International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VI **Month of publication:** June 2017

DOI:

www.ijraset.com

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Human Activity Recognition Using Gaussian Mixtures

Mahantesh K[1], Namita H[2], Harshitha R[3], Madhurya S[4], Nischitha K S[5]
[1,2,3,4,5] Department of ECE, SJBIT, Bengaluru, India

Abstract: ***Objective of recognizing human activity in the growing ubiquity and mobility of video data has emerged as one of the key research topic in computer vision and machine learning. This paper considers the inclusivity of latent variables by extracting missing information with discriminative features in reduced feature space. Multi-modal distribution analysis is performed using Gaussian mixture model in maximum likelihood basis to obtain Gaussian mixtures, PCA is applied to these mixtures to preserve suitable discriminatory features in compressed space. Different distance measure techniques are used to classify and recognize human activity. Proposed subspace mixture model achieved most promising results on KTH dataset in comparison with few state-of-the-art techniques.***

Keywords: Human activity, Latent Variable, Gaussian Mixture, Subspace, PCA, Classification, Distance Measures, KTH Dataset.

I. INTRODUCTION

Recent advances in image sensors and acquisition devices have paved the way to the collection of large video datasets. Due to increase in its potential applications, recognizing human actions in videos has becoming more vital. The difficulty of examining patterns responsible for defining human activity in video data is still empirical and requires lot of efforts in generalizing recognition techniques. Spatio-temporal interest point (STIP) can be classified based on local and global features. Static features such as color, posture, texture etc. is captured using local descriptors. Global descriptors extract dynamic features such as changes in scale, illumination, speed and phase. Spatio-temporal interest point (STIP) detector exploits both static and dynamic information [1]. Discriminative max-margin criteria were used to learn key-pose parameters and embedded in dynamic algorithm to identify spatial relationship between humans involved in interactions [2]. Kinematics and natural science was applied to detect human actions to obtain semantic features and create Bag of Word (BoW) using k-means clustering [3].

Different cameras are used to capture available views; 2-D descriptors are applied to each frame of different views and projected on to the lower dimension feature space to extract posterior probability of different actions. Features obtained are then fused together during classification stage to make final decision about the performed actions [4]. Medical videos were effectively interpreted using HMM in a patient monitoring system of cardiology section [5]. Detailed geometric orientations can be found by computing edge gradients on spatial distribution and R-transform reveals angular and kinematic features, actions are classified based on histogram distance [6]. Though Bag of Words (BoW) representation seems to be successful, fails to define spatial and temporal associations amid visual words. t-BoW proposed captures only temporal associations amid pair of words by counting manifestations at several temporal transformations [7].

Cues extracted from quantization parameters and combined with motion estimators on H.264 compressed video given to SVM classifier is invariant to illumination, scale and appearance of human activity recognition in both outdoor and indoor conditions [8]. Environmental problems and intrinsic noise of spatio-temporal features introduces class imbalance, energy based least square twin SVM introduced in [9] uses two non-parallel hyper planes instead of conventional single hyper-plane to generate different energy levels for different classes and reduce the impact of noise in classification. Three levels of clustering is used on k-means to reduce the required time and memory which increase HAR accuracy on KTH dataset [9].To obtain the outcome trace transform on HAR two feature extraction methods were proposed. In first method trace transforms were extracted from binarized silhouettes and in second method trace transforms were extracted to construct a set of constant features which represent the action sequence [10].

Structural and motion information are discriminative features in different activities. A set of templates were constructed to measure the motion in each activity. This template set obtained the required motion and structural information [11]. Unlike standard data sets like KTH, real time videos comprises of continuous actions that may gradually undergo transitions. Fuzzy segmentation and recognition algorithm was applied to obtain events of the video by segmenting the video and to recognize the human activities in the respective videos [12].

Summary of the basic steps performed in HAR in various approaches is elaborated. A detailed information about actions, requisite

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

of action recognition and possible initial steps to be considered is discussed [13]. Surveillance systems focuses on recognizing abnormal activities, graph kernel support vector machine and graph formulation are used to recognize the abnormal activities. These methods gave better performance percentage compared to the performance of algorithm in [3] using BoW [14].

Dynamic Baysein Network is used for recognizing hand gestures. This method can be utilized in applications like sign language recognition. This method comprises a generic framework for modelling and interfering in complex pattern recognition problems [15]. HAR based on silhouette based method is briefed. The discriminant Spatio-temporal subspace is constructed and unknown human actions are recognised using K–NN classifier. The method is efficient for recognizing various human activities in real time videos [16]. Boosting Eigen Actions algorithm is used for evaluation considering KTH and Weizmann dataset. Need for human tracking and prerequisite knowledge about the sequence of actions is not required [17].

Feature extraction techniques using hull convexity defects is introduced. Classification techniques make use of PCA and minimally encoded neural networks. Comparatively computational complexity is less for this algorithm[18].

II. PROPOSED METHODOLOGY

The methodology is divided into two phases namely low level and high level phase. The low level phase includes Background Subtraction and Human detection. The activity analysis of the human is done in the high level phase. These procedures are carried out in the Training and the Testing phase. The block diagram of the proposed system is shown in Fig.1.

In the training phase, the computer is trained to detect the six actions considered namely boxing, hand clapping, hand waving, running, jogging and walking by extracting the features using the above mentioned methods. In the test phase, it is checked if any of these six actions are being detected by using classification methods. The features are extracted using the following methods.

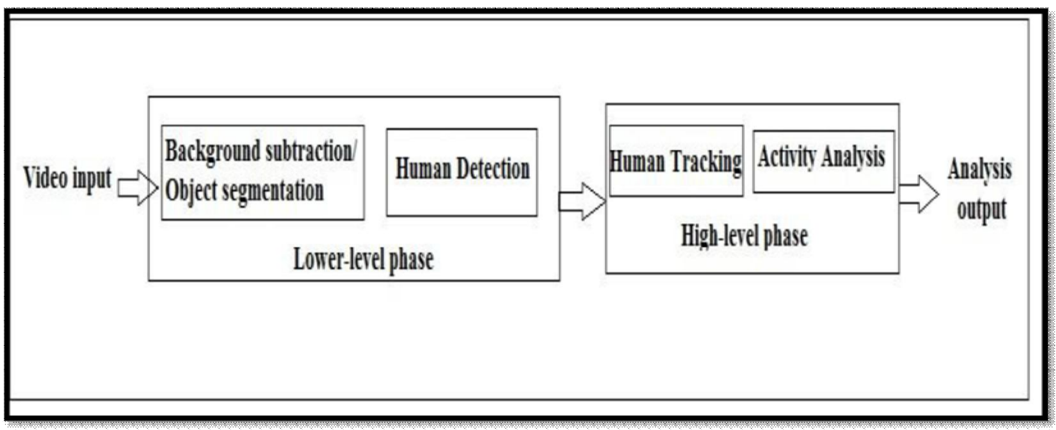


Fig.1. Block diagram of the proposed system

A. Gaussian Mixtures

The frames are initially segmented to obtain the region of interest after which Gaussian mixtures are obtained. The number of Gaussian mixtures to be obtained are decided by obtaining the histogram of the image. The method used to obtain the parameters of the mixtures is Extension Maximization Algorithm (EMA). This helps us to find the hidden information.

EMA is an iterative technique which is operated locally. It consists of two steps namely Expectation and Maximization step.

In the Expectation step, the values of the latent variables are found. In the Maximization step, the parameters are updated using the latent variable calculated to obtain distinct mixtures. Fig.2(a) and Fig.2(b) represent the mixtures at 1st and 2nd iteration (L).

III. SUBSPACE MIXTURE MODEL

The number of features obtained after applying. EMA are high for the computer to handle. Hence, one of the subspace mixture models Principal Component Analysis (PCA) is used. The main purpose of PCA is to reduce the dimension of the data with minimal loss of information. It projects a feature space on to a smaller subspace that represents the data clearly. In PCA, firstly the data is centralized and the covariance matrix is obtained. The eigen vectors of this matrix are found out and the eigen vectors corresponding to the maximum eigen values are considered. These vectors are used to project the data.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

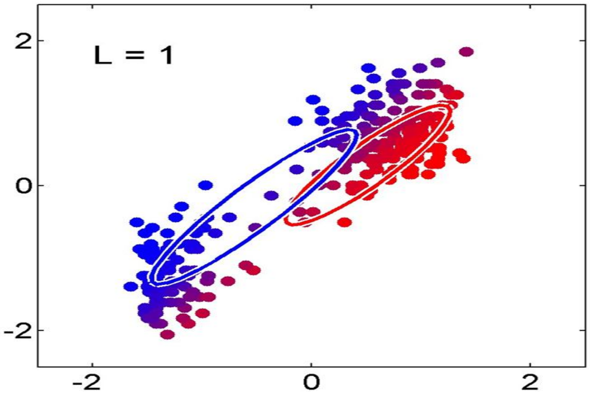


Fig.2(a). Mixtures for L=1

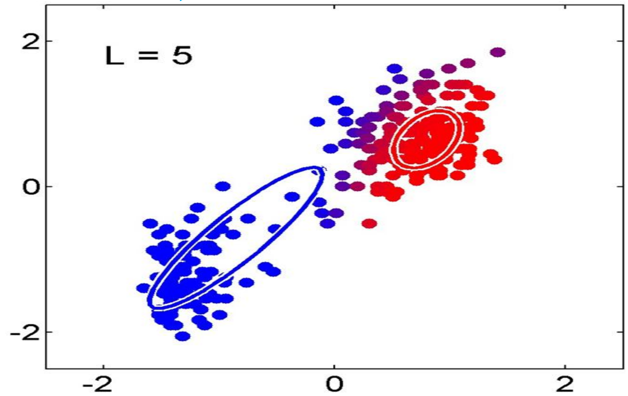


Fig.2(b). Mixtures for L=2

IV. CLASSIFICATION

Classification is used to map the input video into a particular category. The different techniques that can be used for this are Distance Measures, Neural Network, Support Vector Machine (SVM) etc. The distance measures can be done using Eucledian, Manhattan, Modified Eucledian, Angular Based Distance etc. We have used Manhattan Distance Measure for Classification.
The Manhattan Distance is calculated using the formula

$$Sim\ (i) = sim\ (i) + abs(F21(j,1)-D(j,i))$$

V. EXPERIMENTAL RESULTS & DISCUSSIONS

The experiment is done by choosing four videos from each class for training and the remaining videos for testing. The frames considered per video are 50. So, the total frames for training are 1200 (4x6x50). The total frames for testing are 50 (50x1). The number of mixtures considered are four. The proposed method is carried out in two phases namely training and testing phase.

A. Datasets

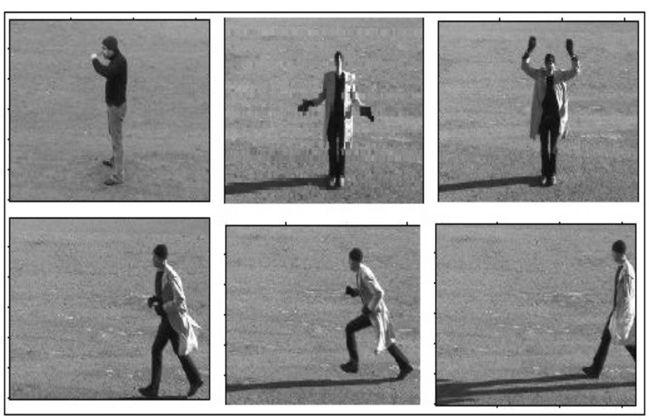


Fig.3. KTH Dataset

B. KTH Dataset

This includes six classes namely boxing, hand clapping, hand waving, jogging, running and walking. Each class has 100 videos performed at 25fps (frames per second) by 25 people in 4 different scenarios namely indoor, outdoor, variant scales and different outfits. Each video has a resolution of 160x120. The evaluation metric is Multiclass Recognition Accuracy. Fig.3 shows the six actions of KTH dataset.

VI. EXPERIMENTAL PROCEDURES

A. Training Phase

1) First videos of all classes are selected as training samples.
2) Read the videos and convert it to frames.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

3) Features are obtained using GMM and EM algorithm.
4) Apply PCA.
5) Store the Feature Matrix in Feature Vector Database.

B. Testing Phase

1) Select a Test Video.
2) Read the video and convert it to frames.
3) Features are obtained using PCA.
4) Compare the features of the trained and the test videos using Manhattan Distance Measure.
5) Label the class obtained.

VII. RESULTS & ANALYSIS

We carried out all our experiments on PC with Intel(R) Core(TM) i3-4030U CPU, 1.90GHz, and 4GB RAM using MATLAB R2013a version.

For every query video of KTH database tested, Table I and Table II show the number of detected frames using Subspace Mixture Model for Manhattan distance measure with number of mixtures k=4.

TABLE I
TOTAL NUMBER OF DETECTED FRAMES FOR K=2

ACTIONS	TOTAL NUMBER OF FRAMES DETECTED
BOXING	50
HAND CLAPPING	3
HAND WAVING	50
JOGGING	16
RUNNING	43
WALKING	25

TABLE II
TOTAL NUMBER OF FEAMES DETECTED FOR K=4

ACTIONS	TOTAL NUMBER OF FRAMES DETECTED
BOXING	56
HAND CLAPPING	19
HAND WAVING	30
JOGGING	51
RUNNING	19
WALKING	54

VIII. CONCLUSION & FUTURE SCOPE

Human behaviour analysis in computer vision is an emerging technology. Interpreting Human behaviour is a challenging task as the behaviour changes with time. In this paper, we have combined a subspace technique (PCA) along with the mixture model in order to recognize human activities in video surveillance system. Since EM algorithm is used, the latent information can also be obtained leading to better results. However, analysis of a human behaviour becomes difficult in a cluttered scenario. In future, algorithms which give better results even in such scenarios must be used.

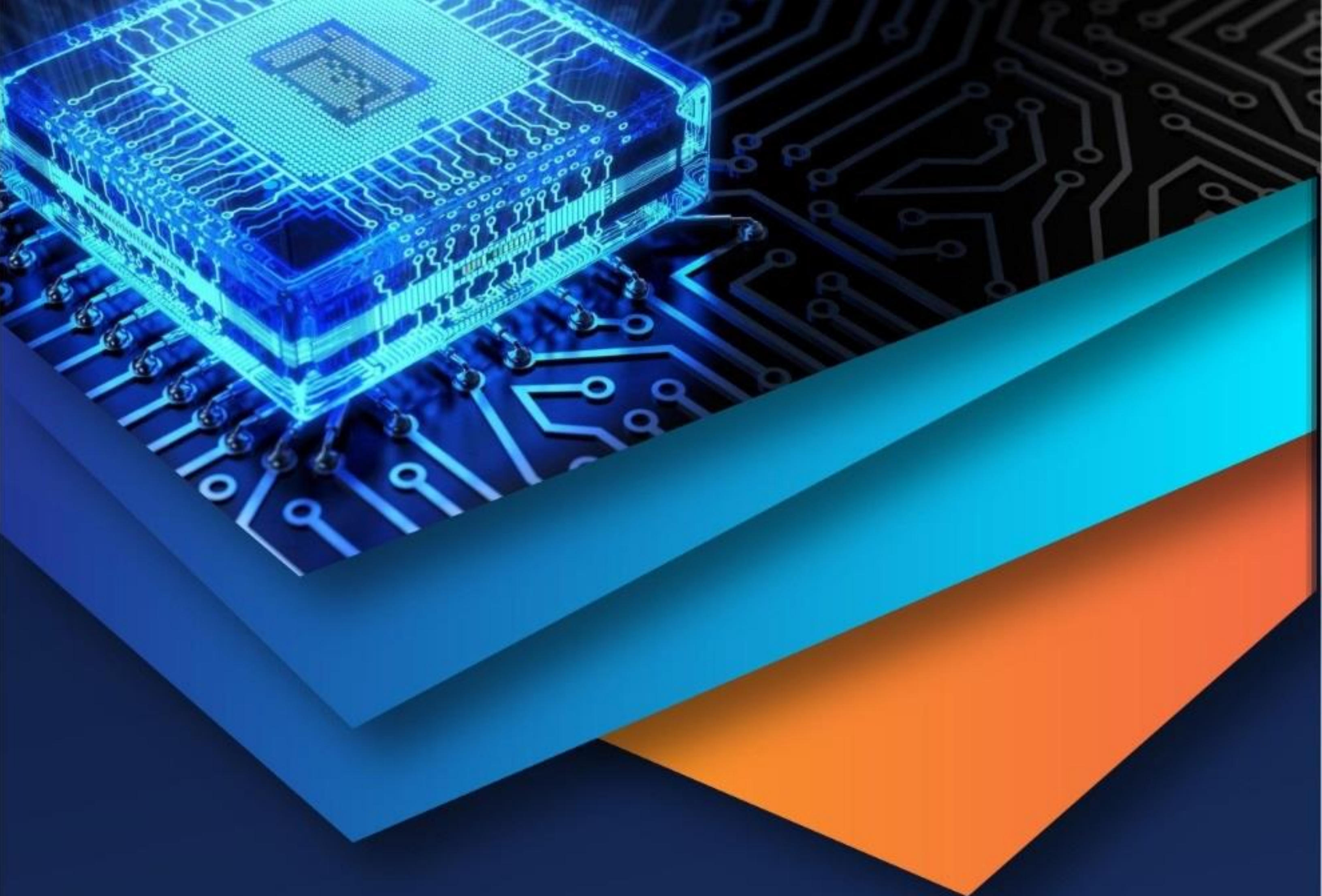
Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

IX. ACKNOWLEDGMENT

This work was partially funded by KARNATAKA STATE COUNCIL FOR SCIENCE AND TECHNOLOGY [Ref No.: 40S_BE_1774], we whole heartedly thank secretary & committee of KSCST – SPP (40th series) for their kind approval of grants in support to our project work.

REFERENCES

[1] Debapratim Das Dawn, Soharab Hossain Shaikh, A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector, Journal Vis Comput, vol. 32, pp. 289–306, 2016.
[2] Arash Vahdat, Bo Gao, Mani Ranjbar, and Greg Mori, A Discriminative Key Pose Sequence Model for Recognizing Human Interactions, IEEE International Conference on Computer Vision Workshops, pp. 1729-1736, 2011.
[3] Yongxiong Wang, Yubo Shi, Guoliang Wei, A novel local feature descriptor based on energy information for human activity recognition, Neurocomputing, vol. 228, pp. 19–28, 2017.
[4] Rodrigo Cilla, MiguelA.Patricio, AntonioBerlanga, Jose M. Molina, A probabilistic, discriminative and distributed system for the recognition of human actions from multiple views, Neurocomputing, vol. 75, pp. 78–87, 2012.
[5] A. S. R. M. Ahouandjino, C. Motamed, and E. C. Ezin, A Temporal Belief_Based Hidden Markov Model for Human Action Recognition in Medical Videos, Pattern Recognition and Image Analysis, vol. 25, no. 3, pp. 389–401, 2015.
[6] Dinesh K. Vishwakarma, Rajiv Kapoor, Ashish Dhiman, A proposed unified framework for the recognition of human activity by exploiting the characteristics of action dynamics, Robotics and Autonomous Systems, vol. 77, pp. 25–38, 2016.
[7] Pau Agusti, V. Javier Traver, Filiberto Pla, Bag-of-words with aggregated temporal pair-wise word co-occurrence for human action recognition, Pattern Recognition Letters, vol. 49, pp. 224–230, 2014.
[8] Manu Tom, R. Venkatesh Babu & R Gnana Praveen, Compressed domain human action recognition in H.264/ AVC video streams Multimedia Tools Appl, vol. 74, pp. 9323-9338, 2015.
[9] Jalal A. Nasiri, Nasrollah Moghadam Charkari, Kourosh Mozafari, Energy-based model of least squares twin Support Vector Machines for human action recognition, Journal of Signal Processing, vol. 104, pp. 248-257, 2014.
[10] Motasem Elshourbgy, Elasyed Hemayed, Magda Fayek, Enhanced bag of words using multilevel k-means for human activity recognition, Egyptian Informatics Journal (2016) 17, 227-237.
[11] Georgios Godelis, Konstantinos Karpouzis, Stefanos Kollias, Exploring trace transform for robust human action recognition, Pattern Recognition 46 (2013) 3238-3248.
[12] Jinhui Hu, Nikolaos V. Boulgouris, Fast human activity recognition based on structure and motion, Pattern Recognition Letters 32 (2011) 1814-1821.
[13] Hao Zhang, Wenjun Zhou and Lyne E.Parker, Fuzzy Segmentation and Recognition of Continuos Human Activities, 2014 IEEE International Conference on Robotics and Automation (ICRA).
[14] Samitha Herath, Mehrtash Harandi, Fatih Porikli, Going deeper into action recognition: A survey, Image and Visin Computing 60 (2017) 4-21.
[15] Dinesh Singh, C.Krishna Mohan, Graaph formulaton of video activities for abnormal activity recognition, Pattern Recognition 65(2017) 256-272.
[16] Heung-II Suk, Bong-Kee Sin, Seong-Whan Lee, Hand gesture recognition based on dynamic Bayesian network framework, Pattern Recognition 43 (2010) 3059-3072.
[17] Chien-Chung Tseng, Ju-Chin, Ching-Hsien Fang, Jenn-Jier James Lien, Human action recognition based on graph-embedded spatio-temporal subspace, Pattern Recognition 45(2012) 3611-3624.
[18] Chang Liu, Pong C.Yuen, Human action recognition using boosted Eigen Actions, Image and Vision Computing 28 (2010) 825-835.
[19] M. M. Youssef, V. K. Asari,, Human action recognition using hull convexity defect features with multi-modality setups, Pattern Recognition Letters 34 (2013) 1971-1979.
[20] Behrouz Saghafi, Deepu Rajan,, Human action recognition using Pose-based discriminant embedding, Signal Processing: Image Communication 27 (2012) 96-111.
[21] Yang Yi, Yikun Lin, Human action recognition with salient trajectories, Signal Processing 93 (2013) 2932-2941.
[22] Swarup, Kumar Dhar, Md. Mahmudul Hassan, Shayhan Ameen Chowdhury, 2nd International Conference on Ekectrical, Computer & Telecommunication Engineering (ICECTE) 8-10 December 2016.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)