Automatic Emotion Detection Model from Facial Expression

[1]Debishree Dagar, [2]Abir Hudait, [3] H. K. Tripathy, [4] M. N. Das
[1,2,4]School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar -751024, India
[3]School of Computing, University Utara Malaysia, Sintok, Kedah, Malaysia
[1]debishreedagar04@gmail.com, [2]abirhudait@gmail.com, [3]hrudayakumar@gmail.com, [4]mndas_prof@kiit.ac.in

***Abstract*—The human face plays a prodigious role for automatic recognition of emotion in the field of identification of human emotion and the interaction between human and computer for some real application like driver state surveillance, personalized learning, health monitoring etc. Most reported facial emotion recognition systems, however, are not fully considered subject-independent dynamic features, so they are not robust enough for real life recognition tasks with subject (human face) variation, head movement and illumination change. In this article we have tried to design an automated framework for emotion detection using facial expression. For human-computer interaction facial expression makes a platform for non-verbal communication. The emotions are effectively changeable happenings that are evoked as a result of impelling force. So in real life application, detection of emotion is very challenging task. Facial expression recognition system requires to overcome the human face having multiple variability such as color, orientation, expression, posture and texture so on. In our framework we have taken frame from live streaming and processed it using Grabor feature extraction and neural network. To detect the emotion facial attributes extraction by principal component analysis is used and a clusterization of different facial expression with respective emotions. Finally to determine facial expressions separately, the processed feature vector is channeled through the already learned pattern classifiers.**

***Index Terms*—Face Detection, Gabor Feature Extraction, Neural Network, Facial Expressions, Emotion Recognition, Facial Attribute Extraction, Principal Component Analysis(PCA), Pattern Classification, K-mean clustering.**

I. INTRODUCTION

The article introduced here has mainly concentrated on the creation of smart framework with the inherent capabilities of drawing the inference for emotion detection from facial expressions. Recently, the notion of emotion recognition is attaining mostly the researcher's mind in the area of exploration on smart system and interaction between human and computer. Based on facial attributes the facial expression recognition can be classified one of the six well known fundamental emotions: sadness, disgust, happiness, fear, anger and surprise [1]. Coren and Russel [1] stated that each emotion is having the property of stereo-scopic-perceptual conflict. So establishing an effective automatic emotion recognition framework is a very challenging task.

Emotion recognition [2][3][5] is useful to make smooth communication between human & computer interaction. The recognition of human emotion can have wide applications in heterogeneous field. The applications are mainly based on the man and machine interaction, patient surveilling, inspecting for antisocial motives etc. Even we can recognize emotion for customers by analyzing their response on seeing certain commodity or advertisement or immediately after getting a message and based on the response from the customers, the resource hub can improve their strategies[1].

The first aim of this work is to incorporate anatomical grip for emotion recognition. Facial behavior is represented using Facial Action Coding System (FACS). FACS couples the transient appearance changes with the action of muscles from anatomical perspective. FACS employs Action Units (AU) and AU represents the muscular activities to describe the facial expressions. Generally, a single muscle is invoked by most of the AUs. However, in some scenarios to express relatively autonomous activity of several segment of one specific muscle, two or more than two AUs [13] are used. FACS has recovered overall 46 Action units which delivers a multifaceted procedure to express a large variety of facial behavior [13].

In rest of the section we have discussed the following-Section II we have tried to report and refer some of the influential work in the domain of emotional intelligence. Section III we have dragged something on emotion taxonomy. Section IV we have described the dataset. Section V discussed mainly methodology, where initially frame extraction from live streaming and well-known face detection through neural network are touched on very light way, then the detail discussion of PCA is taken care and applied K-mean with a little modification for clustering, including a pictorial representation of flow chart and output of each individual step. Section VI the results are shown and experimental analysis is done towards reaching our goal. Finally Section VII and Section VIII extension of our work and conclusion are made betterment.

II. RELATED WORK

In emotional recognition of face a notable advancement has been observed in the field of neuroscience, cognitive and computational intelligence [1][5][6]. It is also proved by Kharat and Dudul that about 55% effect of overall emotion expression is as

ISBN No.978-1-4673-9545-8

facial expression which is contributed during social interactions.

Actually, facial muscle generates monetary adaptation in facial appearance which can be recapitulated by incorporating Action Units. The six common emotions have been considered as globally recognizable as the movements of muscle are very similar of these emotional expressions from the people from various region and society. Therefore, we have mainly concentrated on the automatic recognition of these six fundamental emotions.

In general, emotion recognition is a two steps procedure which involves extraction of significant features and classification [1]. Feature extraction determines a set of independent attributes, which together can portray an expression of facial emotion. For classification in emotion recognition the features are mapped into either of various emotion classes like anger, happy, sad, disgust, surprise, etc [1]. For the effectiveness calculation of a facial expression identification model both the group of feature attributes which have been taken for feature extraction and the classifier that is responsible classification are equivalently significant. For a badly picked collection of feature attributes, in some cases, even a smart classification mechanism is not able to produce an ideal outcome. Thus, for getting the high classification accuracy and qualitative outcome, picking of superior features will play a major role.

The circumflex model by Russell and recognition of six basic emotions are having remarkable contribution in the field of emotion recognition. Other than this, the work by Kudiri M. Krishna, Said Abas Md, Nayan M Yunus [2], where they tried to detect emotion by using the concept of sub-image based features through facial expression. Silva C. DE Liyanage, Miyasato Tsutomu, Nakatsu Ryohei [4], they formed a model for emotion recognition with the help of multimodal information. Maja Pantic, Iounnis Patras [5], they implemented an approach for recognition facial action and temporal segment from fare profile image sequences by considering the dynamic property of facial action. Li Zhang, Ming Jiang, Dewan Farid, M.A. Hassain [13], they modelled an intelligent system for automatic emotion recognition. Happy S L, Routray Aurobinda [7], created an automatic emotion recognition system by using salient features. This article is greatly influenced by all of this contribution in the field of emotion recognition.

III. EMOTION TAXONOMY

According to the emotion theorists and psychologists, different of emotion can be categorized starting from globally showed six fundamental emotions to complicated emotions which are originated from different culture with. From several reported framework in the field of emotion recognition, two models have hold the command in this research domain: Ekman's fundamental set of emotions [1], and Russell's circumflex representation of influence [1]. Ekman and Freisen in 1971 [1] put forward six quintessential basic emotions like disgust, joy, sadness, fear, anger and surprise which are globally presented and identified from facial expressions.

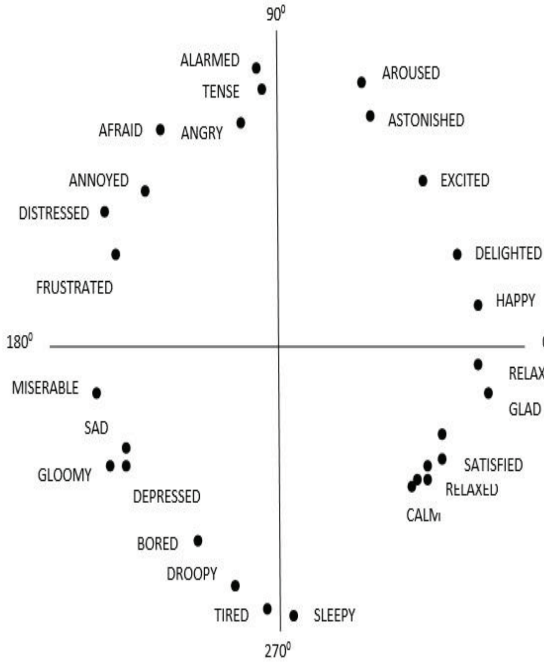


Fig1. The Circumflex Representation of Russell.

Since last five decades, this model with six basic emotion has begun to be the most popular and usual model for estimating the emotions and detection of emotion from their respective facial expression. After certain time a different model of emotion was presented by Russell where emotional states are depicted by a ring having two pole in two dimensional space instead for categorizing each of the emotion distinctly.

IV. DATASET

To simulate our proposed model JAFFE has been used. JAFFE has 213 sample images and 213 lines in this file. Each line contains the position of 77 key points, thus makes a 154 dimensional vector. Also all_labels.txt contains all sample labels in numerical form, where label mapping is done as follows: NEU = 0; HAP = 1; SAD = 2; SUR = 3; ANG = 4; DIS = 5; FEA = 6. Out of 213 samples 20 number of sample is used for training for respective emotion.

V. IMPLEMENTATION

A. *Frame Extraction and Face Detection*

Initially we have taken live video stream and extracts frames from the video. Then we have tried to detect those frames that are having face. To detect face in a frame we have used an existing well-known technique where for feature extraction Gabor method is used and for learning of neural network is used, and combinedly they are used for face detection[14] [15] [16]. We send frame for further processing where the model is already learned for emotion detection.

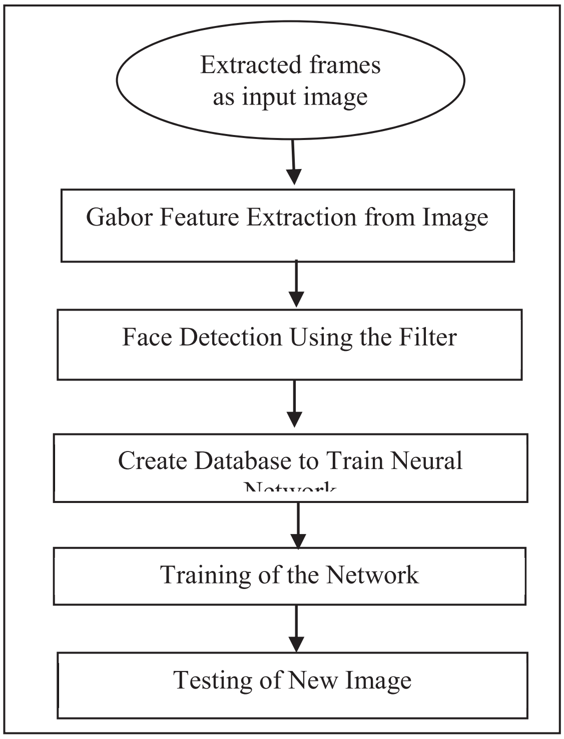


Fig 3. Generalized Model for Face Detection

i)Multi-Layer Perceptron

The feed forward architecture of Multi-Layer Perceptron (MLP) neural network [19][20][21]consists of three layer- input layer, a hidden layer, and an output layer. For an N dimensional input vector there exists N units in the input layer. The input units are fully cascaded to the I hidden layer units, which are in turn, connected to the J output layers units, where J is the number of output classes. A training data of 1 pairs (xi, yi) is assumed to be accessed where xi is the pattern vector, while y_i is the corresponding pattern class. We can code yi as 1 and -1 in a 2-class task.

MLP comprises of 3 layers, the input layer is a vector constituted having n^2 units of neurons (n x n pixel input images). The hidden layer contains n neurons, and the output layer contains a single neuron that is active to 1 when the face is presented and otherwise face is not presented. The activity of j^{th} neuron in the hidden layer can be represented as

$$Sj = \Sigma\, w_{ji} x_i,\ x_i = f(Sj)(1),$$

Where f represents a sigmoid function and W1 represents the set of weights of neuron i, b1 (i) shows the threshold and xi represents an input of the neuron. Similarly the output layer activity is:$Sj = \Sigma\, w_{ji} x_i$

In this network human faces and non-face are presented with 27x18 pixels as the dimension of the retina, the input vector and the hidden layer are having2160 neurons, 100 neurons respectively.

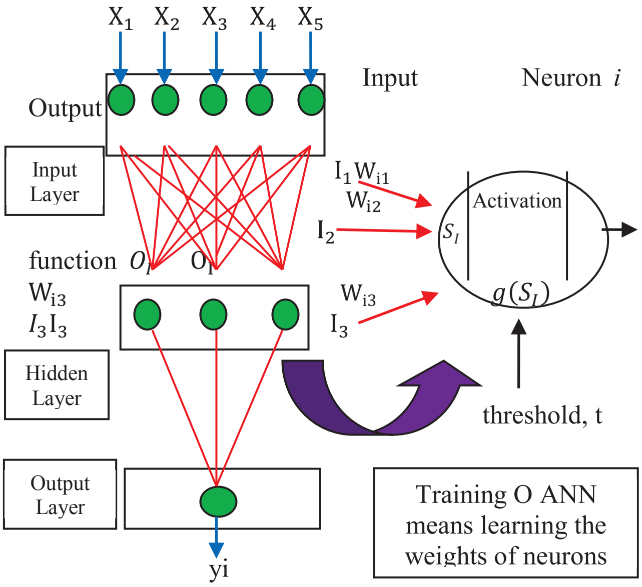


Fig 2. Architecture of Artificial Neural Network

ii).Gabor Feature Extraction

In this work, we have chosen the Gabor features for face detection. The main advantage of the Gabor filters is that they provide optimal simultaneous resolution in both space and frequency domains [21]. The mathematical formulation can be shown as:

$$\Psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\left[\exp\left(j\vec{k_f}\vec{x}\right) - exp\left(-\frac{\sigma^2}{2}\right)\right]$$

The following characteristic wave vector gives the center frequency of f^{th} filter.

$$\vec{k_j} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_x cos\theta_\mu \\ k_y sin\theta_\mu \end{pmatrix}$$

comprises a scale and orientation which is given by (k_{xy}, Θ_μ). The term exp (- σ^2/2)eliminates the bias, where σ is invariable. Convolving the image with complex Gabor filters with 5 spatial frequency (v = 0,…,4) and 8 orientation (μ=0,….,7) captures the whole frequency spectrum, both amplitude and phase (Fig 4.(d)). In broader sense feature extraction system comprises the two following phase-

(1) Localization of feature point
(2) Feature vector generation.

1. Localization of feature point

Feature vectors including special facial features are extracted from the face image by finding the uttermost pixel in a window $W_0=(x_0, y_0)$ of size WxW by the following procedure:

$$R_j(x_0, y_0) = \max_{(x,y)\epsilon W_0}\left(R_j(x, y)\right), \quad (1)$$

$$R_j(x_0, y_0) > \frac{1}{N_1 N_2}\sum_{x=1}^{N_1}\sum_{y=1}^{N_2}(R_j(x, y)) \quad (2)$$

$$j = 1, .., 40$$

Where R_j is the response of the face image to the j^{th} gabor filter. $N_1N_2$is the size of Face image, the center of the window, $W_0$is at (x_0,y_0).

2. Feature vector generation

Feature vectors [15][16] are formed as a composition of Gabor wavelet coefficient sat the feature points. k^{th} feature vector of i^{th} reference face image is defined as.

$$v_{i,k} = \{x_k, y_k, R_j\,(x_k, y_k), \text{j}= 1,\ldots,40\},\ k=1,\ldots,N_f \qquad (3)$$

Where a feature point is presented by the coordinates (x_k,y_k) and the specimen of the Gabor filter reactions at that coordinate is shown by R_{ij} and for image *i,* the number of feature vectors is represented by N_i.

ii. a). Similarity of feature vector

Now for the computation of sameness of two composite feature vectors, we have applied the subsequent similarity function without concerning the period:

$$S_i(k,j) \approx \frac{\sum_l |v_{i,k}(l)||v_{i,j}(l)|}{\sqrt{\sum_l |v_{i,k}(l)|^2 |v_{i,j}(l)|^2}} \qquad (4)$$

I= 3,...,42

Where the similarity of j^{th} feature vector of the test face($v_{i,j}$), to k^{th} feature vector of i^{th} reference face, ($v_{i,k}$) is represented by S_i (k,j). The similarity OS_i of two faces are calculated by

$$OS_i = \frac{\sum_j Sim_{i,j}}{N_i} \qquad (5)$$

Where,

$$Sim_{i,j} = max \begin{bmatrix} S_i(l,j) \\ S_i(N_i,j) \end{bmatrix} \qquad (6)$$

OS_i represents the overall similarity of test face to i^{th} reference face, where N_i, N_j are the number of feature vectors of the i^{th} reference and test faces, respectively.

ii. b). Face Comparison

In this stage [15], firstly both the location and similarity are found for those reference images whose feature vectors are not so adequate to the feature vectors of the test image and then they are eliminated. In the next step, OS_i is used to find the similarity. To fulfil the goal, each of the feature vector of test face reference faces are sorted according to their similarity measure and the number of times that each of the reference face gets the first position is calculated as;

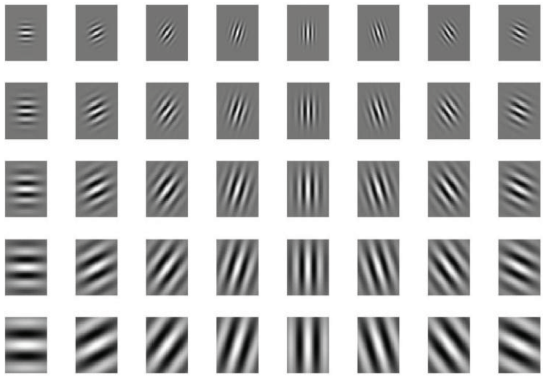
$$C_l= \sum\delta(Sim_{i,j} = \max(Sim_{i,j})), \qquad (7)$$

and to maximize the following, we look for the best candidate match

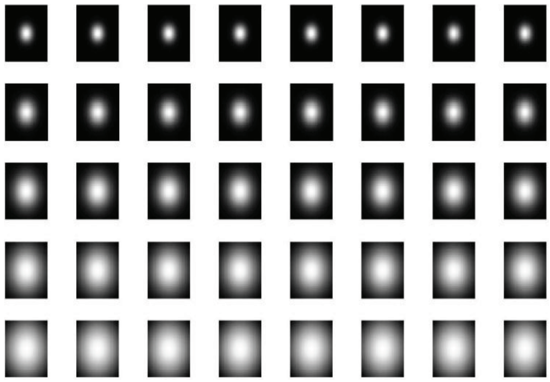
$$FSF_i= OS_i+ \beta(C_i/n_i)$$

where i= l,…… ,shows the number of reference faces, the number of feature vectors of i^{th} reference image is represented by n_i and β represents a weighting factor.

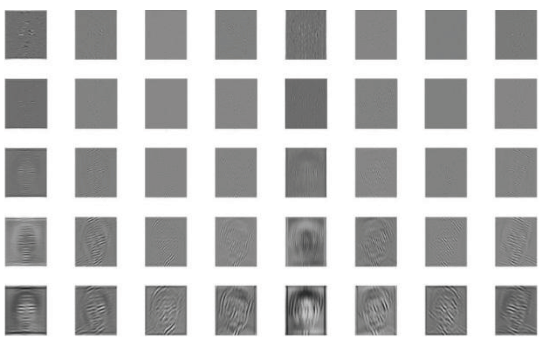
a. Input image for our proposed framework



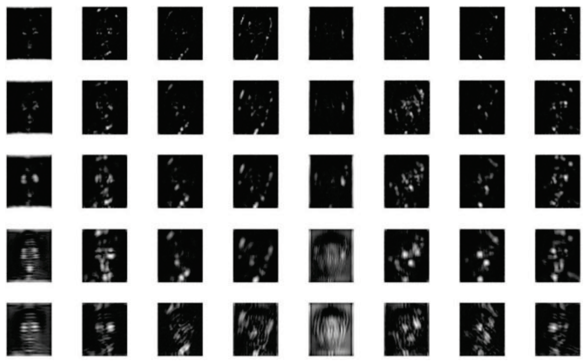
b. Real parts of Gabor filter



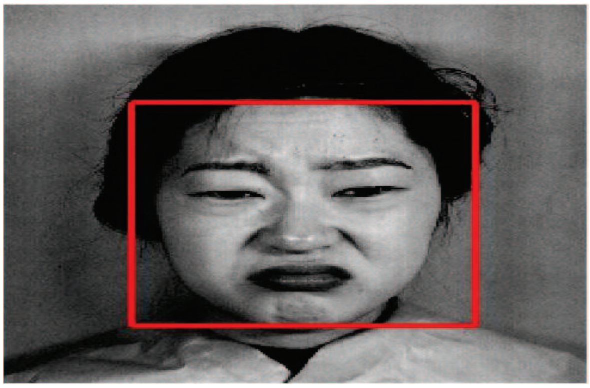
c. Magnitudes of Gabor filter



d. Real parts of Gabor filter of experimented image



e. Magnitudes of Gabor filters of experimented image



f. Face Detected Frame

Fig4. Output of individual steps of Gabor Feature Extraction

B. Learning and Principal Component Analysis

In learning phase, firstly create clusters with respective emotion using PCA [18]. Now whenever a new frame will come it's PCA will be generated & we have compared that PCA in test phase.

Here we have trained our system with different facial expression. Principal component analysis is a technique mainly used for dimensionality reduction, lossy data compression, and feature extraction.PCA can be described by orthogonally projecting the data onto a lower dimensional linear space called as principal subspace in order to maximize the variance of projected data [18].For maximum variance formulation we consider a data set of observation $\{x_n\}$ where n=1, 2, 3 …… N, and x_n is a Euclidean value with dimensionality D. We need to project the data onto a space having dimensionality $M < D$ while maximizing the variance of the projected data. The value of M is assumed.

For the projection we have only considered the space having one dimension ($M = 1$). The direction of this space is defined by a vector u_1 dimension D, then we have considered a unit vector so that $u_1^T u_1$=1. Each data point x_n is then projected onto a scalar value$u_1^T x_n$. The mean of the projected data is where x is the sample set mean given by

$$X = \frac{1}{N}\sum_{n=1}^{N} x_n$$

The variance of the projected data is given by

$$\frac{1}{N}\sum_{n=1}^{N}\{u_1^T x_n - u_1^T x\}^2 = u_1^T S u_1$$

Where we have represented the data covariance matrix by S

$$S = \frac{1}{N}\sum_{n=1}^{N}(x_n - x)(x_n - x)^T$$

C. Testing and Pattern Classification

Then during the test phase we have taken any random image from the dataset and use pattern classification to find the emotion of that particular image. For pattern classification we have used very well-known K-mean clustering with some modified approach. Here, instead of making clusters for each of the individual emotion by creating cluster centre for each of the emotion, we use their principal component as it is. Because for a same human being for similar emotion the output of principal component may vary up to certain extent.

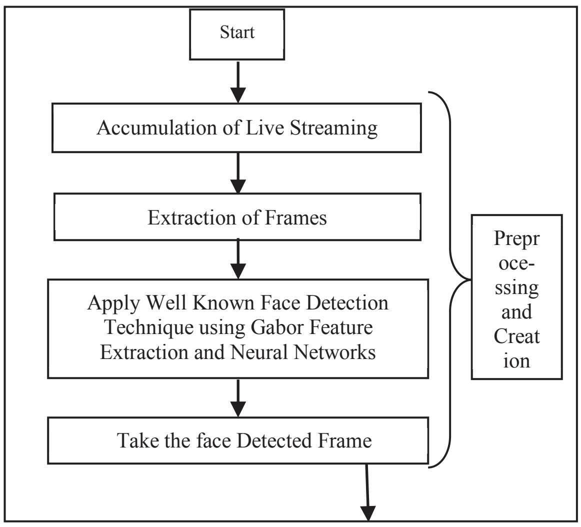
So for testing the emotion we have considered two measure simultaneously-

a) Overlapping measure.

b) Minimum distance measure.

In overlapping measure we have tried to find amount of overlapping of principle component of the test image with the principle component of respective trained emotion, if overlapping occurs.

In minimum distance measure, we have used the K-mean approach. For any principal component of the test image we have measured the distance with the principle component of different or same emotion and considered only the minimum one. Finally comparing these two result we are able to find the emotion of the test image.



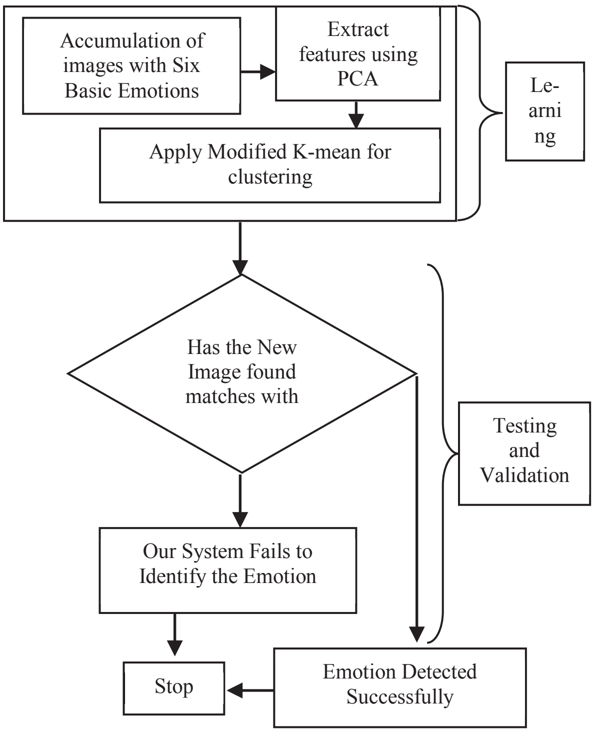
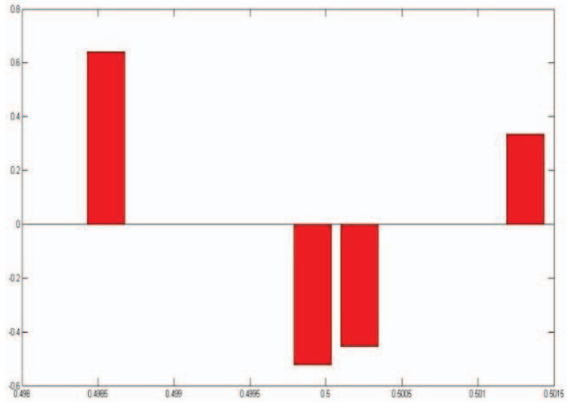


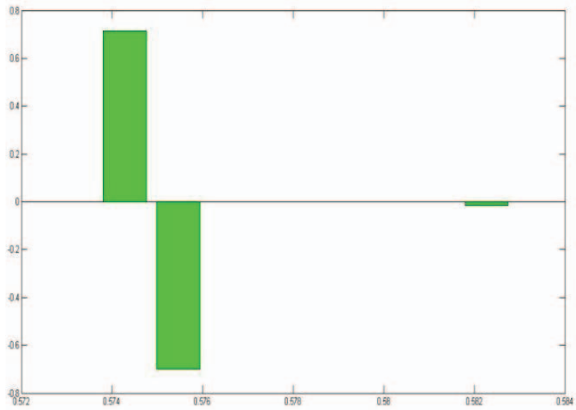
Fig5. Computational Flowchart for Emotion Recognition

VI. RESULTS AND ANALYSIS

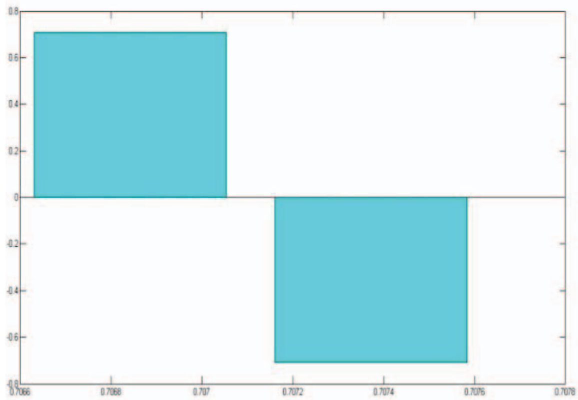
We have represented each of the emotion in Fig 4.(a), (b), (c), (d), (e), (f), (g) with different colour to distinguish them only. For each of the plot we have considered three to four sample for each individual emotion and their principal components and plot those component relatively for better visualization. We can observe that for same person principle components of some cases, the emotions have got overlapped due to their frontal dynamics, depicted in Fig 4.(g), (h). In Fig 4.(g) we have plotted each of the principal component to their respective emotion, but some of them has got overlapped and some of them are not clearly visible due to the interval between each of the plot is very small. To get some justification we taken scatter plot in Fig 4.(h). So we can say one facial expression consists one major emotion and few percentage of other emotion also. We have tried to show the trained output of different emotions individually and together as well with few samples for each of the respective emotion. We have shown the test image to which emotion class it will belong. For testing we have taken a sample with more than one instances (Fig 4.(i)) and evaluated their PCA in Fig 4.(j), then sent that through the trained model. We can observe that in Fig 4.(k) it has overlapped with angry emotion. In Fig 4.(m) we again have considered all the emotion including that overlapping of the experimented image by scatter plotting. We have plotted Fig 4.(k) just for better visibility and overlapping measure and in Fig 4.(m) the plotting was for minimal distance measure to different emotion. The result can be more optimized if we trained our model with large number of versatile input image.



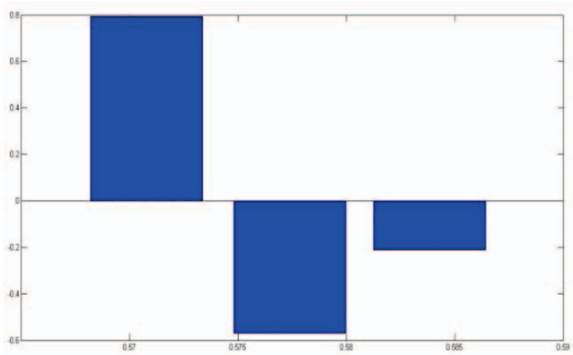
a. Principal component of angry emotion



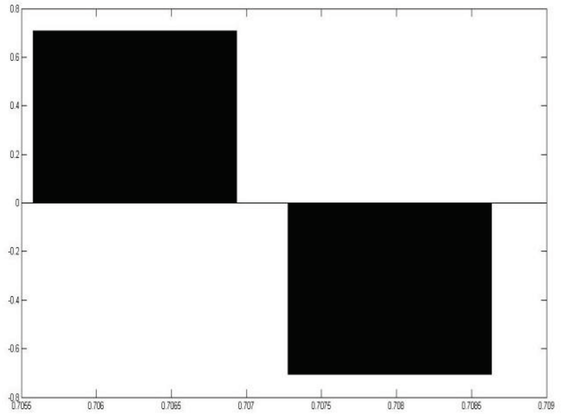
b. Principal component of happy emotion



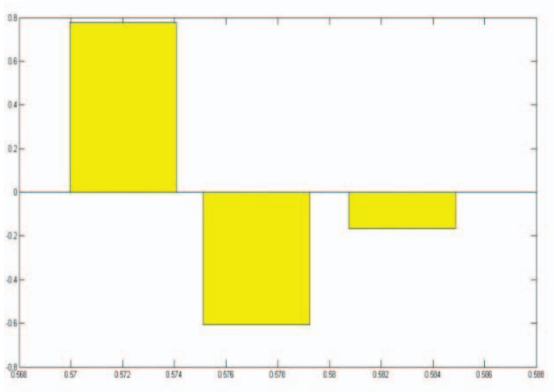
c. Principal component of disgust emotion



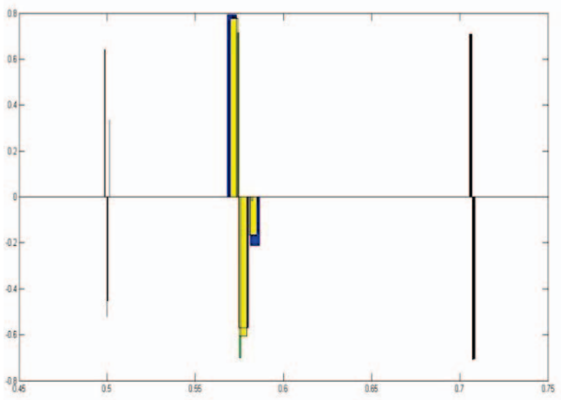
d. Principal component of sad emotion



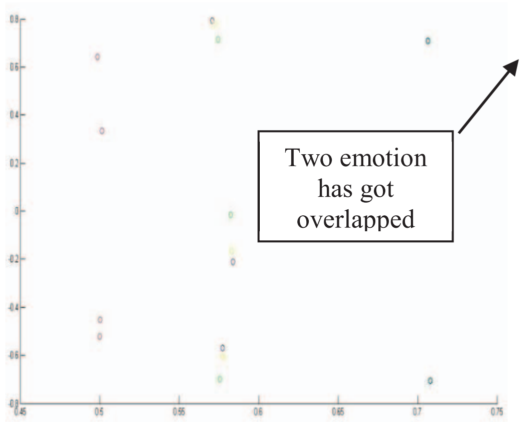
e. Principal component of neutral emotion



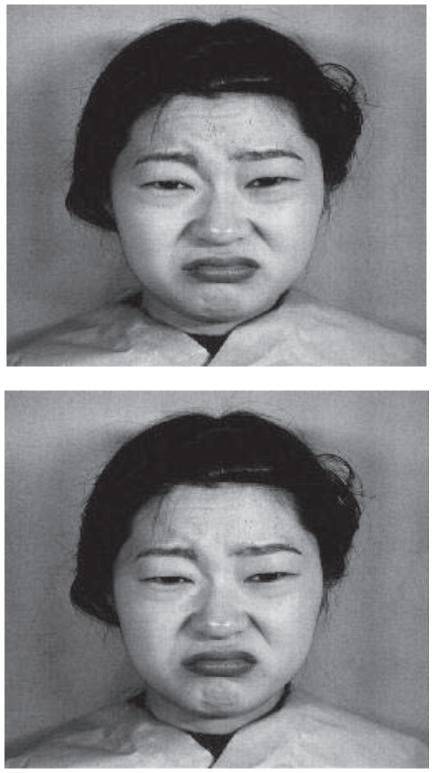
f. Principal component of surprise emotion



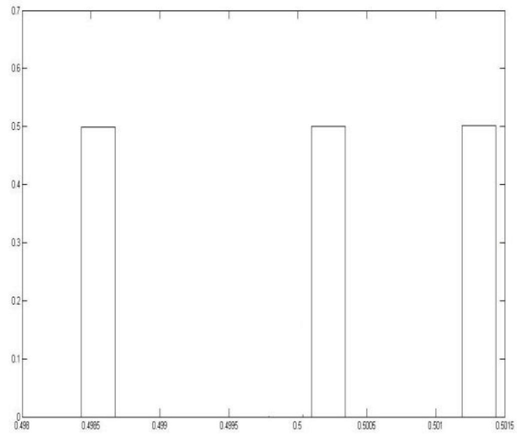
g. Principal component of all emotion



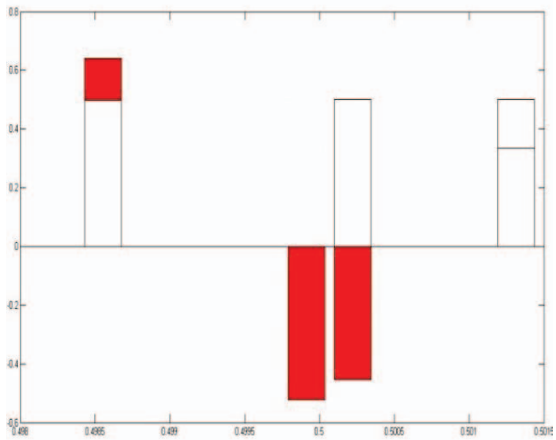
h. Scatter output of all emotion



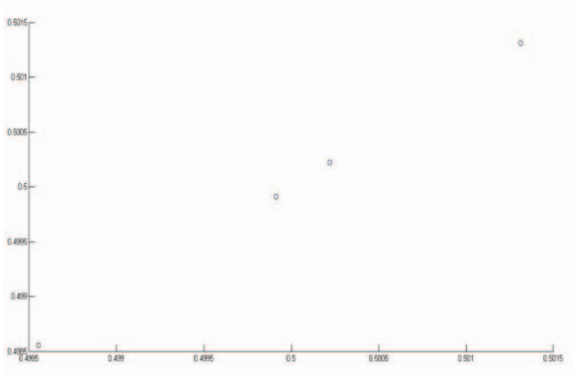
i. Experimented image with different instances



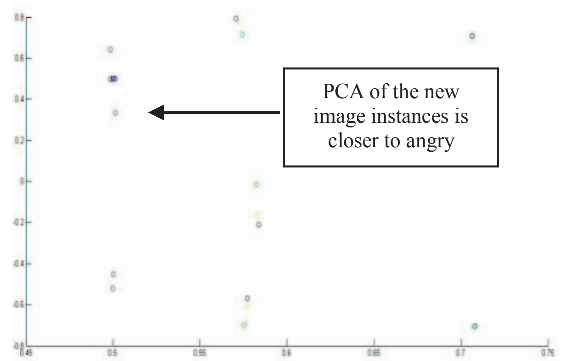
j. PCA of experimental images



k. Overlapping with angry mode PCA



l. Scatter output of the same experimental image



m. Output to show minimum distance measure

Fig 4. Output of each individual steps of our proposed approach

VII. FUTURE WORK

We can make our automated framework for emotion detection more efficient by improving the pattern classifiers by which we will be able to handle more accurately the emotion of new face to which class of emotion-cluster that will belong. However it will be very fascinating if we contemplate by considering both the auditory & visual information and some more attributes like EEG signal, facial color etc. together, for processing with the expectation that this kind of multi-modal information processing will become a datum of information processing in future multimedia era. We can even improve the accuracy by taking the principal component of each individual portion of the face like eye, nose, lips, forehead, cheek etc and then compare with the experimented image.

VIII. CONCLUSION

Till today all of the existing vision system for facial muscle action detection deal only with the frontal-view face images and cannot handle the temporal dynamics of facial actions. Also for some human being, they don't show their emotion and mental state by facial expression, for this kind of situation our proposed model significantly fails to recognize the emotion and provides FALSE POSITIVE result. However, with this shortcoming we have shown based on experimental confirmation that the proposed framework for automatic emotion detection can be well appertained to real time facial expression and emotion characterization task.

REFERENCES

[1] D. Anurag,S. Ashim, '*A Comparative Study on different approaches of Real Time Human Emotion Recognition based on Facial Expression Detection*', International Conference on Advances in Computer Engineering and Applications, 978-1-4673-6911-4/15/$31.00©2015 IEEE.

[2] Kudiri M. Krishna, Said Abas Md, Nayan M Yunus, '*Emotion Detection Using Sub-image Based Features Through Human Facial Expressions*', International on Computer & Information Science (ICCIS). 978-1-4673-1938-6/12/$31.00©2012 IEEE.

[3] Pal Pritam, Iyer Ananth N. and Yantorno Robert E . '*Emotion Detection From Infant Facial Expression*', Internationa Conference on Acoustics, Speech and Signal Processing. 1-4244-0469-X/06/$20.00©2016 IEEE

[4] Silva C. DE Liyanage,Miyasato Tsutomu, Nakatsu Ryohei. '*Facial Emotion Recognition Using Multi-modal Information*', International Conference on Information, Communication and Signal Processing, 0-7803-3676-3/97/$10.00©1997 IEEE.

[5] Maja Pantic, Iounnis Patras. '*Dynamics of Facial Expression : Recognition of Facial Actions and their Temoporal Segments from fave profile Image Sequences*', IEEE Transctions on System, Man and Cybernetis.1083-4419/$20.00©2006 IEEE

[6] Songfan Yang, Bir Bhanu, '*Understanding Discrete Facial expressions in Video Using an Emotion Avatar Image'*, IEEE Tranjactins on Systems, Man and Cybernetis. 1083-4419/$31.00©2012 IEEE

[7] Happy S L, Routray Aurobinda, '*Autamatic facial Expression Recognition Using Features Salient Facial Patterns'*, IEEE Tranjactions on Affective Computing.1949-3045©2014 IEEE.

[8] Mohammad Soleymani , Sadjad Asghari-Esfeden, Yunfu ,Maja Pantic, '*Analysis of EEG signals and facial Expressions for Continuous Emotion detection'*, IEEE Transctions on Affective Computing, 1949- 3045©2015 IEEE.

[9] Leh Luoh, Chih- Chang Huang, Hsueh-Yenhiu. '*Image Processing Based Emotion Recognition*', International Conference on System Science and Engineering. 978-1-4244-6474-6/10/$26.00©2010

[10] F. Abdat, C. Maaoui and A. Pruski, UKSIM 5th European Symposium on Computer Modeling and Simulation. 978-0-7695-4619-3/11$26.00©2011 IEEE.

[11] Kenny Hong, Stephan K. Chalup, Robert A.R. King. ' *A Component based Approach for Classifying the Seven Universal Facial Expressions of Emotion*'. 978-1-4673-6010-4//13/$31.00©2013 IEEE

[12] A. Ghahari, Y. Rakhshani Fatmehsari and R.A Zoroofi, '*A Novel Clustering- Based Feature Extraction Method for an Automatic Facial Expression Analysis System*'. 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 978-0-7695-3762-7/09$26.00©2009 IEEE.

[13] Li Zhang, Ming Jiang, Dewan Farid, M.A. Hassain, '*Intelligent facial Emotion recognition and Semantic- Based topic detection for a Humanoid Robot*'.0957-4174/$ © 2013 Elsevier Ltd.

[14] Lin-Lin Huang, Akinobu Shimizu, and Hidefumi Kobatake. '*Automatic Face and Gesture Recognition. Classification-Based Face Detection Using Gabor Filter Features'*, Sixth IEEE International Conference. 0-7695-2122-3/04 $ 20.00 © 2004 IEEE.

[15] Wang Chuan-xu, Li Xue. '*Face detection using BP network combined with Gabor wavelet transform'*, Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing. 978-0-7695-3263-9/08 $25.00 © 2008 IEEE.

[16] Burcu Kepenekci, F. Boray. '*Occluded Face Recognition Based on Gabor Wavelets*'. 0-7803-7622-6/02/$17.00 02002 IEEE.

[17] http://www.kasrl.org/jaffe_info.htm.

[18] Christopher M. Bishop.'Pattern recognition and Machine Learning', Springer.ISBN 978-0-387-31073-2.

[19] Gupta Bhaskar , Gupta Sushant , Tiwari Arun Kumar. *'Face Detection Using Gabor Feature Extraction and Artificial Neural Network',*International Journal of Computer Science and Technology.IJCST Vol. 1, Issue 1, September 2010

[20] A. G. Ramakrishnan, S. Kumar Raja, and H. V. Raghu Ram. '*Neural Nework- Based Segmentation of texture using gabor feature*'. 0-7803-7616-1/02/$17.00©2002.IEEE.

[21] Muthukannan K, Latha P and Manimaran.'*Implementation of Artificial Neural Network for Face Recognition Using Gabor Feature Extraction.',*ICTACT Journal on Imageand Video Processing, ISSN: 0976-9102 , volume : 04, issue: 02, november 2013.