Academic Science

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 6, Issue 9
September 2017

Human Activity Recognition by Smartphone using Machine Learning Algorithm for Remote Monitoring

[1]Sajanraj T D
Dept. of Computer Science and Engineering
[1]Vidya Academy of Science Technology
Thrissur,Kerala, India

[2]Beena M V
Assistant professor
Dept. of Computer Science and Engineering
[1]Vidya Academy of Science Technology
Thrissur,Kerala, India

Abstract—*Human Activity Recognition has a lot of applications such as patient monitoring, rehabilitation and assisting disabled. When mobile sensors are hold to the subject's body, they permit continuous monitoring of numerous signals patterns from the phone. This has appealing use in healthcare applications. In order to improve the state of global healthcare, numeroushealthcare devices have been introduced that allows doctors to perform remote monitoring and increase users motivationand awareness. Now a days smart phones become a part of our day to day life. The best way to implement the idea throughSmart phones. The smart phones contains various built-in sensing units like accelerometer, gyroscope, GPS, compass sensorand barometer. Using this a system is designed to capture the states of a user. Here the mobile sensor is used as an input deviceand estimate the human motion activity using data mining and machine learning techniques. Here we use the KNN classificationalgorithm in the activity recognition system which support the training and classification using accelerometer data only. We can predict the performance of these classifiers from a series of observations on human activities like walking, running, step up, and step down in an activity recognition system.*

Index Terms—Accelerometer Sensor, Android Smart Phone, Feature extraction, Classification,KNN.

I. INTRODUCTION

Human activity recognition(HAR) has recently drawn extensive attention in various areas such as pervasive, mobile and context aware computing [1], [2], [3], [4], due to the advancement of techniques, including smart mobile devices, wireless communication networks, and machine learning. The main ingredient for successful human activity recognition might be in learning fruitful representations of sensor signals (which corresponds to feature extraction) in such a way that the characteristics of predefined actions are well identified by classifiers in mobile device.

Nowadays smartphones became more and more popular in human daily life. Most of the people used it for searching news, watching videos, playing games and accessing social network but there were many useful studies on smartphones. Activity recognition is one among the most important technologies behind many applications on smartphone such as health monitoring, fall detection, context-aware mobile applications, human based survey systems and home automation etc., Smartphone-based activity recognition system is a progressive area for research because they can lead to new types of mobile applications.

The recent technological movements in the area of sensors, processing units and wireless communication system have led to the development of smart wearable devices that are becoming more interesting and necessary part of our daily life. The unhealthy habitswhich are originated by modern lifestyle results in health degradation and lead us suffering from chronic diseases. The medical literature reveals the positive effects of physical activities and its potential to reduce the risk of chronic diseases. Wearable health care devices could play key role to mitigate the health related issues associated to inactivity.

Human activity monitoring systems monitors daily activities, keeps engage the user in more activities by escalating user's interest and thus helps people to adopt healthier lifestyle. The older HAR systems comprises of different body worn sensors. These sensors are tightly coupled with different parts of the human body and acquire data in accordance with their activity. This bunch of data is processed to classify different activities using machine learning algorithms. Machine learning method for activity recognition aims at less complexity and more accuracy in classification algorithm for real time activity recognition. Feature selection being the work horse of machine learning is considered as one of the key steps for activity classification that efficiently improves the classification accuracy. Smart phone based method is adopted instead of body worn sensors can also be reduce the actual cost 0f additional sensors.

The main purpose of the feature selection step is to select the minimum subset of the most suitable and most relevant features. The proper selection of features significantly improves the accuracy of classification. Redundant and irrelevant features contribute nothing to the classifier result. Besides this, these weak features results in increasing computational complexity and degradation of classification performance. Feature selection method choose the best features, based on correlation among them, for efficient performance of distinguishing one activity from other activities.

Training process is always necessary when a new activity is added in to the system. The same algorithm parameters are needed to be trained and fine-tuned when the algorithm runs on different devices with various built-in sensors. However, labelling a training data (time-series data) is a time consuming procedure and it is not always possible to label all the training data by the users. As a result, we present an active learning technique to accelerate the training process. Given a KNN classifier, an active learning technique intuitively queries the unlabelled training samples and learns the parameters from the correct labels answered by the human. In this way, users will label only the samples that the algorithm demanded to do and the total amount of required training samples is reduced.

II. RELATED WORKS

Human Activity Recognition (HAR) is an important and active area of research. For instance Fernando G.D Silva [5] developed a system for recognition of basic movements of human body using accelerometer data gathered from tri-axial accelerometer sensor integrated with a sports watch. The 19 and 21 most significant features from a group of 31 features were selected by using Fisher Discriminant Ratio (FDR) and Principal Component Analysis (PCA) respectively. The selected features were classified using Multilayer perceptron, K-nearest neighbours and Support Vector Machine (SVM).

Another Machine learning base HAR classifier was proposed by Wallace Ugulino [6] for the classification of five different activities (sitting, standing, sitting down, standing up and walking) using body worn accelerometer data gathered from 4 subjects. Best first selection method was performed to select 12 best features for activity classification. Their experimental result demonstrates that the best classification rate of 99.4% was obtained for C.45 decision tree in connection with AdaBoost. Autoregressive model can also be used to perform activity recognition task

Zhen-Yu [7] presented an autoregressive model (AR) for human activity recognition from tri-axial accelerometer data. AR coefficients were extracted as features for the classification of different activities (run, still, jump, and walk). Recognition rate of 92.25% was achieved by using Support vector machine classifier with five folds cross validation method.

SamadZabihi [8] performed activity recognition by transforming the accelerometer data (x, y, z) to spherical coordinate system (r, Ø, θ) and extracting features from transformed data. 36 derived features with sliding window approach (8 samples) along with 12 accelerometer readings in spherical coordinates were fed into a multilayer perceptron neural network of 48 hidden layers. The results demonstrate that 99.9% activity recognition was achieved. Huafei Wang [9] reduced the dimensionality by Principal component analysis method and selected 3 feature from a dataset of 12 features. Their experimental results illustrates that the pre-processing (zs-core scaling and 0-1 scaling) highly influence the precision and recall. An enhanced precision and recall rate was gained by using SVM and K-mean classifier with 90% training and 10% testing data.

III. PROBLEMS OF PREVIOUS METHODS

Wearable sensor [10] unit for activity recognition is composed of several body worn sensors. These sensors are worn on different parts of the body to collect raw data of physiological signals in accordance with the body motion and transmit the data to central processing unit for activity recognition and classification. The microprocessor unit perform data mining methods and makes conclusion about the user activity (running, walking, sitting, standing etc.). For real time activity recognition system like fall detection and activity tracking in elderly people, the wearable processing unit along with classification algorithm should be fast enough to make real time decision [11], [12]with high accuracy rate. In order to achieve the goal of high accuracy, high computational capability and low complexity for the recognition of activities in daily life, the processing data should be brought to simple form with minimum numbers of features which are free of redundancy and irrelevancy problems. Our aim is to select the best possible minimum features subset with high classification performance that would potentially increase the classification accuracy and computational capability of the wearable devices for daily life activity recognition.

IV. PROPOSED WORK

This research work primarily aims to develop the human activity recognition system for idle

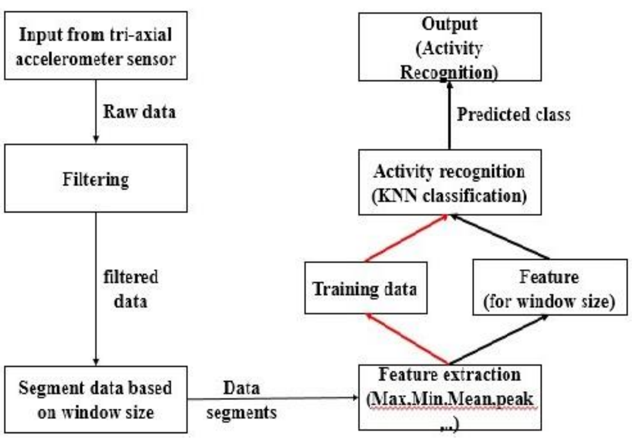

Fig. 1. Block diagram

and walking, step up, step down and the running activities. The framework has been divided into three stages, namely training stage, classification and clustering and activity recognition stage. The training stage is concerned with building of classes based on training data from accelerometer sensors from mobile. Form the cluster for each classes based on minimal distance to increase the efficiency of the system. In the recognition part theKNN is used to find the best suited class.

Block diagram for activity recognition contains following subtasks:

- Data acquisition: In data acquisition accelerometer reading generated during each activities are recorded and sent to the remote system.
- Data Pre-processing: Removal of noise and normalization of data.
- Segmentation: Segment data based on window size. Here the window is the time interval within each activity.
- Feature Generation: In feature generation, steps find out relation within the data values maximum, minimum, standard deviation and the mean.
- Feature Selection: Due to a large number of features, the array is reduced before classifier construction. This involves selection of the most significant features for classification.
- Activity recognition: A nearest-neighbour classification object, where both distance metric (”nearest”) and number of neighbours can be altered. The object classi-

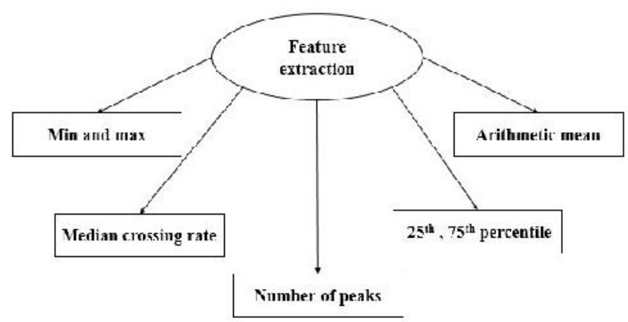

Fig. 2. Feature design

fies new observations using the predict method. The object contains the data used for training, so it is used compute re-substitution predictions.

A. Criteria for Selecting the Sensors

Various advancements are investigated with a specific end goal to locate the best and morally worthy detecting innovation for this endeavour. Accelerometers [13] are truly modest and little in size (couple of millimetres). In moral sense they don't attack protection of the client.

As the name recommends accelerometers are sensors which measure speeding up of an article. As Newton's second law (F = ma) clarifies, the speeding up of the article is relative to the power which the item is under. Each minute, our human body is under some type of power. When we are unmoving the main power which is following up on our body parts are earth's gravitational power. Yet, when we are moving (barring any sort of vehicular transportation) our body parts are under the power which is created by our muscles.

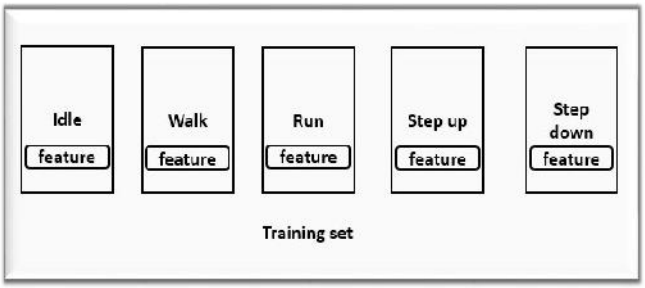

Fig. 3. Design of training class

The top to bottom comprehension of the mentioned subjects are critical to make the human stance acknowledgment framework exact. . Normally human body framework would not encounter steady speed (unless the human ison a vehicle moving in consistent speed, which is not pertinent variable in this study). Along these lines any movement a human makes has increasing speed. In this manner utilizing accelerometers one can re-track the human movement decently and precisely. The accompanying criteria were [14] considered when selecting an accelerometer. Most extreme quantifiable speeding up, linearity of the sensor yield with the genuine increasing speed, recurrence reaction, number of delicate tomahawks, size, mass, cost, sturdiness, safety to outside clamours, for example, attractive electrical commotion and warm affectability. Capacitive accelerometers are delicate in nature and not perfect for high effect applications. However, when the venture objective which is stance following of human elderly subject is taken into thought, the creator can securely respect the capacitive accelerometers can't discover an occasion which will make the accelerometers experience a G power of 7 or more. In this manner beginning uncertainty of strength of the capacitive accelerometers can be disposed of. At the point when the expense, size and recurrence reaction of the capacitive accelerometers are taken into record, they are more than sufficient for this endeavour. After watchful thought the capacitive accelerometer innovation is picked as the most applicable accelerometer innovation for this exploration [16].

V MACHINE LEARNING ALGORITHMS

K-Nearest Neighbour is a supervised learning algorithm where the result of new instance query is classified based on majority of K-Nearest neighbour category and belongs to the most popular algorithms for activity recognition. The purpose of KNN algorithm is to stores the available objects and classifies a new object

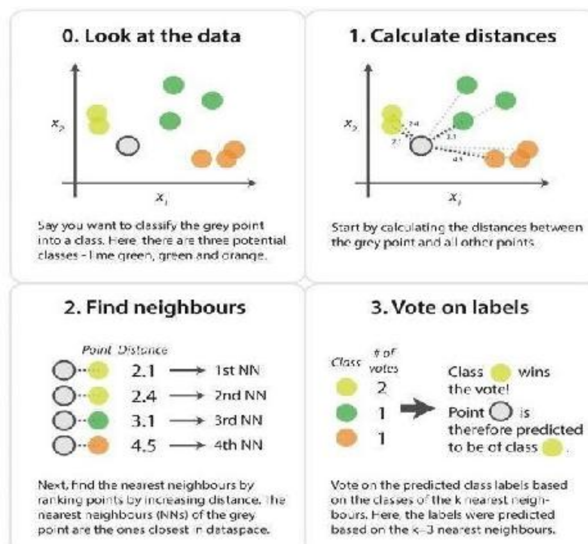

Fig. 4. KNN Algorithm

based on attributes and training samples. The KNN classifiers do not use any model to fit and only based on memory. KNN algorithm used neighbourhood classification as the prediction value of the new query instance. In activity recognition, the k-nearest neighbour algorithm (KNN) is a non-parametric (or distribution free) method for object classification based on nearest training samples in the given feature space.

KNN is a type of memory-based learning, where the function is only approximated locally and all computation is delayed until classification. The KNN algorithm is the simplest in machine learning algorithms.Here, an object is classified by most votedneighbours, with the objects which is assigned to the class amongst its k nearest neighbours (k is small positive integer). If k=1, then the object is merely assigned to its nearest neighbourclass. The vectors in training examplesbelongs to multidimensional feature space, each with a labelled class. The training (data pre-processing) the algorithm phase consists of stored feature vectors and the training samples with class label. In the classification phase of activity recognition system, k is a user-defined constant, and an unlabelled vector is classified by assigning most frequent label among the k training samples.

B. Data Collection

Data is collected through the application we created called activity logger and this component is responsible for collecting the training data for each activity separately from sensors, i.e. accelerometer data. In this application, user will selects the activity to be performed, keep the phone into the packet and starts to perform the related activity. For each activity, this application is responsible for creating different training data files in which raw data from the 3-axes of the accelerometer is being logged.

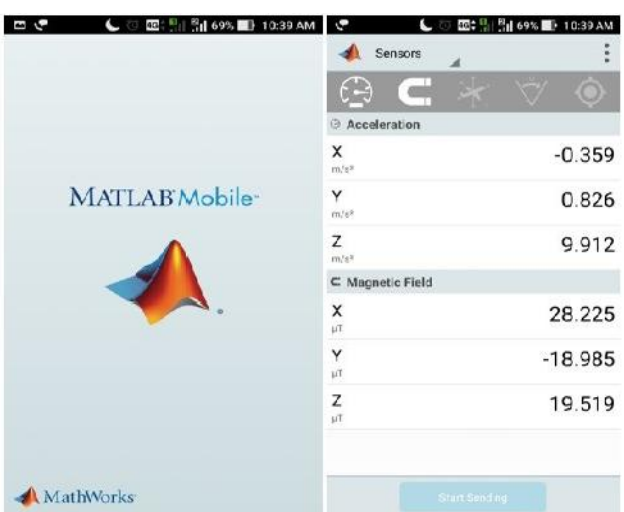

Fig. 5. Matlab mobile application.

The collected data may contain noise and the data collection is processed the data to eliminate the noise by applying noise filters and low-pass filter technique. Time-domain and frequency-domain features [15] have been extensively used to filter relevant information within acceleration and rotation signals. In this paper, we used four features: MIN, MAX, MEAN and Standard deviation. By these statistical operations features were calculated.

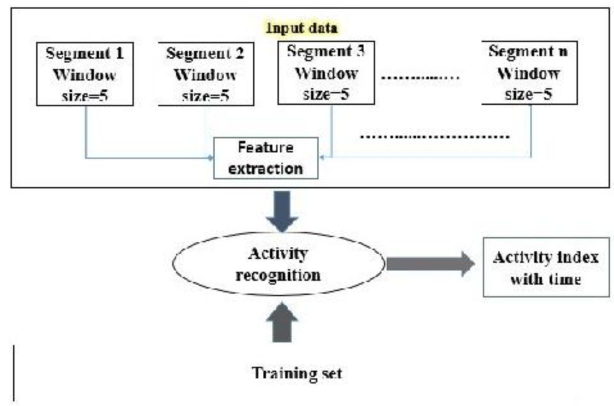

Fig. 6. Design of input for activity recognition

C. Data Pre-processing in KNN

The main objective of the pre-processing step is to define activity sets from the training data based on the mentioned features. Instead of comparing all the data in the training set, we compare the test data only with the compact training data set that we selected from the original training set. During the pre-processing step, for each feature and for each activity, compact training sets are created. For each feature, except the standard deviation, K datapoints are selected from the training data. For instance, for the minimum feature set, K minimum data points are selected from the training data. Likewise, we create a maximum set by selecting the K maximum data points. The average value of the training data is calculated and the nearest K data points are included in the average set. For the standard deviation set, standard deviation value of training data for each activity is calculated. However, at the same time accuracy of the results are expected to decrease with smaller value of K, so that there is an important

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 6, Issue 9
September 2017

trade-off between accuracy and execution time considering the value for K.

D. Feature Selection

Feature construction [15],[16] is the key action in pre-processing the data and it is important not to lose any information in this process. To improve the accuracy of the Machine Learning algorithms, we perform statistical calculations on raw accelerometer data before using the data for training the classifier. The Feature selection [17] will reduce the processing cost by remove irrelevant and redundant features, whereas ensuring the accuracy of recognition. The redundant features are those which do not contribute information to the recognition process. In other words, the features which provide no/irrelevant to the selected current features set should be eliminated. As a result, it will improve model interpretability, shorter training time, and enhance generalization by reducing over fitting.

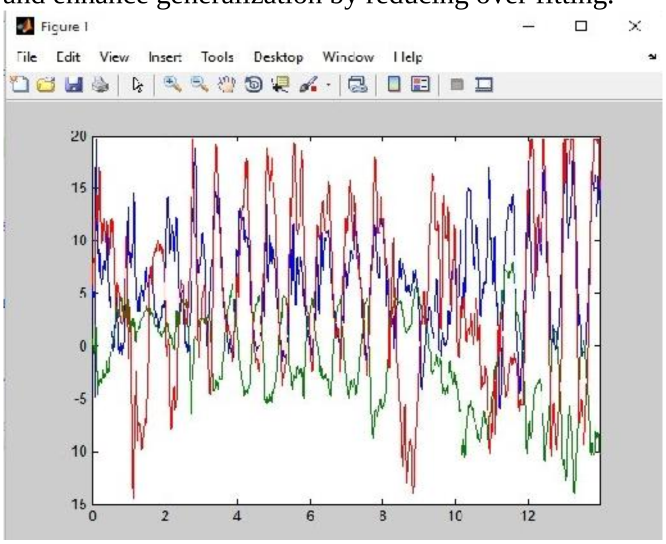

Fig. 7. Input data plotted against acceleration and time axis.

In addition to the raw accelerometer data we calculate Vector Magnitude V where x(i), y(i), and z(i) are the ith acceleration sample of the x, y, and z axis in each window respectively. The definition of the time domain features for a given window W = d1, d2…dn are listed below:

- Maxd1, d2 …dn and Mind1, d2, ..dn.
- Arithmetic mean for each accelerometer axis and V.
- Number of Peaks
- Median crossing rate which is the number of times the signal changes from below the median to above the median or vice versa.
- The 10th, 25th, 50th (Equation 3) , 75th, 90th percentile
- Correlations between Accelerometer axes and Vector Magnitudes

E. Classification in KNN

In the classification step, we collect test data, in other words we segment the data during a window with a predefined size. After the window is filled, classification phase will starts, and average, minimum, maximum, standard deviation values of the data in the window is calculated and these values are compared one by one with the values in the compact training sets which were created during the pre-processing step. K nearest sample to test data is selected from training sets and voting is done by looking at the final list of activities. We label the data in the related window as the activity for which we have maximum amount of data in the final K set. The one which is closer to the standard deviation of that particular window is selected as the recognized activity by the standard deviation feature. At the end, we have four labels from voting results of each feature. We label the window as the activity for which we have the highest vote and finalize the classification.

When the window is filled, classification phase will starts, and average, minimum, maximum, standard deviation values of the data in the window is calculated and these values are compared one by one with the values in the compact training sets which were created during the pre-processing step. K nearest sample to test data is selected from training sets and voting is done by looking at the final list of activities. We label the data in the related window as the activity for which we have maximum amount of data in the final K set. The one which is closer to the standard deviation of that particular window is selected as the recognized activity by the standard deviation feature. At the end, we will get four labels from voting results of each feature. We label the window as the activity for which we have the highest vote and finalize the classification.

Academic Science

International Journal of Innovations & Advancement in Computer Science
IJIACS
ISSN 2347 – 8616
Volume 6, Issue 9
September 2017

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

Implementation of system is carried out using MATLAB 2017a. The proposed methodology is tested for students in the college for walk 90% of the data in dataset is used for training and the remaining 10% for validation and testing. The recent technological advances in the area of sensors, processing units and wirelesscommunication [18],[19] system have led to the development of smart wearable devices that are becoming more interesting and necessary part of our daily life. Here the system is implemented using LAN to recognize the activities.

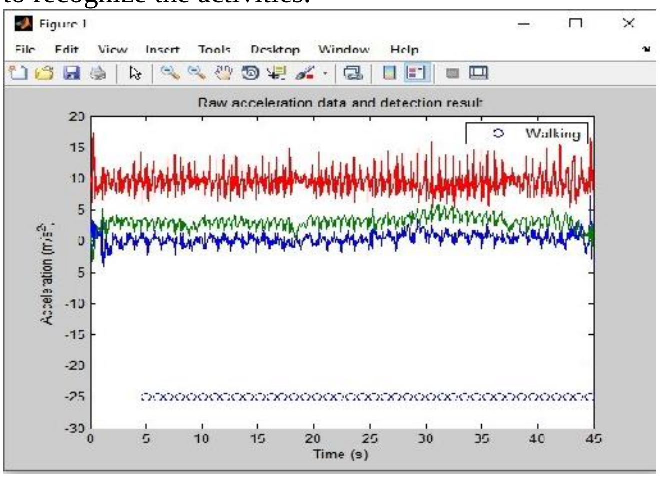

Fig. 8. Output data plotted against acceleration and time axis.

The result generated using the system is plotted below. The system works in different areas like inside buildings, road and straightway with an error varying from 5 to 15%.The output generated when tested in campus is shown in the figure8.

TABLE 1
Confusion Matrix

Window size	0.5			1		
Sampling Interval	10	50	100	10	50	100
k	88.5	87.2	87.1	89.1	89.5	88.7
	90.1	92.5	96.3	96.5	98.4	95.6

F. Performance Parameters

In predictive analytics, a table of confusion (sometimes also called a confusion matrix), is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives. This allows more detailed analysis than mere proportion of correct guesses (accuracy). Accuracy is not a reliable metric for the real performance of a classifier, because it will yield misleading results if the data set is unbalanced (that is, when the number of samples in different classes vary greatly)[20].

Confusion matrix is a visualization tool typically used in supervised learning techniques. One advantage of confusion matrix is that it is easy to see if the system is confusing two classes. Each column of the confusion matrix represents the instances in a predicted class, while each row represents the instances in an actual class. A confusion matrix contains information about known class labels and predicted class labels. Compared to the performance of activities like running, lying down, standing and sitting, the KNN classifier presents slightly worse performance for walking, where this activity is sometimes classified as running or standing. However, the overall performance for clustered KNN classification is around 96% accuracy considering all activities obtained from different environment.

TABLE 2
Accuracy for Each Test Samples

Activity	NO.of Samples	Accuracy
Walk	50	99.05
Run	60	98.95
Step up	30	98.26
Step down	45	97.35

VII. CONCLUSION

In this paper, we proposed an activity recognition system working on Android platforms by an application called Matlab Mobile App that supports on-line training and classification while using only the accelerometer data for classification. The GPU system provides more efficiency in the case of time complexity as compared to CPU systems .The performance of on-line classification of KNN classifier is evaluated first then a clustered KNN method is used.

The clustered KNN classification exhibit a much better performance than the KNN classifier in terms of accuracy on android platforms with limited resources. We also evaluated the performance of clustered KNN in terms of execution times and obtain 98.4% accuracy. As we expected, classification execution times are considerably reduced as K parameter is decreased. Additionally, classification times are highly dependent on the device model and capabilities as well.

MATLAB provides the better solution for Activity recognition. Developing such system for helping remote monitoring of patients.

REFERENCES

[1] T. Chouhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, A. Rahimi, A. Rea, G. Borriello, B. Hemingway, P. Klasnja, K. Koscher, J. A. Landay, J. Lester, D. Wyatt, and D. Haehnel, The mobile sensing platform: An embedded activity recognition system, *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 3241, 2008.

[2] C. R. Wren and E. M. Tapia, Toward scalable activity recognition for sensor networks,*inLocationa and Context Awareness LNCS*, vol. 3987. Springer, 2006, pp. 168185.

[3] K. Van Laerhoven and K. Aidoo, Teaching context to applications, *Personal and Ubiquitous Computing*, vol. 5, pp. 4649, 2001.

[4] M. Stikic and B. Schiele Activity recognition from sparsely labeled data using multi-instance learning, in*Location and Context Awareness LNCS* , vol. 5561. Springer, 2009, pp. 156173.

[5] G.D Silva, Fernando Ginez, and ElisabeteGaleazzo. ”Accelerometer based intelligent system for human movement recognition.”*Advances in Sensors and Interfaces (IWASI), 2013 5th IEEE International Workshop on. IEEE*, 2013

[6] Ugulino, Wallace, et al. ”Wearable computing: accelerometers data classification of body postures and movements.” *Advances in Artificial Intelligence-SBIA 2012. Springer Berlin Heidelberg*, 2012. 52-61.

[7] He, Zhen-Yu, and Lian-Wen Jin. ”Activity recognition from acceleration data using AR model representation and SVM.” *Machine Learning and Cybernetics, 2008 International Conference on*. Vol. 4. IEEE, 2008.

[8] Zabihi, Samad, PouyanNasrollahi, and Mohammad Bazmara. ”Human Activity Recognition Using Accelerometers Values in Different Coordinate Systems.” *World Appl. Program* 4.12 (2014): 237-240.

[9] Huafei Wang, Jennifer Wu”Classification of Human Posture and Movement Using Accelerometer Data ” in *international Conference on Innovations in Computing & Networking (ICICN16), CSE, RRCE* ,ISSN: 0975-0282

[10] DelaramYazdansepas, Anzah H. Niazi, Jennifer L. Gay, Frederick W. Maier, LakshmishRamaswamy, Khaled Rasheed, Matthew P. Buman”A Multi-Featured Approach for Wearable Sensor-based Human Activity Recognition” in *IEEE International Conference on Healthcare Informatics*,2016 978-1-5090-6117-4/16.

[11] A. Ferrero, Strategic plans, Presidents Perspectives, IEEE Instrum. Meas. Mag., vol. 11, no. 3, p. 6, 2008.

[12] R. San-Segundo, J.M. Montero, R. Barra-Chicote, F. Fernndez, and J.M. Pardo, Feature extraction from smartphone inertial signals for human activity segmentation, Signal Processing, vol. 120, pp. 359-372, Mar. 2016..

[13] R. San-Segundo, J. M. Montero, J. Moreno-Pimentel, and J. M. Pardo, HMM adaptation for improving a human activity recognition system, Algorithms, vol. 9, no. 3, p. 60, 2016, doi:10.3390/a9030060.

[14] A. Mannini and A. M. Sabatini, ”Machine Learning Methods for Classifying Human Physical Activity from OnBody Accelerometers,” Sensors, vol. 10, pp. 1154-1175, Feb 2010.

[15] Zhang, Shumei, Paul McCullagh, and Vic Callaghan. ”An Efficient Feature Selection Method for Activity Classification.” Intelligent Environments (IE), 2014 International Conference on. IEEE, 2014.

[16] Tonchev, Krasimir, et al. ”Recognition of Human daily activities.” Communication Workshop (ICCW), 2015 IEEE International Conference on. IEEE, 2015.

[17] Ghasemzadeh, Hassan, et al. ”Power-Aware Computing in Wearable Sensor Networks: An Optimal Feature Selection.” Mobile Computing, IEEE Transactions on 14.4 (2015): 800-812.

[18] A. Varshavsky, M. Y. Chen, E. De Lara, J. Froehlich, D. Haehnel, J. Hightower, A. Lamarca, F. Potter, T. Sohn, K. Tang, et al. Are gsm phones the solution for localization? In Mobile Computing Systems and Applications, 2006. WMCSA06. Proceedings. 7th IEEE Workshop on, pages 3442. IEEE, 2005.

[19] M. Youssef and A. Agrawala. The horuswlan location determination system. In Proceedings of the 3rd international conference on Mobile systems, applications, and services, pages 205218. ACM, 2005.

[20] E. Foxlin. Pedestrian tracking with shoe-mounted inertial sensors. Computer Graphics and Applications, IEEE, 25(6):3846, 2005.