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 focuses on detecting activity of an individual by using different sources of information - sensors. This finds great utility in the real world not only for purposes of tracking our daily activities but also in monitoring activities of others - like the elderly, patrol officers, etc for purposes of health-care, security, entertainment. 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. Apart from this we found that a major amount of focus has been given on the global motion of the entity(walking, running, etc) and not on the local motion (typing, writing, other hand movement activities). This is an area we would like to explore if time permits. At the end of the project we aim at achieving human activity recognition with better performance using a fusion of these sensors to train different models for learning.**

***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 wear where wearable sensors like accelerometers, gyrometers mounted on smartphones and smartwatches are used and other where egocentric cameras are used. The use of egocentric cameras on human activity detection has just cropped up. We here perform data fusion between these two domains in an attempt to improve the performance of the existing architecture.

II. Literature Survey

We performed a comprehensive literature survey as a part of this project which has been summarised and tabulated in table 2. In this first part of the project we are going with the classifiers proposed by the authors in the literature survey. We will be providing a detailed comparison among classifiers in the next segment of the project. RDF and Decision trees have proven to be succesful 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.
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. Chimeric Data Set

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

We divide our classification problem into two hierarchical stages:

A. Stage 1

We first go for the easier problem of classifying a persons state into two broad categories:

1) **Static**
2) **Motion**

We predict this stage to have a high accuracy based on the literature survey as well as after visualizing the data obtained from the sensors. 1

B. Stage 2a

Once an activity is classified as **Static**. We further perform a binary classification into:

1) **Standing**
2) **Sitting**

C. Stage 2b

Once an activity is classified as **Motion**. We further perform a binary classification into:

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

Performing the classification in the above hierarchy saves us both runtime and memory as compared to directly running a 7 class classification problem. It also allows us to have a host of different classifiers at different stages with customized parameters to give us better results. This has also been depicted in figures ahead.

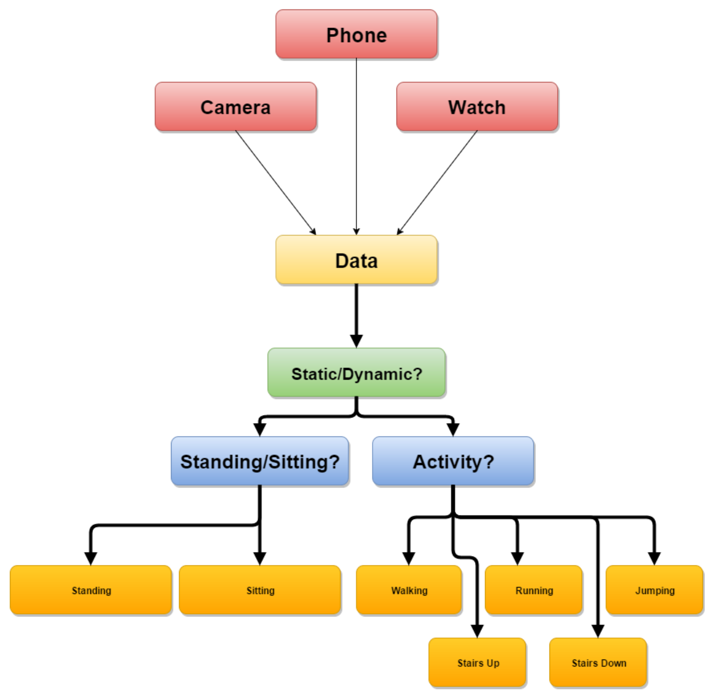

Fig. 1. Classification of Activities

D. Procedure

Our approach towards the problem in this project can be explained as follows:

1) We first perform individual tasks of solving the entire problem using each available sensor individually. Do initial testing on public datasets to obtain the achievable results.
2) Within each individual approach, we work on the available dataset and the one we will be collecting in course of this project.
3) We try a varied range of classifiers and grid searches to come up with the best performers for specific classifications.

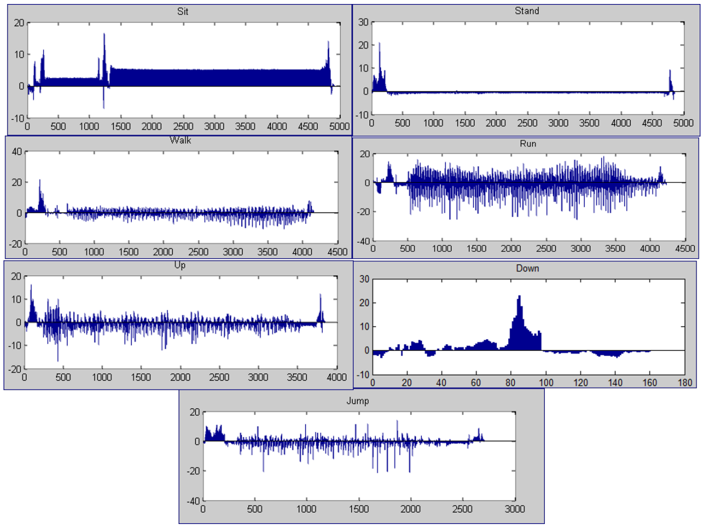

Fig. 2. Plotted accelerometer raw data for different activities

V. Smart Phone Based Classification

A. Highlights

The smartphone 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 we performed manual cleaning of the data. There was a lot of noise in the data, for which we performed one dimensional fourth order median filtering on the collected dataset. The dataset was divided into windows and over each window the features are extracted. Keeping a window size of 40 and an overlap of 50% between the windows, the total dataset comprises of around 9000 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 are enlisted below.
A visualization of the acclerometer data on various activities in our dataset is depicted in figure 2.

C. Features

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

1) Time Domain: Mean, Standard Deviation, Maximum, Minimum, Signal Magnitude Area, Inter-quartile range, Entropy, Autorregresion 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

Extracting these features gave us a a total of 148 features.

D. Results

The results for different stages among the tested classifiers are shown in Table II.

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	55.83
SVM(Polynomial)	26
NB	39.05

VI. Smart Watch Based Classification

A. Highlights

The smartwatch is worn at the wrist. This ensures that we record activity of the arms carefully to capture the signature movements for any activity.

B. Process

The raw data from the smartwatch was cleaned in the same way as smartphone data. The same median filter was applied. The same method of feature extraction was used. Keeping a window size of 40 and an overlap of 50% between the windows, the total dataset comprises of around 3000 samples. For our dataset, raw data and the features were extracted the same way as above.

C. Features

For both the databases the features listed in smartphone were extracted and used.

D. Results

Extracting these features gave us a a total of 148 features.

E. Results

Results among the tested classifiers 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

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

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 increments by including the visual data in the decision making process. Computer vision provides a source of information which comes handy in detecting motion in real time called optical flow. Optical flow or optic flow is the pattern of apparent motion of objects, surfaces, and 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. Since our approach towards extracting the features built upon Optical Flow are based on the same library tool (openCV) we avoid spending time at this stage on running tests on the dataset and go by their word in their paper[**?**] which is appreciably high. We will be evaluating it in the second evaluation of the project and using it for training our final classifier as well. For now we focus on our own collected dataset.

2) Collected Dataset: The Dataset consists of annotated 175 videos of roughly 1 minute each (specific activity) collected by 25 individuals performing the above 7 activities (mentioned in the dataset section above).We have cleaned 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 it's 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 preprocessing is done to remove highly abnormal values and smoothen the optical flow by applying a smoothing filter. (to remove noise)

D. Results

Following the literature survey we chose SVM as our first choice for a classifier for stage 1. We are providing 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 tableIV (Stage 1). We next try out naive bayes and RDF for Stage 1 on the camera data to obtain better accuracies. as shown in table IV. For stage 2 we again perform the same above mentioned steps. For Stage2(a) i.e. intra-static classification we obtain best results on using a RBF SVM as depicted in tableIV. For Stage2(b) i.e. intra-dynamic classification Though the results are overly pleasing we must take in to consideration that only 21 videos were used at this stage from 3 subject (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 an FPS of **6** frames per second.

VIII. Sensor Fusion

Based on the above obtained accuracies and classification performances by the various classifiers on the various sensors we propose some ways to combine these results:

A. Hierarchical Greedy Tree

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

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

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

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.

B. *Majority Voting Tree*

In this model we do a majority voting among all classifiers at each step to decide the classification being proposed. In case of all three classifiers return different output, we choose the output given by the classifier with maximum training accuracy.

IX. FINAL EVALUATION

The confusion matrix for different stages is given in 4. We performed data fusion using two methods. One is majority voting which has a disadvantage that even if a sensor predicts the correct class but the other two predict the wrong class, the final predicted class would be the wrongly predicted one. We need a way to somehow use a weighted combination of classifiers. For now we simply use class predicted by the most successful classifier over the dataset.

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. 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.