Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 5, May 2019, pg.109 – 118

Recognition of Human Facial Expression using Machine Learning Algorithm

Lamia Zuha A[1]; Anand M J[2]
[1]Department of ECE, PESCE, Mandya, India
²Assistant Professor, Department of ECE, PESCE, Mandya, India
[1] ylamiya23@gmail.com; [2] anamysore@gmail.com

***Abstract**— In the vision of the computer, Human Action Recognition (HAR) plays a huge role in the research area. The human facial expressions convey a lot of information visually rather than articulately. Hence, to identify these facial expression by computer with high recognition rate was a challenging task. To overcome these problem, this paper presents a new technique for human face recognition using the real time videos. The aim of this paper is to recognize the Human Facial Expression like joy, sadness, surprise, fear, anger and disgust from the input video having homogeneous background. The extracted features are given to the pre-processing stage to remove noise and the required features are obtained from the Pre-loaded video sequence. For this gage of work we are using Viola-Jones Algorithm to recognize face, matching algorithms is used for training data, Extreme Sparse Learning and Kernel Extreme Sparse Learning algorithm to recognize the expression of human face. The main application of this works are information security, authentication, biometric identification, video surveillance, data privacy, Human Action Recognition (HAR), Human Computer Interface (HCI), Health Care etc. It also presents the application of the machine learning algorithm to recognize and classification of facial expressions in real time video. Here the pre-loaded video sequence of the human facial expression is given as the input to the MATLAB software to obtain the ideal output. This paper can be implemented using the High Configuration MATLAB simulation software.*

***Keywords**— Human Facial Expression, HAR, HCI, Viola - Jones Algorithm, Matching Algorithm, ESL and KESL.*

I. INTRODUCTION

As the technology is growing rapidly, everything is getting automated. Human beings naturally use facial expression as an important and powerful modality to communicate their emotions and to interact socially [1]. There is a huge need to recognize Human Facial Expression as it has wide scope of applications such as information security, authentication, video surveillance, biometric identification, data privacy, Human Behaviour Interpretation (HBI), Human Computer Interface (HCI), and Health Care etc. In Human Machine Interaction (HMI) a natural Way of communication is introduced by Facial Expression Recognition (FER). Emotions are the positive or negative state of a person's mind which is related with a pattern of any physiological activities [1]. Mental state of the person is described by Emotions. Basics emotions are categorized into six classes: JOY, SADNESS, SURPRISE, FEAR, ANGER and DISGUST. These are the basic six emotions that blend to form complex emotions. The Viola-Jones algorithm is real-time processing system for face detection. It is designed by applying AdaBoost algorithms for rectangular features classifier in "cascade" stages [2]. The primitive concept allows background image are discarded quickly and their face detection proceeded in 15 frames per second (fps) [1]. The main objective of this project is to recognition/analyse the problems existed in HAR and to develop a robust human facial expression recognition algorithm which helps to recognize many actions such as: joy, sadness, surprise, fear, anger and disgust etc., from the input video sequences. It also check the performance in the fusion of activities which includes emotions of person are mapping and testing for HAR System [3].

II. PROPOSED WORK

The proposed work is as shown in Figure.1 and it consists of any input video sequence that is been considered for the detection of the expression is first split into frames. These frames are converted into images. The converted images are pre-processed in order to remove unwanted noise. The feature that are needed for the processing is been extracted by the Viola-Jones Algorithm [4]. The desired expressions are identified and the output is obtained. The proposed block diagram of Human Activity Recognition (HAR) system is as shown below in Figure 2. In order to have a robust recognition, the system should be invariant to scale, homogeneous background. There are two sets of data 1) Training and 2) Testing. The training datasets are used to train the complete system with relevant data of the actions present in the current video sequences and the testing dataset is used for performance evaluation of the system.

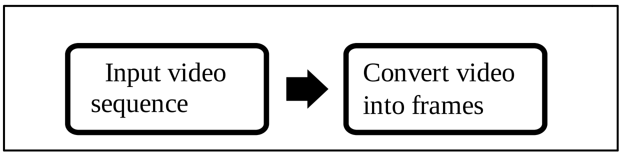


Figure 1: A System for Image Extraction

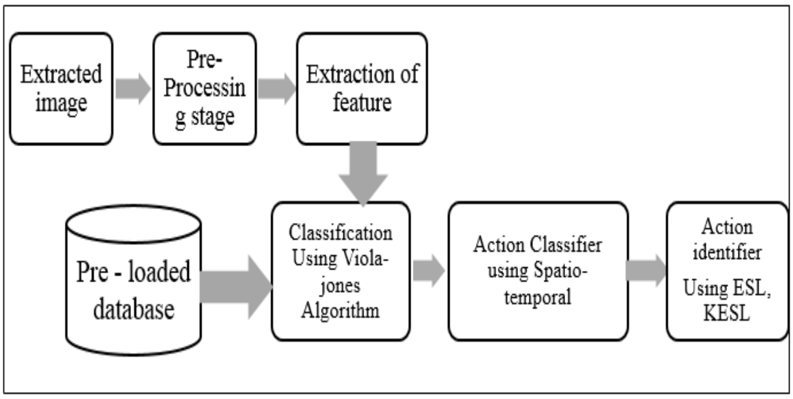

Extracted image
Pre-Processin g stage
Extraction of feature
Pre - loaded database
Classification Using Viola-jones Algorithm
Action Classifier using Spatio-temporal
Action identifier Using ESL, KESL

Figure 2: General Human Activity recognition System

The details of the each block of the proposed system from Figure.2 is given below.

Image Extraction Model:

The video sequence is used as the input to the developed system. The video should be analysed frame by frame to obtain the appearance of the object details such as shape and motion characteristics of the video. Hence, the video sequence is converted into frames and in turn into images [4].

Pre-processing:

In general, due to the presence of unwanted noise the input video, which is affected while capturing, lighting variations, changes in background clutters like, trees and so on, due to which the required details cannot to be extracted satisfactorily [5]. Hence, before feature extraction process there is a need of pre-processing the input image to remove unwanted noise in order to preserve the required details.

Feature Extraction:

In order to perform the recognition task in HAR system there is a necessary for the details which characterize the motions of the object. There are many types of features which can be extracted to characterize the details of motions of human action [5]. These features which represent the characteristics of the motions are necessary to train and test the algorithm which is used to recognize the different types of actions of human body.

III. DESIGN AND IMPLEMENTATION

a) Viola – Jones Algorithm

b) Spatio- Temporal Algorithm

c) Active Appearance Model

d) Extreme Sparse Learning Algorithm

e) Kernel Extreme Sparse Learning Algorithm

a) VIOLA JONES ALGORITHM:

To identify whether it is face or not we are using viola jones algorithm. It consist of four step

- Harr feature extraction.
- Integral image
- Adaboost method
- Cascade classifier

Harr feature extraction:

All human faces have some similar property, to match these similarity we are using Harr feature. For ex: Eye region is darkening than the upper cheeks, Nose bridge region is brighter than the eyes. Each feature is correlate to a particular position in the sub- window [6].

We having three types of Rectangular feature i.e. Two, Three, Four rectangular feature. Viola jones uses two rectangular features.

Value=Σ (pixel in black area) -Σ (pixel in white area).

Integral image:

Image representation is known as integral image, it finds the value of rectangular feature in constant time. Integral image at location (X, Y), is the sum of the pixels above and to the left of (X, Y).

Adaboost method:

It is used to select the best feature and to train the classifiers. It built a "strong" classifier as a linear combination of weighted simple "weak" classifier.

H(x) = sign [$\boldsymbol{j}$ ($\boldsymbol{x}$)] $\boldsymbol{Mj}$ =$\mathbf{1}$

Each weak classifier is a threshold function based on the feature f j.

Hj (x) = $\boldsymbol{-sj\ if\ fj}$ < $\boldsymbol{\theta j}$
$\boldsymbol{sj}$; $\boldsymbol{otherwise}$

The threshold value θj and the polarity sj ∈ ±1 are determined in the training as well as the coefficient αj.

Cascade classifier:

The first classifier in the cascade known as attention operator. It uses two features to attain a false negative rate of about 0% and a false positive rate of 40%. In cascading, all phase consists of a strong classifier [6]. So all the features are grouped into some phases where each phase has certain number of feature. The task of each phase is to find whether a given sub-window is face or not. A given sub-window is straightway discarded as not a face if it loss in any of the phases.

The false positive rate for an entire cascade is given by

F= $\boldsymbol{fi.\ ki}$ =$\mathbf{1}$

Detection rate is given by

D= $\boldsymbol{di\ ki}$ =$\mathbf{1}$

b) SPATIAL TEMPORAL ALGORITHM:

The variation of space with respect to time is given by Spatio Temporal Algorithm. The extension of spatial domain is Spatio Temporal, which includes geometry over time and location of object moving over invariant geometry [7].

In this paper we have considered the Spatio-Temporal Algorithm that is based on the pose invariant Optical Flow that is capable of expressing the expression even when there is a head movement. There are two proposed descriptors which uses the histogram to extract the features from the Optical Flow.

The encoded motion information of facial components is proposed for the pose-invariant features

- **The divergence of the flow field**

This is used to measure the amount of facial muscle expansion or contraction.

- **The Optical Flow perpendicular to the local spin around**

This referred as Curl which is useful to measure the dynamics of the local circular motion of the facial components.

c) ACTIVE APPEARANCE MODEL (AAM):

AN AAM is a computer vision algorithm in order to match a statistical model of an object shape and appearance to a new image. They are built during a training phase. A set of images, together with coordinates of landmarks that appear in all of the images, is provided to the training supervisor [7]. In AAM we learn the relationship between error and parameter adjustments. The use of multiple multivariate linear regressions generates training set by moving object parameters for training images, which includes small displacements in position, scale, and orientation. Image difference in the moving object is recorded. It is important to consider the reference frame when computing image difference which is used to shape-normalized representation (warping).

d) EXTREME SPARSE LEARNING:

Sparse is a powerful tool for high-dimensional data compression, representation and extraction. The sparse representation has the ability to unveil the important information from dictionary atoms of the signals which makes it a successful technique to represent noisy signals [8]. Extreme Sparse Learning is used for the classification of the expression. By Adding the Extreme Learning Machine (ELM) error into the objective function of the conventional sparse representation to learn a dictionary that is both reconstructive and discriminative defines ESL. The objective functions that are combined here has both linear and non-linear terms which is solved by an approach called **Class Specific Matching Pursuit (CSMP).**

e) KERNEL EXTREME SPARSE LEARNING:

The kernel extension of the Extreme Sparse Learning framework is Kernel Extreme Sparse Learning [8]. It increases the robustness of extreme sparse learning machine (ELM) by turning linearly the non-separable data in a low dimensional space into a linearly separable one. However, the internal power parameters of ESL are initialized at random, causing the algorithm to be unstable. We use the active operator’s particle to obtain an optimal set of initial parameters for KESL, thus creating an optimal KESL classifier. Moreover, KESL algorithm has good stability and convergence, and is shown to be a reliable and effective classification algorithm.

IV. RESULT AND DISCUSSIONS

The expression consider for processing was joy and the result obtained is the same. The input image is converted from RGB to gray scale and the features are extracted from the Gray Scale image to recognize the input expression of the video sequence [8]. Following are the results obtained.

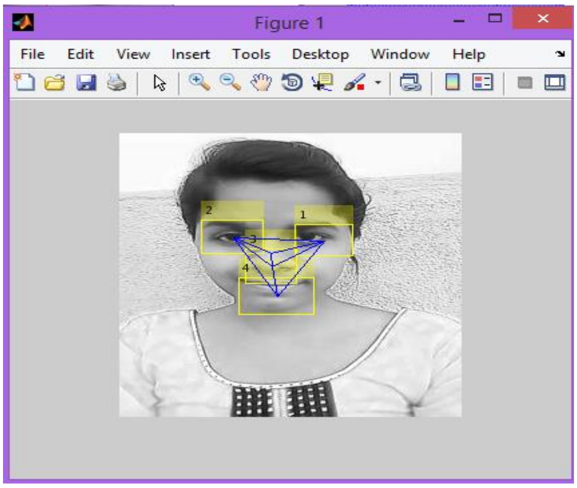


Figure 3: Extraction of features

- The Figure 3 shows the extraction of feature by Viola Jones algorithm. With the help of this algorithm the extracted features will be represented in rectangular box. Here nose is consider as the centre point. Each feature is correlated to a particular position in the sub- window.

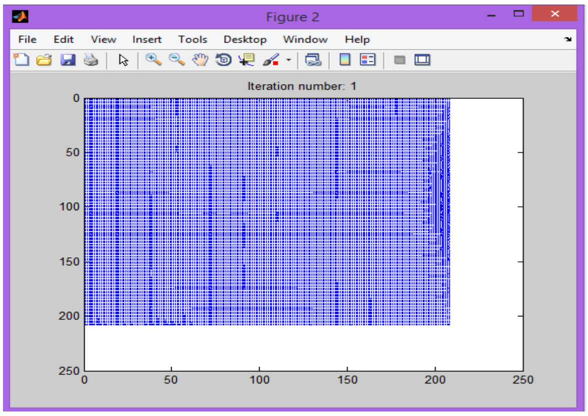


Figure 4: Vector for Extracted features

- Figure 4 shows the pixel to pixel representation in the graph. The extracted features are represented in vector by using support vector machine.

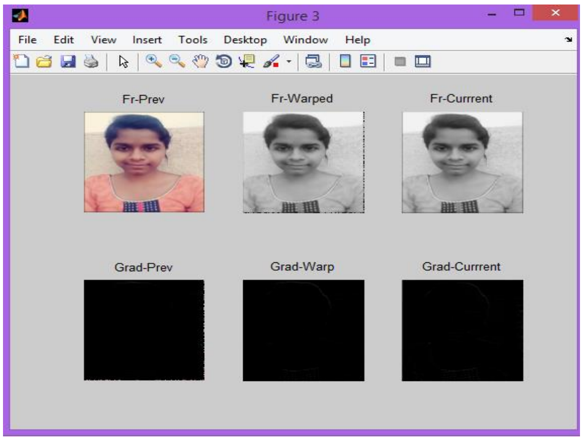


Figure 5: Preview, Warped & Current frames of the Feature extracted frames

- From the Figure 5, the subtracted frame from the averaged frames are indicated in the warped and current frames. The image is reduced from maximum to minimum and the input image is converted from RGB to gray scale and the features are extracted from the Gray Scale image to recognize the input expression.

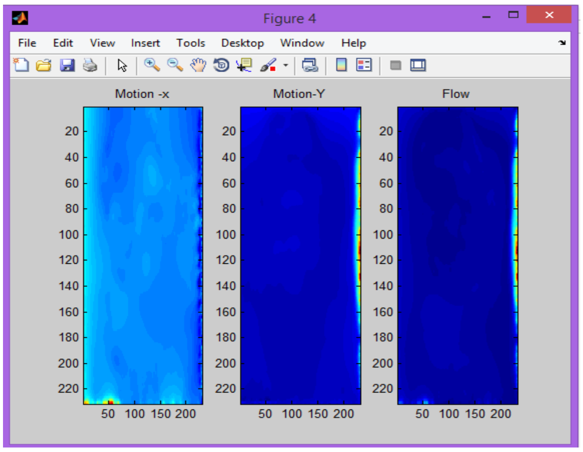


Figure 6: Motion with respect to X & Y axis

- Figure 6 represents the Variation of motion to X and Y axis and flow with respect to time for the extracted frame. Motion-X represents the amount of facial expansion and contraction, Motion-Y represents the face neural and flow represents the background clutters.

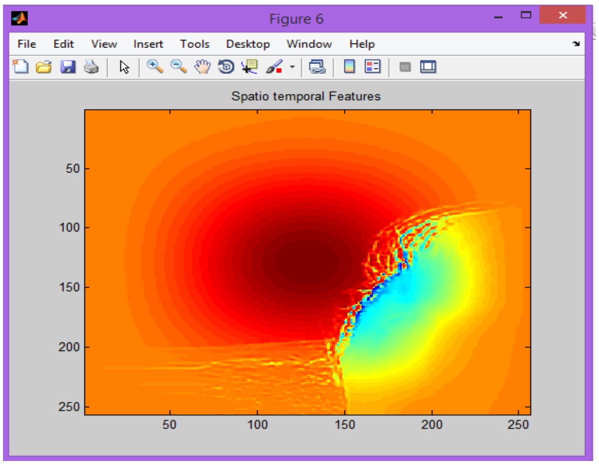


Figure 7: Spatio temporal features of the best feature extracted frame

- The Figure 7 gives the Spatio temporal features of the best feature extracted frame. Most contraction and expansion muscle variation is depicted in various color by using Spatio temporal feature.

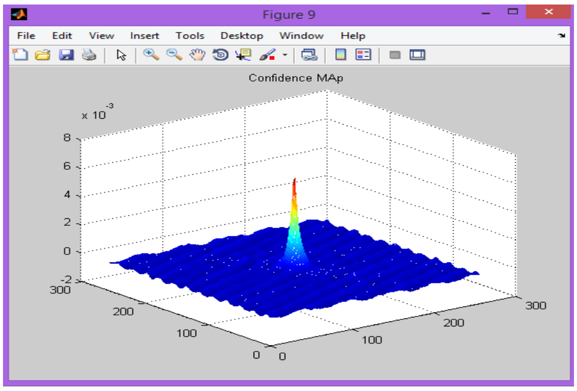


Figure 8: Confidence Map

- Figure 8 represents the Confidence map in which the Extracted features from the Spatio temporal are indicated in 3D view. The best feature obtained is represented at its peak with respect to the neutral and background clutters on the base.

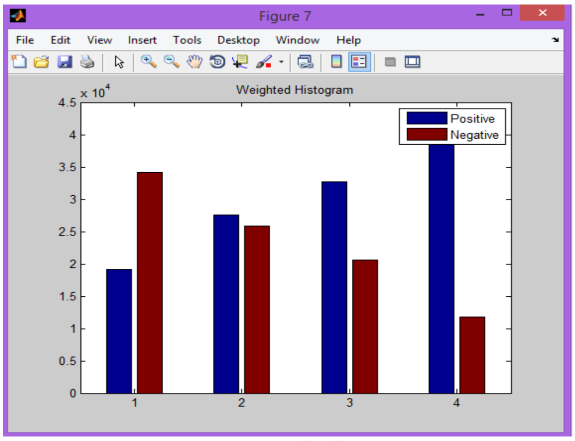


Figure 9: Weighted Histogram

- Figure 9 shows the Weighted Histogram which contains only magnitude with positive and negative bins. Here above the face contraction and expansion is consider as positive and below the face contraction and expansion is consider as negative.

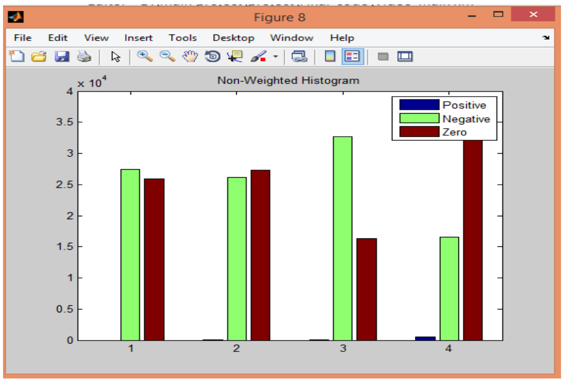


Figure 10: Non-Weighted histogram

- Figure 10 shows the Non-Weighted Histogram which contains only sign of features with positive, negative and zero bins. Here positive and negative are indicated same as weighted histogram and zero is consider as the movement of the mouth.

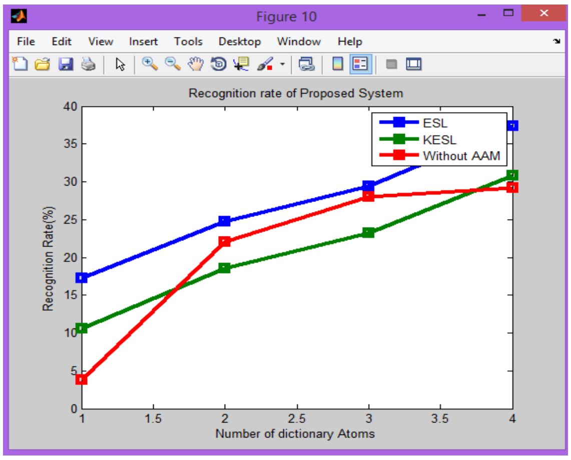


Figure 10
File Edit View Insert Tools Desktop Window Help
Recognition rate of Proposed System
Recognition Rate(%)
40
35
30
25
20
15
10
5
0
1
1.5
2
2.5
3
3.5
4
Number of dictionary Atoms
ESL
KESL
Without AAM

Figure 11: Recognition rate of proposed system

- Figure 11 represents the graph of Recognition rate of the proposed system. It is obtained with respect to recognition rate in percent and number of dictionary atoms.

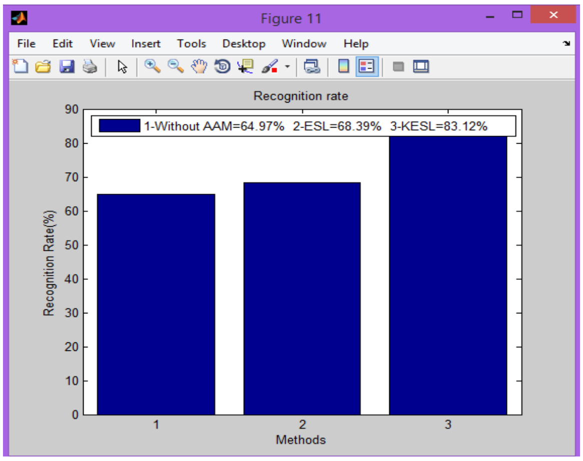
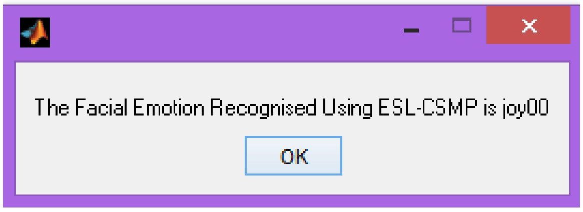


Figure 11
File Edit View Insert Tools Desktop Window Help
Recognition rate
1-Without AAM=64.97% 2-ESL=68.39% 3-KESL=83.12%
Recognition Rate(%)
90
80
70
60
50
40
30
20
10
0
1
2
3
Methods

Figure 12: Recognition rate

- Figure 12 represents Recognition rate. It shows with respect to Recognition rate in percent and different methods like without AAM, ESL and KESL.



The Facial Emotion Recognised Using ESL-CSMP is joy00

Figure 13: Expected Result for the preloaded video (joy expression)

- Figure 13 displays the final result. An Input expression will be Joy and we get the output expression indicated will also be Joy**.**

Table 1: Outputs of different expressions

Expression	**Without AAM**	**ESL**	**KESL**
Joy	62.34%	65.45%	83.12%
Anger	51.89%	61.94%	80.59%
Fear	61.16%	66.69%	91.07%
Disgust	61.19%	66.66%	91.54%

Table 1 shows the ideal outputs of different expressions using various algorithms. The KESL algorithm have high stability and convergence. Therefore the efficiency of KESL algorithm is high when compared to other two algorithms.

V. CONCLUSION

Human Facial Expression place a vital role in inters personal communication containing extremely abundant information of human behaviour. Viola-Jones Algorithm is used to determine whether the given extracted frame is a face or not. Spatio-Temporal Algorithm is used determine the variation of space with respect to the varying time. For the purpose of matching we use Active Appearance Model (AAM), Extreme Sparse Learning (ESL) and Kernel Extreme Sparse Learning (KESL) Algorithms. The main application of this work are information security, authentication, biometric identification, video surveillance, data privacy, Human Action Recognition (HAR), Human Computer Interface (HCI), Health Care etc.

ACKNOWLEDGEMENT

Centre of Excellence in VLSI design, Department of Electronics and Communication Engineering PESCE, Mandya. First I would like to express my gratitude to my Guide Mr. Anand M J and also our HOD Dr. K A Radhakrishna Rao who gave me the opportunity to do this project on the topic Recognition of Human Facial Expression Using Machine Learning algorithm, I would also like to thanks to my parents and friends.

REFERENCES

1) Luca Turchet and Roberto Bresin, "Effects of Interactive Sonification on Emotionally Expressive Walking Styles": IEEE Transactions on affective computing, Vol.6, No.2 April-June-2015.
2) Muhammad Babar Rasheed, Haris Baidar Raja and Turki Ali Alghamdi, "Evolution of Human Activity Recognition and Fall Detection Using Android Phone", 2015 IEEE 29th International Conference on Advanced Information Networking and Applications.
3) R. V. Darekar, A. P. Dhande, "Emotion Detection with Multimodal Fusion Using Speech - A Review" International Journal of Computer Science and Communication Engineering Volume 3 issue 1(February 2014 issue)
4) V. Kulkarni and Savitha S. Raut, "Emotion Recognition by Using Speech and Facial Expressions". Proceedings of 9th IRF International Conference, Pune, India, 18th May. 2014, ISBN: 978-93-84209-20-9. PNCDI No. 339/2007.
5) Lu Xia and J. K. Aggarwal,"Spatio-Temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera", Computer & Vision Research Center / Department of ECE, The University of Texas at Austin in 2013
6) Harish Kumar Dogra , Zohaib Hasan , Ashish Kumar Dogra "Face expression recognition using Scaled-conjugate gradient Back-Propagation algorithm" International Journal of Modern Engineering Research (IJMER) Vol. 3, Issue. 4, Jul - Aug. 2013 pp-1919-1922
7) Michael B. Holte, "Human Pose Estimation and Activity Recognition from Multi-View Videos": Comparative Explorations of Recent Developments, IEEE Journal of Selected Topics In Signal Processing, Vol. 6, No. 5, September 2012
8) M. S. Ryoo J. K., Aggarwal, "Stochastic Representation and Recognition of High-level Group Activities", Robot Research Department, Electronics and Telecommunications Research Institute, Korea, in 2009