iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45630

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Human Activity Recognition

Lohitha Maradani
Electronics and Communication Engineering (ECE), Chaitanya Bharathi Institute Of Technology (CBIT)

Abstract: ***Human Activity Recognition (HAR) is one of the active research areas in computer vision as well as human computer interaction. However, it remains a very complex task, due to unresolvable challenges such as sensor motion, sensor placement, cluttered background, and inherent variability in the way activities are conducted by different humans. Human activity recognition is an ability to interpret human body gesture or motion via sensors and determine human activity or action. Most of the human daily tasks can be simplified or automated if they can be recognized via HAR system. Typically, HAR system can be either supervised or unsupervised. A supervised HAR system required some prior training with dedicated datasets while unsupervised HAR system is being configured with a set of rules during development. HAR is considered as an important component in various scientific research contexts i.e. surveillance, healthcare and human computer interaction.***
Keywords: supervised, unsupervised, datasets, human computer interaction, sensor motion

I. INTRODUCTION

Human activity recognition has been studied for years and researchers have proposed different solutions to attack the problem. Existing approaches typically use vision sensor, inertial sensor and the mixture of both. Machine learning and threshold-base algorithms are often applied. Machine learning usually produces more accurate and reliable results, while threshold-based algorithms are faster and simpler. One or multiple cameras have been used to capture and identify body posture. Multiple accelerometers and gyroscopes attached to different body positions are the most common solutions.

Human activity recognition and prediction is a technology of automatically detecting human activities observed from a given video. Human activity recognition is applied to surveillance using multiple cameras, dangerous situation detection using a dynamic camera, human-computer interface. However, it is important to prevent dangerous activities and accidents such as crimes or car accidents from occurring, and the recognition of such activities after occurred is insufficient. With the recent progress in wearable technology, unobtrusive and mobile activity recognition has become reasonable. With this technology, devices like smartphones and smartwatches are widely available, hosting a wide range of built-in sensors, at the same time, providing a large amount of computation power. Overall, the technological tools exist to develop a mobile, unobtrusive and accurate physical activity recognition system. Therefore, the realization of recognizing the individuals' physical activities while performing their daily routine has become feasible. So far no-one has investigated the usage of light-weight devices for recognizing human activities. An activity recognition system poses several main requirements. First, it should recognize activities in real-time. This demands that the features used for classification are computable in real-time. Moreover, short window durations must be employed to avoid delayed response.

II. RELATED WORK

Local image and video features have been successfully used in many action recognition applications such as object recognition, scene recognition and activity recognition. Local space-time features capture characteristic shape and motion information for a local region in video. They provide a relatively independent Local feature methods representation of events with respect to their spatio-temporal shifts and scales as well as background clutter and multiple motions in the scene.

These features are usually extracted directly from video and therefore avoid possible dependencies on other tasks such as motion segmentation and human detection. In the following, we first discuss existing space-time feature detectors and feature descriptors. Methods based on feature trajectories are presented separately, since their conception from space-time point detectors. Finally, methods for localizing actions in videos are discussed.

The proposed method includes extracting space-time local features from video streams containing video information related to human activities. clustering the extracted space-time local features into multiple visual words based on the appearance of the features. computing an activity likelihood value by modelling each activity as an integral histogram of the visual words. The visual words may be formed from features extracted from a sample video by using a K-means clustering algorithm. Integral bag-of-words is a probabilistic activity prediction approach that constructs integral histograms to represent human activities. In order to predict the

ongoing activity given a video observation 'O' of length t, the system is required to compute the likelihood value $P(O|A_p,d)$ for all possible progress level d of the activity A_p. What is presented herein is an efficient methodology to compute the activity likelihood value by modelling each activity as an integral histogram of visual words. The integral bag-of-words method is a histogram-based approach, which probabilistically infers ongoing activities by computing the likelihood value $P(O|A_p,d)$ based on feature histograms. The dynamic back of wars method according to the present invention utilizes an integral histogram for calculating the similarity between inner segments (ie, P ($O^{\Delta t}$ | A, Δd)). Fig. 1 shows the integral histogram which enables efficient construction of the histogram of the behavior segment Δd and the histogram of the video segment Δt for any possible (Δd, Δt). Let [a, b] be the time interval of Δd. The histogram corresponding to Δd is calculated as follows.

$$F_p(t,d) = \sum_{\Delta t}\sum_{\Delta d} [F_p(t-\Delta t, d-\Delta d) \cdot M\left(h_{\Delta t}(O), h_{\Delta d}(A_p)\right)]$$

Fig.1 Formula of integral histogram.

III. METHODOLOGY

A. *Activity Model*

We model each activity based on the sliding window strategy. Specifically, we divide the continuous sensor streams into fixed length windows. By choosing a proper window length, we assume that all the information of each activity can be extracted from each single window. The information is then transformed into a feature vector by computing various features over the sensor data within each window. Here, a window of length 2 seconds with a 50% overlap is used. We now describe two sets of features we incorporated in our recognition framework.

B. *Statistical Features*

The first set of features are statistical features computed from each axis (channel) of both accelerometer and gyroscope. Some of them have been intensively investigated in previous studies and proved to be useful for activity recognition. For example, variance has been proved to achieve consistently high accuracy to differentiate activities such as walking, jogging, and hopping. Correlation between each pair of sensor axes helps differentiate activities that involve translation in single dimension such as walking and running from the ones that involve translation in multi-dimension such as stair climbing. We also consider statistical features that have been successfully applied in similar recognition problems. Examples are zero crossing rate, mean crossing rate, and first-order derivative. These features have been heavily used in human speech recognition and handwriting recognition problems.

C. *Physical Features*

The second set of features are called physical features, which are derived based on our physical interpretations of human motion. In this work, Motion Node is placed at the subject's front right hip, oriented so the x axis points to the ground and is perpendicular to the plane formed by y and z axes. We assume that the sensor location and orientation are known a priori. Some of our physical features are computed and optimized based on this prior knowledge. Although this assumption limits the generalization of our physical features to be applied to other locations and orientations to some extent, it simplifies the problem and allows us to focus on designing features with strong physical meanings so as to better describe human motion1. It should be noted that the way to compute physical features is different from statistical features.

For statistical features, each feature is extracted from each sensor axis (channel) individually. In comparison, most of the physical features are extracted from multiple sensor channels. In other words, sensor fusion is performed at feature level for physical features

D. *Data Preparation*

1) *Extracting :* extracting space-time local features from video streams containing video information related to human activities.
2) *Clustering:* clustering the extracted space-time local features into multiple visual words based on the appearance of the features.
3) *Computing:* computing an activity likelihood value by modeling each activity as an integral histogram of the visual words.
4) *Predicting:* predicting the human activity based on the computed activity likelihood value.

Fig.2 indicates the flow chart representation of the above processes

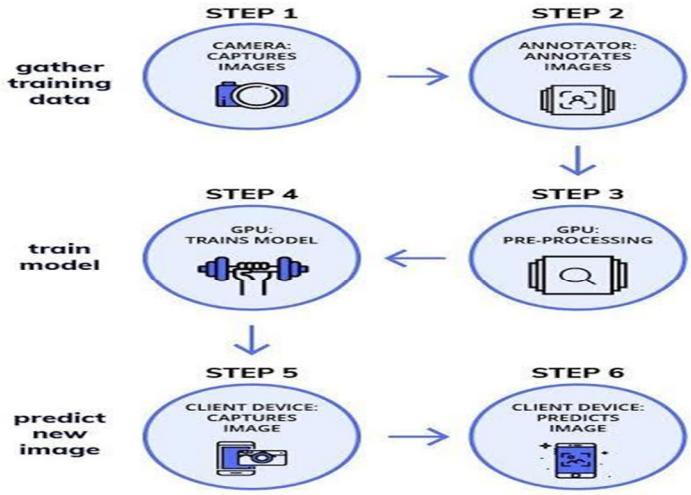


Fig.2 Data Preparation Process

IV. IMPLEMENTATION AND RESULTS

A. Data Collection

Machine learning is the new big thing in the world of computer science. The motivation behind this project is to implement machine learning algorithms in real-world data sets so that their accuracy can be studied and effective conclusions can bedrawn. This informational index has been gathered from chronicles of 30 human subjects caught by means of cell phones empowered with installed inertial sensors. Many AI courses utilize this information for educating purposes. This is a multi-arrangement issue. The informational collection has 10,299 lines and 561 segments. There are thirty volunteers of age group 18-50 years and examinations done in that. Every individual performed physical activities like WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a cell phone on the midsection. Utilizing its implanted accelerometer and gyrator, we caught 3-pivotal straight increasing speed and 3- hub rakish speed at a steady rate of 50Hz. The analyses have been video- recorded to name the information physically.

The sensor signals (accelerometer and whirligig) were preprepared by applying clamour channels and after that tested in fixed-width sliding windows of 2.56 sec. The sensor quickening signal, which has gravitational and body movement parts, was isolated utilizing a Butterworth low-pass channel into body speeding up and gravity. The gravitational power is accepted to have just low recurrence parts, subsequently a channel with 0.3 Hz cut-off recurrence was utilized. From every window, a vector of highlights was acquired by computing factors from the time and recurrence area.

B. Method Of Implementation

Convolutional neural network (CNN), one of well-known deep learning structures, was an innovation inspired by cat's visual cortex system, overcoming the vanishing gradient problem and the problem of unconnected weights in each layer of neural networks. The overall structure of CNN will be specified as following The first layer of CNNs is a Convolutional layer. If matrix f is a convolutional filter or kernel and matrix X represents input data, the processing of convolutional computation is that 13 filter f will be sliding along input data x with fixed stride, in which the operation of dot product is computed at each step and the output of each slide called feature map will be sent to next layer as input. Each layer owns totally different filters with the same functions, lowering dimensions and extracting essential information. More specifically, convolution operation originally generated from signal processing. Looking at the formula in the Fig.3, convolution is a sum(t) of a series of weighted values with a weighting function w(a) in which weights change with the variation of the value of (t-a) at the point of t

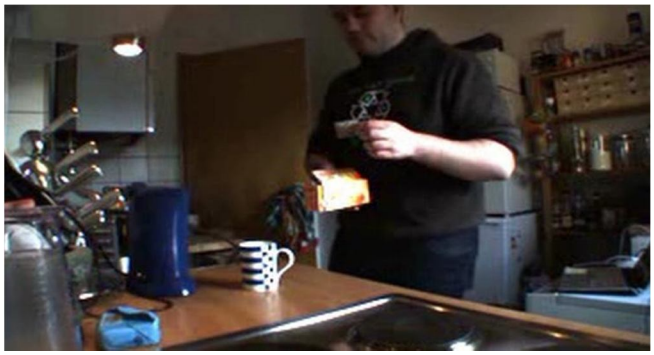
$$A^* = \arg\max_p \frac{\sum_d F_p(t,d)P(t \mid d)P(A_p,d)}{\sum_i \sum_d F_i(t,d)P(t \mid d)P(A_i,d)}.$$

Fig.3 Convolution Formula

C. Output Screens

Prediction: Add a Tea bag

Output : Future Event



V. RESULT ANALYSIS

The function of the pooling layer is reducing the computation and the numbers of parameters in the whole network, in another word, reducing dimensions. The general rule for the pooling layer is to keep the maximum or compute the average in each sliding window. Generally, behind the convolutional layer, the next is the activation layer (Rectified Linear layer), in which there are Rectified Linear Units with a nonlinear activation function in CNN structure. The most commonly used nonlinear activation function is ReLu, a simple thresholding operation as shown in Fig.4. If the ReLu function does not work well, Leaky ReLu and ELU functions are better recommendations. This layer is indispensable since it can accelerate the convergence of the whole neural network. Therefore, a good choice of nonlinear activation function would influence the performance of training neural networks. Figure shows six popular activation functions and their function plots

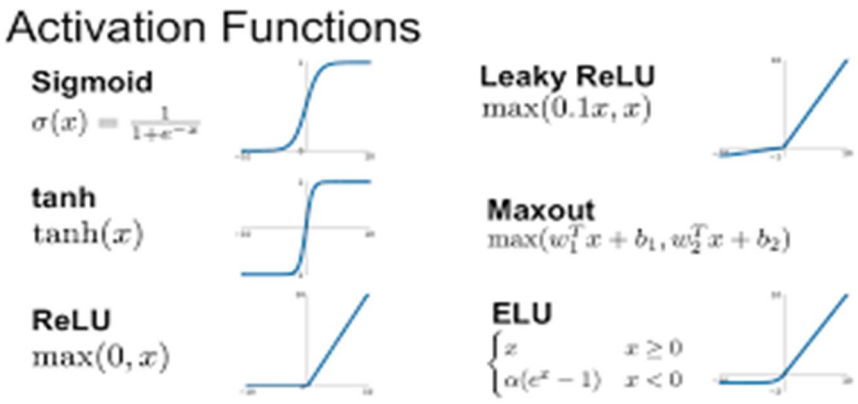


Fig.4 Simple Thresholding Operation

Next layer in CNN is the polling layer whose goal is to reduce dimensions, and summarize or refine representative information and features. There are usually two approaches to achieve this step. The first one is to select the maximum from each sliding block along input data, another one is averaged.

VI. TESTING AND VALIDATION

A. Design of test cases and scenarios

Convolutional neural network models were developed for image classification problems, where the model learns an internal representation of a two-dimensional input, in a process referred to as feature learning.

This same process can be harnessed on one-dimensional sequences of data, such as in the case of acceleration and gyroscopic data for human activity recognition. The model learns to extract features from sequences of observations and how to map the internal features to different activity types.

The benefit of using CNNs for sequence classification is that they can learn from the raw time series data directly, and in turn do not require domain expertise to manually engineer input features. The model can learn an internal representation of the time series data and ideally achieve comparable performance to models fit on a version of the dataset with engineered features.

This section is divided into 3 parts; they are

1) *Load Data:* The first step is to load the raw dataset into memory.There are three main signal types in the raw data: total acceleration, body acceleration, and body gyroscope. Each has three axes of data. This means that there are a total of nine variables for each time step. Further, each series of data has been partitioned into overlapping windows of 2.65 seconds of data, or 128 time steps. These windows of data correspond to the windows of engineered features (rows) in the previous section.This means that one row of data has (128 * 9), or 1,152, elements. This is a little less than double the size of the 561 element vectors in the previous section and it is likely that there is some redundant data. The signals are stored in the /*Inertial Signals*/ directory under the train and test subdirectories. Each axis of each signal is stored in a separate file, meaning that each of the train and test datasets have nine input files to load and one output file to load. We can batch the loading of these files into groups given the consistent directory structures and file naming conventions.
2) *Fit and Evaluate Model:* Now that we have the data loaded into memory ready for modeling, we can define, fit, and evaluate a 1D CNN model. We can define a function named *evaluate_model()* that takes the train and test dataset, fits a model on the training dataset, evaluates it on the test dataset, and returns an estimate of the models performance. This is exactly how we have loaded the data, where one sample is one window of the time series data, each window has 128 time steps, and a time step has nine variables or features. The output for the model will be a six-element vector containing the probability of a given window belonging to each of the six activity types. These input and output dimensions are required when fitting the model, and we can extract them from the provided training dataset.
3) *Summarize Results:* We cannot judge the skill of the model from a single evaluation. The reason for this is that neural networks are stochastic, meaning that a different specific model will result when training the same model configuration on the same data. This is a feature of the network in that it gives the model its adaptive ability, but requires a slightly more complicated evaluation of the model.

B. Validation

Validation means observing the behaviour of the system. The verification and validation mean, that will ensure that the output of a phase is consistent with its input and that the output of the phase is consistent with the overall requirements of the system. This is done to ensure that it is consistent with the required output. If not, we apply certain mechanisms for repairing and thereby achieved the requirements.

VII. CONCLUSIONS

The Human activity recognition has broad applications in medical research and human survey systems. In this project, we designed a smartphone-based recognition system that recognizes five human activities: walking, limping, jogging, going upstairs and going downstairs. The system collected time series signals using a built-in accelerometer, generated 31 features in both time and frequency domain, and then reduced the feature dimensionality to improve the performance. The activity data were trained and tested using 4 passive learning methods: quadratic classifier, k-nearest neighbour algorithm, support vector machine, and artificial neural networks.

To collect the acceleration data, each subject carries a smartphone for a few hours and performs some activities. In this project, five kinds of common activities are studied, including walking, limping, jogging, walking upstairs, and walking downstairs. The position of the phone can be anywhere close to the waist and the orientation is arbitrary. According to a previous study, body movements are

constrained within frequency components below 20Hz, and 99% of the energy is contained below 15 Hz. According to Nyquist frequency theory, a 50 Hz accelerometer is sufficient for our study. A low-pass filter with 25Hz cut-off frequency is applied to suppress the noise. Also, due to the instability of the phone sensor.

VIII. ACKNOWLEDGMENT

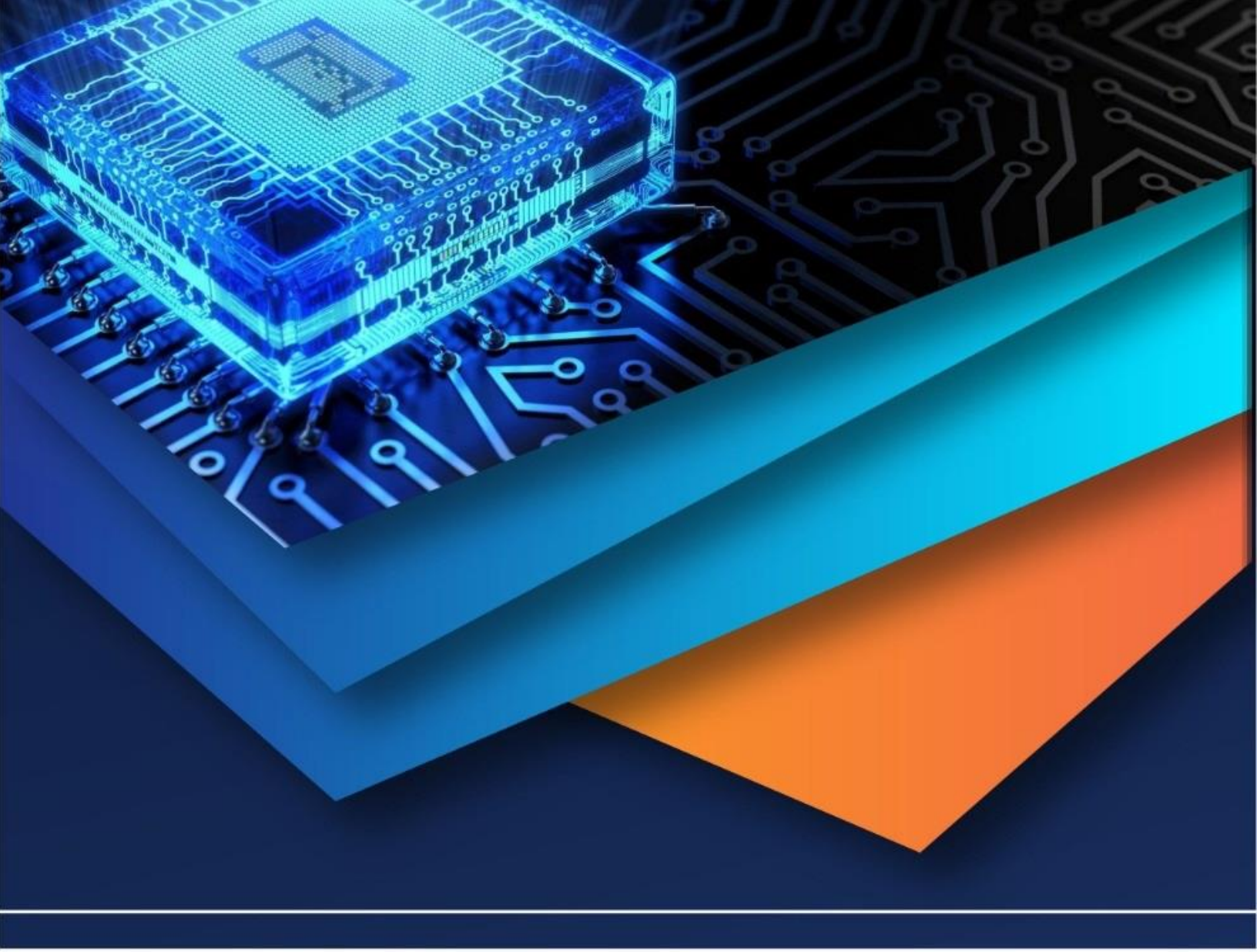
The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely fortunate to have got this all along the completion. Whatever we have done is due to such guidance and assistance. We would not forget to thank them.

I thank Mr. Pradeep Kumar for guiding us and providing all the support in completing this project. We thank the person who has our utmost gratitude is Dr. Krishna Reddy, Head of ECE Department.

I am thankful to and fortunate enough to get constant encouragement, support and guidance from all the staff members of ECE Department.

REFERENCES

[1] Morris, M., Lundell, J., Dishman, E., Needham, B.: New Perspectives on Ubiquitous Computing from Ethnographic Study of Elders with Cognitive Decline. In: Proc. Ubicomp (2003).
[2] Lawton, M. P.: Aging and Performance of Home Tasks. Human Factors (1990)
[3] Consolvo, S., Roessler, P., Shelton, B., LaMarcha, A., Schilit, B., Bly, S.: Technology for Care Networks of Elders. In: Proc. IEEE Pervasive Computing Mobile and Ubiquitous Systems: Successful Aging (2004).
[4] http://www.isuppli.com/MEMSandSensors/News/Pages/Motion-Sensor-Market-forSmartphones-and-Tablets-Set-to-Double-by-2015.aspx.
[5] http://www.comscore.com/Press_Events/Press_Releases /2011/11/comScore_Reports_September_2011_U.S._M obile_Subscriber_Market_Share.
[6] S.W Lee and K. Mase. Activity and location recognition using wearable sensors. IEEE Pervasive Computing, 1(3):24–32, 2002.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY