)	26
NB	25.9

VII. Egocentric Camera Based Classification

A. Highlights

The biggest drawback of fitness sensors in smartphones and smartwatches is that they are very susceptible to motion often leading to wrong results. The idea of using an egocentric camera is to prevent such false sudden changes by including the visual data in the decision making process. Computer vision provides a source of information called optical flow that is useful in detecting motion in real time . Optical flow or optic flow is the pattern of apparent motion of objects, surfaces, and

Fig. 3. Optical Flow: from top- Standing, Sitting, Walking, Running, Jumping, Stairs Up, Stairs Down

edges in a visual scene caused by the relative motion between an observer (an eye or a camera) and the scene. The Lukas-Kanade method is a popular method to obtain the optical flow. It is based on solving a set of linear equations using least squares approximation on the intensity and 2D motion constraint of an image pixel as explained in [6].

B. Process

1) Available Dataset: In the Chetan/Huji Dataset[5] activities used were:

1) Standing
2) Walking
3) Sitting
4) Wheels
5) Static
6) Car
7) Bus

The dataset consists of 35 videos, of about 12 minutes each collected by 5 individuals.

2) Collected Dataset: The dataset we collected consists of annotated 175 videos of roughly 1 minute each (specific activity) collected from 25 individuals performing the seven activities (mentioned in the dataset section above).We have pre-processed and annotated these videos and performed the following classification on them.

C. Features

Following the approach in [5] we use cumulative displacement curves of the optical flows by calculating them for every element in an overlaying grid. This approach towards calculating the optical flow readily deals with noise and errors that may be caused by instantaneous head motion or jerks. A total of 13 features are calculated upon the obtained optical flow at each frame which characterize the optical flow on the basis of its smoothness, variance, mean and standard deviation. Another interesting feature calculated is the radial projection response over the focus of expansion. This basically gives us a quantitative estimation of how the optical flow vectors of each cell align with an outward vector from each cell. For a person in motion these would have a high value as the optical flow vector will try and grow perpendicular to the outward vector while vice versa would happen in a static context. Figure 3 shows how the optical flow looks during various human activities. (on our dataset). One can clearly see the values of optical flow(both) . It can also be imagined how the optical flow will show high variance over a time period during running when the image is jerky as compared to when one is walking. Besides just extracting the above features a certain amount of pre-processing is done to remove highly abnormal values and smoothen the optical flow by applying a smoothening filter. (to remove noise)

D. Results

Following the literature survey we chose SVM as our first choice for a classifier for stage 1. We provide results of a **10-cross validation** on our collected dataset on various kernels with grid search obtained parameters using the [7] libsvm library(**degree : 3, gamma : 1/13, coef0 : 0, cost : 1**). The results obtained are shown in the table IV (Stage 1). We also performed classification using Naive Bayes and Random Decision Forest for Stage 1 on the camera data to obtain better accuracies as shown in table IV. For stage 2 we performed the same steps mentioned above. For Stage2(a) i.e. intra-static class classification we obtain best results on using a RBF SVM as depicted in table IV. For Stage2(b) i.e. intra-dynamic class classification,we observed that though the results are overly pleasing we must take into consideration that only 21 videos were used at this stage from three subjects (which may have introduced a bias). The length of the videos was **1 minute** each which led to **73779** feature vectors when frames were taken at **6** frames per second.

TABLE IV. CAMERA RESULTS

Classifier	Accuracy(per)
Stage1	
RDF	95.36
SVM(Polynomial)	94.20
NB	56.35
Stage2(a)	
RDF	55.40
SVM(Linear)	61.0
NB	56.35
Stage(b)	
RDF	43
SVM(RBF)	56
NB	43

VIII. SENSOR FUSION

Based on the results obtained above and classification performances by the various classifiers using various sensors, we propose the following approach to combine these results:

A. Hierarchical Greedy Tree

In this model we use the best accuracy classifier at each tree joint which gives us:

1) Use RDF Phone to classify between static and motion.
2) Use Phone RDF to distinguish between standing and sitting.
3) Use Camera SVM with RBF to distinguish between the in motion activities.

Stage 1

Class	Static	Dynamic
Static	354	1
Dynamic	0	1539

Stage 2

Class	Stand	Sit
Stand	289	0
Sit	4	62

Class	Jump	Walk	Run	Up	Down
Jump	226	0	8	1	0
Walk	2	228	1	92	2
Run	43	3	241	60	0
Up	0	64	1	245	0
Down	28	162	16	116	0

Fig. 4. Confusion Matrix at the two stages

IX. Final Evaluation

The confusion matrix for different stages is given in 4. We successfully combine the best accuracies of each sensor to create a more superior classifier. While the phone sensors prove to be efficient in detecting static motion, the egocentric camera proved to be superior in classifying the various motion activities. This combination of sensors hence, is better than using any of the involved sensors, independently. As part of future work we will look to add a feature of direction/orientation of the sensor to improve our results. We observe that our classifier fails to separate the motion of going down the stairs with the motion of going up. This can be understood by the fact that all features extracted from the sensors are independent of direction considering that the sensor may be placed with any orientation.

A. Code

Link to our Code. It contains all of the following:

1) Smart Phone Data Collection Code
2) Smart Watch Data Collection Code
3) Main Code

Our data is available on request from here.

Acknowledgment

We would like to thank Dr. Richa Singh under whom guidance we successfully completed this project. We would also like to thank Dr. Chetan Arora for lending his GoPro Camera, Deepali Kishnani for lending her Smart Watch and the 25 students of IIIT-Delhi who volunteered to participate in our Data Collection Drive.

References

[1] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Computers in Human Behavior*, vol. 15, no. 5, pp. 571–583, 1999.

[2] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *Trans. Info. Tech. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006. [Online]. Available: http://dx.doi.org/10.1109/TITB.2005.856863

[3] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*, ser. BSN '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 113–116. [Online]. Available: http://dx.doi.org/10.1109/BSN.2006.6

[4] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor," in *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*, ser. ISWC '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 1–4. [Online]. Available: http://dx.doi.org/10.1109/ISWC.2007.4373774

[5] Y. Poleg, C. Arora, and S. Peleg, "Temporal segmentation of egocentric videos," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 2537–2544. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2014.325

[6] "Optical flow." [Online]. Available: http://crcv.ucf.edu/REU/2014/p4_opticalFlow.pdf

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