International Conference on Machine Learning and Computational Intelligence-2017
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 7 | ISSN : 2456-3307

Multiple Instances Based Emotion Detection Using Discriminant Feature Tracking

Abhishek Kilak, Namita Mittal
Malaviya National Institute of Technology, Jaipur, Rajasthan, India

ABSTRACT

Automatic recognition of emotions from the facial expressions continues to be an important aspect in the field of evolution of new age computing systems. Emotion detection capable computer systems is an ongoing research field that has a numerous applications ranging from recommender systems, human-computer interaction, robotics and affective systems. The various features that increase the complexity of emotion recognition systems include ethnicity, gender, pose, occlusion, beard, moustache etc. The type of database used for learning by systems is of crucial importance. Many databases exist for this purpose but none of them is for posed Indian faces. In this paper,we bridge this gap by providing Bharat Database which contains facial images of Indian people. The database has posed expressions of 102 participants and has 896 images. The participants were asked to pose for different emotions by showing them images eliciting those emotions as well as with the help of expert artists. The annotation was done using polling by a panel of three experts.Experiments were conducted on the database using different algorithms and results are presented for reference. This database will further help the community involved in developing of algorithms for emotion recognition. In this paper we propose two emotion detection approaches, the first one is based on Compact Local Binary Pattern and is used for construction of hybrid features which increases emotion detection accuracy. Second approach is Enhanced Feature Extraction using multiple Patches face on images from indigenously developed Bharat Database of Indian Faces, Japanese Female Facial Expression Database and Karolinska Directed Emotional Faces. The results show its applicability for construction of emotion detection systems.

Keywords : Emotion detection, facial expression, emotion recognition, facial images, Indian faces.

I. INTRODUCTION

An acceptable approximation of emotion of people automatically is indispensible information for affect enabled systems. There are various approaches for emotion detection but still challenging is their availability, unintrusiveness and accuracy.

Emotional states of human beings can inferred from the gestures, text analysis, voice analysis, brain mapping, breath analysis, keyboard pressure information and face expressions. The keyboard and mouse approach for gathering log of user action is one of the simplest ways of predicting human emotions, but nonetheless it lacks inaccuracy [26].Also keyboard pressure analysis for classification of emotions suffers from having a limited scope only as people are not constantly using keyboards and also all but lab equipment are not equipped with necessary hardware to deliver the required information required for mapping them to emotional states. On the other hand there are approaches that based on sensors that are intrusive and gather physiological signals by way of EEG [27] [28] and ECG [29]. In between the two extremes are approaches that rely on breath analysis [30], gesture analysis [31], [32] and facial information [33].

Exhibition of emotions through gestures and face expressions by humans is the most common aspect. The reverse engineering of analyzing emotions from these expressions is a very common fundamental for humans. They are capable of communicating among themselves through the exchange of these mutually dependent parameters. Gestures may be controlled by people voluntarily and thus may not be coherent with the emotional state. Text analysis is generally a binary analysis of either positive or negative state and thus may provide information only about valence. While emotional state comprises of not only valence but orthogonally arousal also. Brain mapping is an important technique to predict emotional states based on surges of neurons of human brain. This technique is more lab based and not practical as the subjects under consideration are always aware of their examination and thus actual emotional state may be not be predicted. Also this technique involves we ring of monitoring caps which is obtrusive, limited to lab environment and impractical in real life situations. Breath analysis method for emotion detection is again an obtrusive technique and sufferers from the same limitations as of brain mapping technique. Humans are capable of reading gestures and facial features and understanding them near precision. But when it comes to computers, the research is novice and there is a great scope of improvement.

Face expressions provide an insight into the emotional state of human beings. Face expressions change in tandem with emotional states and people generally have limited or no control over their facial movements. Also if people might try to control their expressions in some specific temporal space but expressions are generally spontaneous and in tandem to their emotional state. Therefore emotion detection based on analysis of facial features is undoubtedly the best way to detect emotions as it is not limited to lab environment, unobtrusive, coherent to emotional state and practically applicable.

The real life applications of this research are enormous and include but not limited to Artificial Intelligence, Affective computing, Driver assistance systems, Robotic interactions, Surveillance systems, Mob control systems, Military applications, health care support systems, personalized recommenders for online as well as real life shopping, Content delivery for social media, Movie Target Audience analysis, sentiment analysis of public towards political and business decisions, tailored support solutions for differently abled persons, anxiety bipolar and depression detection, crowd behavior analysis and Smart City as well as IoT applications.

Detection of emotions from face expressions involves extracting mathematical information from spatial variations of the faces. This spatial information also invariably involves ethnicity, gender, scale correction as well as noise features involving occlusion submounting to beard, moustache, hairlocks, face Rotation, resolution and physical hindrances towards full face detection.

Development of systems capable of emotion classification invariably requires training to catch the strains that effectively and categorically map them to different categories. This in turn requires data for the systems to train, which has the following two aspects. Either the training data can be too personalized to serve for only particular subject involved. For eg: Personal response systems based on specific requirement of individuals. This involves catching individual data and catering tailor made solutions bases on them which cannot be applied to pro rata basis. Or, there may be robust general systems that are capable enough to provide services to the unknown masses based on rigorous training provided that involves non ideal conditions of data parity, ever unknown individuality and other parameters mentioned in the previous paragraph.

This requires creation of database that is essential to train the systems, test them, validating of applicability of algorithms for emotion classification systems and building of robust systems. Database creation is toilsome and tedious task. But the less is the fact that it is a prerequisite to build systems that are capable of detecting emotions. Creation of posed expression database requires proper guidance and training of the subjects by experts. Thereafter validation of database is an equally important concern. This involves labeling of the images with categories that they belong to.

In literature Paul Ekman [1] reported six basic emotions that are valid across human species irrespective of gender and ethnicity. These are happy, sad, angry, fear, surprise and disgust. Humans can predict the emotional state of people of different ethnicity. But such is not the case with computers as they need to be trained for different for the texture, shape and appearance of the target subjects. Because of different shape and texture of people across globe, headway in the field of affective computing requires different databases covering ethnicity. To our knowledge there is no posed expression database for Indian faces that catches all the above mentioned emotion classes.

For capturing and building of affective systems it is a prerequisite to capture these spatio-temporal displacements happening in facial features due to underlying muscles. The spatial and temporal assessment of prominent facial features may be utilized for categorization of emotions. Many facial emotion techniques have been proposed which have considered 2D images, 3D images [34], [35], expressions exhibited by infants, AAM [36] based systems and AU based systems [37].

Non deliberate and deliberate ace expressions are the two categories of face expressions defined by Battocchi et al. [38], Expression which are deliberate are expressed under the absence of speech. Whereas those expressions exhibited along with speech are termed as non-deliberate. Also Valence, Arousal and Dominance multidimensional space is sometimes used for separating the emotions categorically.

This requires creation of database that is essential to train the systems, test them, validating of applicability of algorithms for emotion recognition and classification systems for building a robust systems. Database creation is toilsome and tedious task. But the less is the fact that it is a prerequisite to build systems that are capable of detecting emotions. Creation of posed expression database requires proper guidance and training of the subjects by experts. Thereafter validation of database is an equally important concern. This involves labeling of the images with categories that they belong to.

We propose two approaches to automatic recognize six basic emotions. First is compact local binary pattern representation of the images for extraction of features which are used for construction of hybrid feature vectors for classification of emotions. The other approach is Enhanced Feature Extraction using multiple Patches face on images from indigenously developed Bharat Database of Indian Faces, Japanese Female Facial Expression Database and Karolinska Directed Emotional Faces. The results show its applicability for construction of emotion detection systems.

The main contributions of this work include:

1. Construction of Database of Indian Faces with emotion annotation.
2. Propose a novel method for Compact Local Binary Pat- tern.
3. Integrate it with Histogram of Oriented Gradients for classification.

4. Propose a novel technique for extraction of features using multiple patches face on images.
5. A challenging facial expression database of Indian faces is introduced with benchmark for comparative analysis.
6. The results from the experiments shows robust performance by macro analysis of person independent emotion detection.

The rest of the paper is organized as follows. Section 2 introduce the relevant works on existing database for facial emotion and techniques for dynamic analysis of facial expression. Section 3 described the procedure of creation of proposed database. Section 4 contains the description of dataset. Section 5 and 6 illustrated the proposed work of feature extraction and emotion detection respectively. Further, experimental evaluation and classification results are given in section 7. Section 8 conclude the paper with a summary and discussion.

II. RELATE DWORK

As presented work participated in both database creation as well as emotion detection techniques, to maintain the information related to both of the domains we have presented the related work in two fold. First we present different techniques used for dynamic facial expression detection. Thereafter we discuss about the various databases available for facial expression detection.

2.1 Existing Techniques used for Facial Expressions

Various techniques have been proposed in recent times to classify emotions based on facial expressions. The features are extracted from the images or videos using appearance based and geometric approaches.

Geometric feature based approaches dependent upon fiducial points of the facial images [14]. The classification is based upon recognition of movement of facial points in spatial domain. The temporal parameters can also be used for recognition of expressions by analyzing shapes across frames of videos. Many researches thrive upon posed facial expressions [15] for emotion recognition.

Texture in images can be captured using Local binary patterns [12]. LBP is computationally simpler and also is tolerant to illumination conditions. LBP features on orthogonal planes of space and time are applied by Zhao et al. which led to introduction of temporal features augmenting the spatial textural features [13]. Local Phase Quantization [16], Histogram of Gradients (HoG) [17], LBP and SIFT are the commonly used low level features for feature extraction and classification and may be extended to include temporal features for real time applications. Being simpler computationally, robust in extraction of textural features and having the potential extendibility to include features from temporal domain makes LBP a popular choice foe feature extraction [18]. An extended approach of LBP comprising circular neighbors of different radius and showing a further discriminating presentation has also been proposed [19].

Technique based on blocks is generally used taking into account the locations of subregions and co-occurrences [20], [21], [24]. Use of this technique was done primarily for recognition of faces [22]. Image was partitioned in regions that were non overlapping and Local Binary features were derived and weighted according to their importance in contributing to identification and then appended for formation of augmented representation. This method was further extended by Zhao et al. by considering overlapping blocks for the experiments [23]. However, person related bias and resultant over fitting was not care of in their experiments. Weight based approach was considered by Shan et al. [22], wher only expressions were considered and not identities [25]. However it lacked the leverage of using overlapping blocks.

For detection of emotions localization of mouth and eye part has been performed in many approaches [39]. This technique suffers from loss of data as the rest portion of the face is not considered. Face motion tracker method has been proposed by Shen et al [40] which is composed of combination of multiple models for construction of cost function. The constituent models of the cost function are varied according to the application scenario.

Viola Jones is widely used algorithm for detection of face part in the images [41]. Thereafter Gabor filters are applied at different scales and orientations for building of feature vector [42]. This feature vector is then used for classification of emotions in different categories. Active Shape model is used by some researchers which has the drawback of using only shape constraint information. This is overcome by a widely used approach for tracking of faces and emotion detection is the Active Appearance Model which is a geometrical approach that uses texture information also. AAM is used for the matching of a statistical model of the shape of facial features upon actual facial image. Supervised training is done using landmark coordinates which are posted on the training image set [43].

The manual location of feature points was done by Wang et al [44] at prominent features nearing eyes, mouth and eyebrows for detection of emotions. However it is not practical as it lacks positioning of these lines and dots in the impending images after training. Kernel methods based on time series have been used for recognition of emotions upon landmark data by Lorinez et al [45]. The have shown that Global Alignment Kernel or Dynamically Time warping are required for emotion categorization. A system registering facial expressions based on series of steps that detects emotions has been proposed by Sariyanidi et al. [46]. The registration of rigid, parts and point parameters is done for encoding the image sequences frame by frame. Shapes are represented using coordinate pairs of a sequence of facial points. The bulky information is then reduced dimensionally and the resultant is used for emotion recognition. Various statistical measures are used which are complex in nature and recognition is done on posed data only.

Multiscale learning based upon high and low level spatio- temporal facial expressions has been proposed by Liu et al [47]. The high level features constitute different gestural events which are assessed for different duration. Low level features are comprised of information obtained from head pose, appearance and face geometrical features. However the system is very complex to be deployable in real life situations.

2.2 Existing Databases of Facial Expressions

Automatic recognition of emotions from the facial expressions continues to be an important aspect in the field of evolution of new age computing systems. The various features that increase the complexity of emotion recognition systems include ethnicity, gender, pose, occlusion, beard, moustache etc. Therefore there had been an emergence of several databases covering various features that are available publicly emotion recognition. The different databases are discussed in brief in this section.

One of the most widely used database is Cohn-Kanade Action Unit Coded Facial Expression database. It comprises of 486 image sequences posed by 97 subjects was released in year 2000. The image sequence proceeds from neutral face image to extreme expression. The peak expression images were coded using Facial Action Coding System and annotation is provided in form of emotion labels. It had the limitation as the emotion labels were not validated. Rather the emotion labels were those which were asked from the subjects to perform.

This dataset was extended to address as Extended Cohn- Kanade (CK+) database [2]. The images of 123 subjects corresponding to 593 sequences were taken which were coded using Facial Action Coding System for the last or peak frame of the sequence. The Action Units and their intensity were provided for the peak expression images. Also the images were tracked using Active Appearance Model for 68 landmark points. Out of the total sequences only 327 were having corresponding emotion files. This was because only these sequences were validated. The emotion labels were neutral, anger, contempt, disgust, fear, happy, sadness and surprise.

Japanese Female Facial Expression (JAFFE) database has 213 images 10 female Japanese models that posed for both neutral and six basic expressions. JAFFE database was created by Lyons et. al. [3]. The images are in grayscale.

Binghamton University-3D Dynamic Facial Expression database [4] has 2500 3D face expressions of 100 subjects having the six basic expressions having four intensity levels. The different aspects considered include age, race and culture.

The MMI Database was initially conceived in 2002 with the goal of serving as a source that can be used across facial expression recognition community [5]. It has videos that has sequence from neutral to apex and back to neutral expression. It has over 2900 videos of 75 subjects as well as still images. The videos are annotated for presence of Action Units.

The Belfast Database [6] has different sets of over 250 coloured video clips depicting natural emotions at different resolutions. Multimedia Understanding Group (MUG) has 1462 posed color sequences of 86 subjects which are annotated with emotion labels. The Radboud Faces Database (RaFD) has posed color images of 67 subjects at five different camera angles and three different gaze directions for eight emotion labels.

Indian Spontaneous Expression Database (ISED) [7] has 428 spontaneous color videos of 50 subjects having emotion labels of sad, happy, surprise and disgust only.

Denver intensity of spontaneous facial action database (DIFSA) [8] is a color video database of 27 subjects whereby each video sequence is of 4845 frames of spontaneous reactions while viewing video of 4 minute duration. Six intensity level annotation of Action Units is provided for the facial expressions.

III. CREATION OF THE DATA BASE

The BDIF has still posed facial images for various emotions. Subjects were asked to pose for all the basic as well as neutral emotion. The participants were asked to pose for different emotions by showing them images eliciting those emotions as well as with the help of expert artists. The photographs were taken in well lit conditions.

The BDIF was carefully constructed by showing the subjects valid labeled emotion images from different databases and also with the help of expert artists training them how to elicit the said emotions. The effectiveness in expressing emotion was due changing the mental state of the subjects by showing them visual cues and also narrating real life situations which pertain to those particular emotions.

The annotation of images was done using polling by three annotators who were familiar with Facial action coding system. They were shown images and told to classify them as belonging to one of the classes. Only those images were labeled in which a consensus arrived among the annotators.

3.1 Experimental Setup

The subjects participated voluntarily for the posing of expressions. They were made comfortable by telling as how to pose and showing visual cues as well as real life situations which are associated with the particular emotional state. The subject being comfortable with the environment and counseled properly led to the capture of expressions effectively.
These images were captured in ideal conditions which was free from noise and other distractions. The images were captured in ambient light conditions. According to experience obtained from preliminary studies, the light conditions were as subjects are generally accustomed to and not being very bright or dull. These made the subjects comfortable with the experiment environment.

Closed rooms were used for the conduction of shoot sessions. The subjects were made well aware about the experiment and also as how it will help in future scientific studies.

The subjects were asked to stand comfortably taking the sup- port of wall. They were allowed time to ease them for elicitation of emotions. Thereafter images were captured by posing for neutral, anger, contempt, disgust, happy, fear, sadness and surprise. In order to avoid disturbance the rooms were kept close during the different shooting sessions. In first part of the experiment it was found after annotation that the sample of anger and disgust emotion was the least. The statistics of first phase is shown in Figure 1.

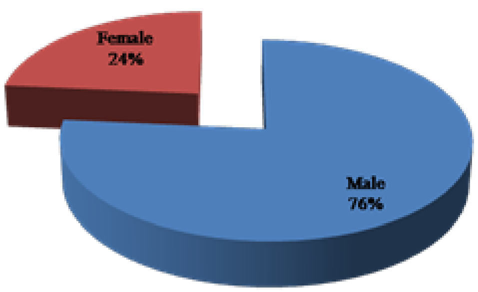

Figure 1. Male Female Ratio.

The images were taken comprising of 4320×3240 pixels at 300 dpi horizontal and vertical resolution using Nikon Coolpix 120. The images were taken with compulsorily no flash. Distance maintained between subject and the camera was around 1 meter. The emotion label categories for the complete database is shown in the Figure 2.

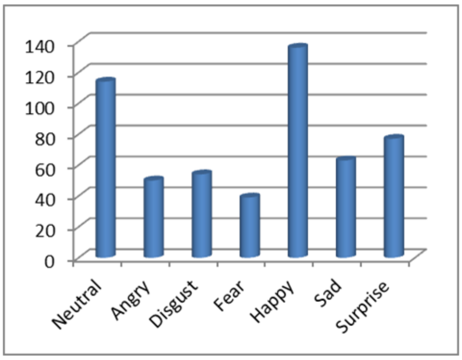

Figure 2. Graph showing the emotion label categories for the complete database.

3.2 Obstructions

Gathering information from facial expressions may be hindered by the presence of glasses, moustache, beard etc. These act as noisy features in implementation of automatic emotion detection algorithms. However they are a regular feature in the real life situations. Therefore these features were deliberately included in the formation of database so that more robust emotion detection algorithms can be built using such images also from the database.

3.3 Annotation

The images obtained for the database were subjected to labeling for different emotions. This is particularly important as it serves as ground truth for comparison with results obtained by applying automated algorithms. Therefore the technique of validating emotion labels is substantial significance. The images were labeled by a panel of three annotators who were familiar with facial action coding system. The labeling was done for neutral, happy, angry, disgust, contempt, surprise, sad, fear and invalid emotions. For validation polling was done for each image by gathering the emotion labels of all the three annotators . Only those images which had a majority for particular emotion label were validated for that particular emotion. Rest of the images were not annotation validated for the particular emotion labels.

The resultant database consisted of:-

Total Subjects-59

Images Collected-436

Total Classes of Valid Emotions: 7

1) Neutral : 71
2) Angry: 24
3) Disgust: 14
4) Fear: 3
5) Happy:91
6) Sad: 40
7) Surprise: 47

The results of the first phase of database creation led to a finding that even when the subjects were made comfortable with the environment, they could not provide fear and disgusted motions. This was because people do not feel comfortable to exhibit these emotions willingly and publically. Therefore in the next phase, the subjects were given privacy and shoots were performed in confined places so that they can comfortably exhibit those emotions. Also subjects were provided visual cues as what that expression looks like. Assistance was provided to them with the help of professional artist to help them exhibit those emotions.

The database was extended in the next phase by including new subjects and taking care of findings of the first phase. The extension led to:

New Subjects-43

Images Collected-460

Total Classes of valid Emotions: 7

a. Neutral : 43
b. Angry: 23
c. Disgust: 39
d. Fear: 36
e. Happy:41
f. Sad: 23
g. Surprise: 36

3.4 Consent from Subjects for publication

Initially the subjects were reluctant to pose for the Database. The subjects involved in the collection of dataset were given incentive like chocolate, juice and ice cream. The participants were informed that their images may be used for publication for research purpose. Ethically the subjects were asked to fill up their consent for publication of images for research purpose. Images of those subjects who did not provide consent were deleted from the final database.

IV. BHARAT DATA BASE

BDIF initially had expression images of 59 subjects for seven classes of emotions viz. neutral, happy, angry, fearsome, disgust, surprised and sadness. It initially had 436 images. Thereafter it was extended by addition of expression images of 43 new subjects exhibiting the seven classes of emotions. 460 new images were added in this phase. The complete database thus has total 102 subjects and a total of 896 images. The class wise distribution is shown below:-

1) Neutral : 114
2) Angry: 50
3) Disgust: 54
4) Fear: 39
5) Happy:136
6) Sad: 63
7) Surprise: 76

Total Male : 78

Total female : 24

Sample images of different class of emotions can be seen in the Figures 3 - 9.

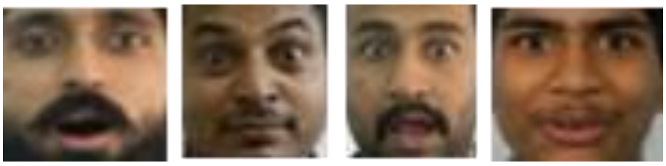

Fig. 3. Images from Bharat Database of Indian Faces exhibiting Surprise emotion.

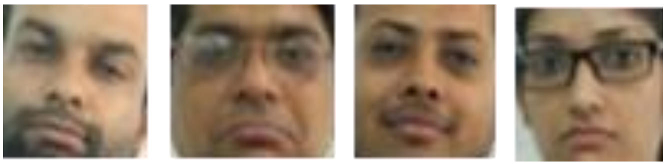

Figure 4. Images from Bharat Database of Indian Faces exhibiting Sad emotion.

Figure 5. Images from Bharat Database of Indian Faces exhibiting Neutral face.

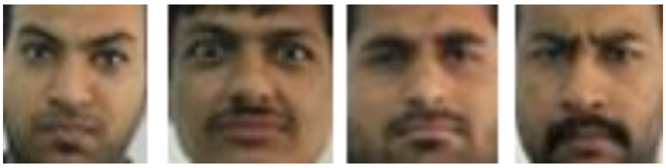

Figure 6. Images from Bharat Database of Indian Faces exhibiting Anger emotion.

Access to this database may be made available for research purpose only by sending an Email to the author at 2012RCP9516@mnit.ac.in.

V. PROPOSED WORK I

The steps involved in emotion detection involve the following steps:

1) Image Preprocessing
2) Extraction of Feature Vector
3) Feature Vector Vector Size Reduction
4) Emotion Classification
5) Compact Local Binary Pattern extraction
6) Statistical moment and other feature calculation and ap- pending
7) Classification

5.1 Preprocessing

Viola Jones [9] algorithm was used for detection of the face part from the images of Bharat Database of Indian Faces which resulted in 100% accuracy on our dataset by adjusting minimum permissible height and width to 600 pixels. Cropping of images was done so that the images may contain only the face part. The images of the database are colored and they were converted to grayscale for the conduction of experiments. Sample of images which were converted to grayscale and cropped are shown in The Figure 10. The images were resized to 1024*1024 pixel size.

5.2 Extraction of Features

Histogram of Oriented Gradients were calculated for the grayscale cropped images that had corresponding emotional tags. Edge directions distribution or gradient intensity are able to catch shape of an object and local appearance. Division of image into cells of uniform length and breadth was done in the experiment for cell sizes ranging from 10×10 pixels to 512×512 size.

Gradient directional histograms were then computed for the cells. HOG blocks of $2 * 2$ cells with 50% overlap were used for the construction of feature vector comprising of orientation binned cell histograms. Sample Histogram visualization is shown in the following Figure 11.

$$T_{Blocks} = floor\left(\frac{Len}{Cell} - 1\right) \mathrm{X}\, floor\left(\frac{Bre}{Cell} - 1\right)$$

Where used to denote total number of blocks in an image. Since every block is composed of $2 * 2$ cells, therefore

$$T_{Cells} = T_{Blocks} * 4 \qquad (2)$$

Where T_{Cells} represents total number of cells that participate forthe computation. The total number of histograms per cell used in this work is given in Equation 3.

$$Hist_{Cell} = 9 \qquad (3)$$

5.3 Reduction of Feature Size

Histogram features for gradient orientations [10] were extracted for cell sizes ranging from 64*64 pixels to 512*512 pixels. Experiments were conducted on this feature vector and reduced feature vector which was done by applying principal Component Analysis [11] covering variance of 0.95.

5.4 Extraction of features using Proposed Compact LBP

A new approach is proposed as Compact LBP in which the neighbouring pixels of the previous were discarded in the resultant image. This resulted in image size reduction from 1024*1024 pixels to 341*341 pixels.

The conversion process is shown in the following figures with the help of sample pixel values for an image subsection.

The resultant Compact LBP Images are shown in the following figure.

5.4.1. Calculation of Statistical moments and other features and appending

For each Compact local pattern image mean and median were calculated for each pixel row. This provided feature vector of length 341 + 341 i.e. 682. To this 51 histogram counts were added for each image. Also mean of the whole image and standard deviation were calculated. This resulted in feature vector of length 735 for each image. The results were calculated for classification using different classifiers. In the next phase results were classified appending this feature vector to the feature vector histograms of oriented gradients.

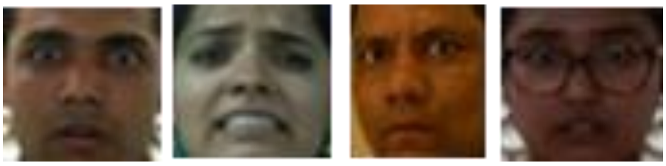

Figure 7. Images from Bharat Database of Indian Faces exhibiting Fear emotion.

Figure 8. Images from Bharat Database of Indian Faces exhibiting Disgust emotion.

Figure 9. Images from Bharat Database of Indian Faces exhibiting Happy emotion.

Figure 10. Grayscale Cropped Image samples.

VI. PROPOSED WORK II

The procedure for computing structural Features is illustrated in Algorithm 1.

The first database used for the experiments was Bharat Database of Indian Faces which is composed of total 102 subjects and a total of 896 images. This database has images of school going children of 11th and 12th class, subjects doing graduation, post-graduation and research scholars, staff members from different offices and random people. There are occlusions of beard, moustache and spectacles.

The class wise distribution is shown below:-

Neutral: 114; Angry: 50; Disgust: 54; Fear: 39; Happy:136; Sad: 63; Surprise: 76

Total Male : 78

Total Female : 24

6.1 Marking of points on Neutral faces

This was followed by marking of points on the neutral faces of these subjects. The neutral face in the images was subjected to tracking points manually. These points covered eyebrows, eyes, nose and lips of the subjects. The reason for this allotment was that these parts of the faces convey the most information. In literature associated to physiology these parts contemplate to Action Units. Sample image with point marked is shown in Figure 17.

6.2 Tracking of the points in corresponding images

The points established in the previous phase on the neutral images were averaged in this phase. Matrices corresponding to 5×5 grid around pixel points of considered. The pixel values of these points were taken and average value was calculated.

R[ ]= R(r – I R + 3, c – I C + 3)

G[ ]= G(r – I r + 3, c – I c + 3)

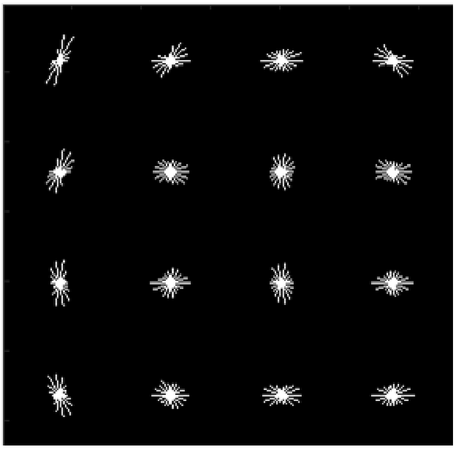

Figure 11. Cropped Image Histogram visualization for cell size 256.

B[ ]= B(r – I r + 3, c – I c +3) Where r = I R – 2 to I R + 2 and c = I C – 2 to I C + 2

Where I R and I C correspond to initial Row and Column. Average Color Channel values were calculated for each color.

These Average Color Channel values were used to track these points in the corresponding emotion images of the subjects. Maximum permissible distance of 15 pixels in both directions was used in searching of probable points within the periphery of initial points of consideration.

$$AvgR1 = \sum_{i+2;j=2}^{i-2;j-2} \frac{R1}{25}$$

For each pixel in periphery of 15×15.

6.3 Calculation of distance vector

In this step the Euclidean distance between the initial point positions and the tracked point positions. This is particularly important as the displacements capture information that may be

115	121	214	56	78	98
87	98	76	211	58	67
123	124	178	189	111	48
76	78	67	90	99	111
99	100	102	104	88	76
244	235	68	89	96	89

Figure 12. Initial image sample pixels.

1	1	1	0	1	1
0	98	0	1	58	1
1	1	1	1	1	0
0	0	0	1	1	1
0	100	1	1	88	0
1	1	0	1	1	1

Figure 13. Applying Binary pattern on the immediate neighbors of the pixels shown in red color.

used to categories among various emotions. The distance vector is composed of these corresponding displacements.

$$D = \sqrt{(X_i - X_j)^2 - (Y_i - Y_j)^2}$$

Algorithm 1 Mark Point Technique

Ensure: A – Emotion Accuracy

Require: Ii -Image i = 1......T ∗ t

1: For each image Ii do

2: Take a netural emotion image of the subject.

3: Crop and resize

4: Mark the points(N)

5: Read coordinatorof all the points

6: Get R, G and B channel values in windowof −2 to 2 x and y values around

7: Calculate Average R, G and B values for each window around N points

8: end for

9: For i = 1 to T

10: for For j = 1 to all images in i folder do

11: Find face, Crop and Resize

12: Get R, G and B channel values in widow of −9 to +9 x and y values around

13: end for

14: for For each N (x, y) in emotion image and in periphery of 3 to 17 pixels do

15: Calculate Average R1, G1 and B1 for widow size −2 to +2 in x and y direction

16: Slide the widow

17: end for

18: for For each Value obtained do

19: Calculate the Difference=Average R - Average R1

20: Repeat for other 2 color channels

21: end for

22: Calculate the point having minimum difference out of values obtained.

23: Obtain coordinates of the pixel whose window has minimum Difference.

24: Calculate signed difference in coordinates of Neutral Image and Emotion Image for all N points

25: Classification

Calculation of the points having minimum difference out of values obtained was done. Thereafter coordinates of the pixel whose window had minimum difference were obtained. Also signed difference in coordinates of Neutral Image and Emotion Image for all N points was calculated in this step.

VII. CLASSIFIC AT ION AND RESULTS

The feature vector obtained in the previous step was then used for classification. The results were matched with the emotion labels provided in the database for obtaining percentage of ac- curacy. The resultant vector of Histogram gradients was subjected to classification. Carrying the experiment further this vector was subjected to principal component analysis and subsequently then subjected to classification. In the next phase resultant vectors of statistical moments of Compact LBI images were classified to obtain accuracy percentage. Further on this vectors was appended to the feature vector obtained by HOG and then classified to obtain accuracy percentage at five fold cross validation is shown in Table1 and 2 and in Figure 17.

The results in terms of obtained accuracy at 5 fold cross validation (HOG) is shown in the following Table 1 and Table 2.

Table 1. Results at various cell sizes from 64 to 104

Classifier | Cell Size	64	72	80	88	96	104
SVM	66.2	65.5	64.9	66.2	67	64
Linear SVM	59.6	61.9	67.4	69.1	70	75.2
Quadratic SVM	82.4	83.1	85	86.6	86.3	87.3
Cubic SVM	80.5	81.8	82.4	82.4	84	85.3
Ensem ble Subspace Discr iminant	89.6	89.3	89.9	88.9	89.3	90.9
Multilayer Perceptron	80.2	80.9	81.1	83.4	83.4	81.4

Table 2. Results at various cell sizes from 112 to 512

Classifier| Cell Size	112	120	128	256	512
SVM	64.7	62.8	61.9	50.8	39.9
Linear SVM	83.7	84	86	84.7	85
Quadratic SVM	88.3	85	88.3	85.7	87.3
Cubic SVM	86	84	87.6	84.4	84.7
Ensemble Subspace Discriminant	88.6	89.9	91.5	89.6	91.9
Multilayer Perceptron	85.3	85.3	87	85.3	85.3

0	0	0	0	0	0
0	1101110	0	0	01110111	0
0	0	0	0	0	0
0	0	0	0	0	0
0	00010110	0	0	11101111	0
0	0	0	0	0	0

Figure 14. Resultant image LBP shown in binary.

206	**119**
22	**239**

Figure 15. Compact LBP.

In the next phase only Quadratic SVM was used for the hybrid feature which comprised features of both HOG and Compact LBP as shown in Figure 18.

The number of points marked on the neutral images were 14 to 32 at a gap of 2 points for each subject of three databases namely

Bharat Database of Indian Faces (BDI F), The Japanese Female Facial Expression Database(J AF F E) [20] and Karolinska Directed Emotional Faces(K DEF) [21]. J AF F E has posed seven facial expressions of 10 Japanese female models and has a total of 213 images. KDEF has images of 70 amateur actors comprising of 35 males and 35 females. The total number of images in the database is 4900. The subject images have no occlusions in the form of earrings, eyeglasses, moustache, beard and also no visible make-up. It has images taken from five different angles. However only frontal images were used in the experiments.

The points were marked to be in conformity to action units of Facial Action Coding System as described by Paul Ekman. These points covered eyebrows, eyes, nose and mouth region of the images. The tracking of these points was done in the all the emotion images of those subjects.

The accuracy of emotion detection is then calculated:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Figure 16. Sample Compact Lbp extracted from BDIF.

Figure 17. Marked points on sample image.

Support Vector Machines are used in machine learning as supervised learners with algorithms for data analysis for classifi- cation. Training examples are categorized into different categories based on the classification labels. SV M builds model based on the training examples as falling into different categories. After build- ing model, new examples are categorized for different classes. It is non-probabilistic classifier that represents examples as points in space. Mapping is done to categorize the examples of the different categories so that there is a clear distinction between them. After model building the test examples are mapped into the same space. Thereafter prediction is done about their class based on which side of space they come into.

Multilayer Perceptrons maps input data to different output categories based on feedforward artificial neural network. It has directed graph of nodes in multiple layers, the nodes of one layer are connected fully to nodes next layer. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. It has one or more hidden layers that have nonlinear activating nodes.

Navie Bayes provides least percentage of accuracy of the three classifiers used in the experiment. This is particularly true because it is probabilistic classifier with the assumption of independency among features. But in the feature vector used in the experiments, there is a correlation among the various features associated with the points marked. SVM provides high degree of accuracy because of its ability to construct hyperplane between the classes. Multilayer Perceptron also provide high accuracy because of limited number of input nodes, maximum of 34, mapped to total seven output nodes. The results show that SVM and Multilayer perceptron's are best suited for this approach.

Where T P is True positive, T N is true negative, F P is False Positive and F N is False Negative.

The classifiers used for classification are NaÃrve Bayes, Sup- port Vector Machine (SVM) and Neural Network.

Naive Bayes are probabilistic classifiers based on applying Bayes' theorem with assumption of independence between features. It is used for classification of examples as falling into different categories.
Results are shown in Table 3 corresponding to the number of points considered for tracking.Across the databases, the accuracy percentage of KDEF is highest followed by BDIF and JAFFE is at last. KDEF has the highest accuracy as the images of the database.

Table 3. Accuracy of emotion detection in percentge(%)

Database	Classifier	Number of Points										
		14	16	18	20	22	24	26	28	30	32	34
BDIF	Naive bayes	52.1	54.1	50	55.1	56	58.6	61.3	62.8	65.2	66	65.2
	SVM	78.4	84	87.4	89.3	89.8	91.9	91.9	93	93.4	94.5	94.4
	Multilayer Perceptron	84.6	88	90.4	91.2	92.1	92.9	92.9	91.9	92.1	93	92.5
JAFFE	Naive bayes	45.2	46	47.2	47.2	47.6	48.1	54.1	56.9	59	61.8	61.5
	SVM	75.3	80.7	84.2	86	86.4	88.2	87.7	90.1	90.3	92.1	91.9
	Multilayer Perceptron	81.1	84.7	88	88.2	89.3	89.5	89.3	88.2	87.9	89.5	88.2
KDEF	Naive bayes	53.2	54.7	53	56.1	57.4	60.1	62.4	64.6	66.9	67.6	66.7
	SVM	79.1	85.3	88.9	91	91.4	93.1	93.1	94.3	94.6	95.8	95.6
	Multilayer Perceptron	85.8	89.2	92	93.1	93.5	94.1	94.3	93.1	93.5	94.5	94.1

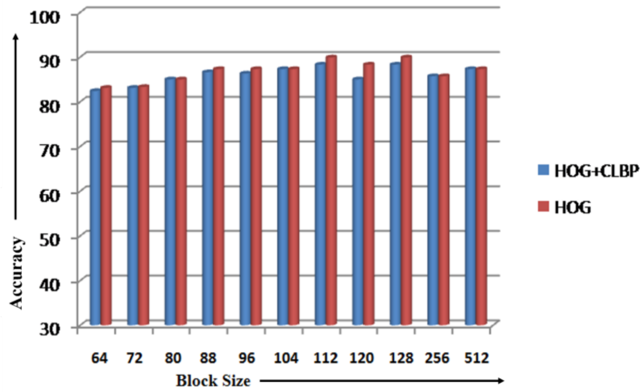

Figure 18. Comparison of result achieve using HOG and HOG+CLBP.

Gathering information from facial expressions may be hindered by the presence of glasses, mustache, beard etc. These act as noisy features in implementation of automatic emotion detection algorithms. However they are a regular feature in the real life situations. Therefore these features were deliberately included in the formation of database. In order to show its effectiveness and applicability quantitative analysis of accuracy is done.The results show that it is robust to gender, occlusions and ethnicity.

Are free from occlusions in the form of earrings, eyeglasses, moustache, beard also no make-up. BDIF stands second because of the presence of occulisons. JAFFE stands last because it is composed of low resolution images. Therefore the calculation of distance vector within the periphery of original point location that has minimum intensity difference is not always that accurate. This results in feature vector that is less accurate towards classification of emotions.

VIII. CONCLUSION

Automatic recognition of emotions from the facial expressions continues to be an important aspect in the field of evolution of new age computing systems. The various features that increase the complexity of emotion recognition systems include ethnicity, gender, pose, occlusion, beard, moustache etc.

The type of database used for learning by systems is of crucial importance. Many databases exist for this purpose but none of them is for posed Indian faces. We bridge this gap by providing Bharat Database which contains facial images of Indian people. The wide variety of subjects and their emotion labeling may help researchers in developing robust algorithms for futuristic artificially intelligent systems. Several evaluations of accuracy were done to behave as a baseline by researchers to develop more.

High degree of accuracy is obtained for all the point tracking for multiple classifications. Extensive experiments were done to show that a window size ranging from 14 to 34 points is sufficient to categorize the emotions. This may be attributed to points conveying the information inline with action units described in physiological studies.

Exhaustive experiments may be conducted involving different models for hybrid vector creation over various other datasets for increasing robustness and accuracy in categorization of emotions.

IX. ACKNOWLEDGMENT

We are grateful to all participants who have provided their time and also helped by ascertaining the advice provided for creation of this database which was a herculean task. Also photography expert and professional artist deserve a great thanks for making this project successful. Last but special thanks to the annotators who provided their time to validate the emotion labels associated with the database.

X. REFERENCES

[1]. Ekman, Paul. "An argument for basic emotions." Cognition & emotion 6, no. 3-4 (1992): 169-200.

[2]. Lucey, Patrick, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression." In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp. 94-101. IEEE, 2010.

[3]. Lyons, Michael J., Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba, and Julien Budynek. "The Japanese female facial expression (JAFFE) database." In Proceedings of third international conference on automatic face and gesture recognition, pp. 14-16. 1998.

[4]. Yin, Lijun, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J. Rosato. "A 3D facial expression database for facial behavior research." In Au- tomatic face and gesture recognition, 2006. FGR 2006. 7th international conference on, pp. 211-216. IEEE, 2006.

[5]. Valstar, Michel, and Maja Pantic. "Induced disgust, happiness and surprise: an addition to the mmi facial expression database." In Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect, p. 65. 2010.

[6]. Sneddon, Ian, Margaret McRorie, Gary McKeown, and Jennifer Hanratty. "The belfast induced natural emotion database." IEEE Transactions on Affective Computing 3, no. 1 (2012): 32-41.

[7]. Happy, S. L., Priyadarshi Patnaik, Aurobinda Routray, and Rajlakshmi Guha. "The Indian Spontaneous Expression Database for Emotion Recog- nition." IEEE Transactions on Affective Computing 8, no. 1 (2017): 131-142.

[8]. Mavadati, S. Mohammad, Mohammad H. Mahoor, Kevin Bartlett, Philip Trinh, and Jeffrey F. Cohn. "Disfa: A spontaneous facial action intensity database." IEEE Transactions on Affective Computing 4, no. 2 (2013): 151-160.

[9]. Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. I-I. IEEE, 2001.

[10]. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 886-893. IEEE, 2005.

[11]. Jolliffe, Ian. Principal component analysis.John Wiley & Sons, Ltd, 2002. 12Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on pattern analysis and machine intelligence 24, no. 7 (2002): 971-987.

[12]. Zhao, Guoying, and Matti Pietikainen. "Dynamic texture recognition using local binary patterns with an application to facial expressions." IEEE transactions on pattern analysis and machine intelligence 29, no. 6 (2007): 915-928.

[13]. Rudovic, Ognjen, Maja Pantic, and Ioannis Patras. "Coupled Gaussian processes for pose-invariant facial expression recognition." IEEE trans- actions on pattern analysis and machine intelligence 35, no. 6 (2013): 1357-1369.

[14]. Tong, Yan, Jixu Chen, and Qiang Ji. "A unified probabilistic framework for spontaneous facial action modeling and understanding." IEEE trans- actions on pattern analysis and machine intelligence 32, no. 2 (2010): 258-273.

[15]. Ojansivu, Ville, and Janne HeikkilAd'. "Blur insensitive texture classi- fication using local phase quantization." In International conference on image and signal processing, pp. 236-243. Springer Berlin Heidelberg, 2008.

[16]. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 886-893. IEEE, 2005.

[17]. Huang, Di, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. "Local binary patterns and its application to facial image analysis: a survey." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 41, no. 6 (2011): 765-781.

[18]. Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on pattern analysis and machine intelligence 24, no. 7 (2002): 971-987.

[19]. Valstar, Michel F., Bihan Jiang, Marc Mehu, Maja Pantic, and Klaus Scherer. "The first facial expression recognition and analysis challenge." In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pp. 921-926. IEEE, 2011.

[20]. Jiang, Bihan, Michel F. Valstar, and Maja Pantic. "Action unit detection using sparse appearance descriptors in space-time video volumes." In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pp. 314-321. IEEE, 2011.

[21]. Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face descrip- tion with local binary patterns: Application to face recognition." IEEE transactions on pattern analysis and machine intelligence 28, no. 12 (2006): 2037-2041.

[22]. Zhao, Guoying, and Matti Pietikainen. "Dynamic texture recognition using local binary patterns with an application to facial expressions." IEEE transactions on pattern analysis and machine intelligence 29, no. 6 (2007): 915-928.

[23]. Taheri, Sima, Qiang Qiu, and Rama Chellappa. "Structure-preserving sparse decomposition for facial expression analysis." IEEE Transactions on Image Processing 23, no. 8 (2014): 3590-3603.

[24]. Shan, Caifeng, Shaogang Gong, and Peter W. McOwan. "Facial expres-sion recognition based on local binary patterns: A comprehensive study." Image and Vision Computing 27, no. 6 (2009): 803-816.

[25]. Lv, Hai-Rong, Zhong-Lin Lin, Wen-Jun Yin, and Jin Dong. "Emotion recognition based on pressure sensor keyboards." In Multimedia and Expo, 2008 IEEE International Conference on, pp. 1089-1092. IEEE, 2008.

[26]. Soleymani, Mohammad, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. "Analysis of EEG signals and facial expressions for continuous emotion detection." IEEE Transactions on Affective Computing 7, no. 1 (2016): 17-28.

[27]. Matlovic, Tomas, Peter Gaspar, Robert Moro, Jakub Simko, and Maria Bielikova. "Emotions detection using facial expressions recognition and EEG." In Semantic and Social Media Adaptation and Personalization (SMAP), 2016 11th International Workshop on, pp. 18-23. IEEE, 2016.

[28]. Kim, Kyung Hwan, Seok Won Bang, and Sang Ryong Kim. "Emotion recognition system using short-term monitoring of physiological signals." Medical and biological engineering and computing 42, no. 3 (2004): 419- 427.

[29]. Boiten, Frans A., Nico H. Frijda, and Cornelis JE Wientjes. "Emotions and respiratory patterns: review and critical analysis."

International Journal of Psychophysiology 17, no. 2 (1994): 103-128.

[30]. Glowinski, Donald, Antonio Camurri, Gualtiero Volpe, Nele Dael, and Klaus Scherer. "Technique for automatic emotion recognition by body ges- ture analysis." In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, pp. 1-6. IEEE, 2008.

[31]. Castellano, Ginevra, Santiago D. Villalba, and Antonio Camurri. "Recog- nising human emotions from body movement and gesture dynamics." In International Conference on Affective Computing and Intelligent Interaction, pp. 71-82. Springer, Berlin, Heidelberg, 2007.

[32]. Tian, Yingli, Takeo Kanade, and Jeffrey F. Cohn. "Facial expression recognition." In Handbook of face recognition, pp. 487-519. Springer London, 2011.

[33]. Bowyer, Kevin W., Kyong Chang, and Patrick Flynn. "A survey of ap- proaches and challenges in 3D and multi-modal 3D+ 2D face recognition." Computer vision and image understanding 101, no. 1 (2006): 1-15.

[34]. Jeni, LAa˛szlAs¸ A., Jeffrey F. Cohn, and Takeo Kanade. "Dense 3D face alignment from 2D video for real-time use." Image and Vision Computing 58 (2017): 13-24.

[35]. Kamarol, Siti Khairuni Amalina, Mohamed Hisham Jaward, Heikki KAd'lviAd'inen, Jussi Parkkinen, and Rajendran Parthiban. "Joint facial expression recognition and intensity estimation based on weighted votes of image sequences." Pattern Recognition Letters92 (2017): 25-32.

[36]. Valstar, Michel F., Enrique SAa˛nchez-Lozano, Jeffrey F. Cohn, Laszlo A. Jeni, Jeffrey M. Girard, Zheng Zhang, Lijun Yin, and Maja Pantic. " FERA 2017-Addressing Head Pose in the Third Facial Expression Recognition and Analysis Challenge". arXiv preprint arXiv:1702.0417 (2017).

[37]. Battocchi, Alberto, Fabio Pianesi, and Dina Goren-Bar. "A first eval- uation study of a database of kinetic facial expressions (dafex)." In Proceedings of the 7th international conference on Multimodal interfaces, pp. 214-221. ACM, 2005.

[38]. Baron-Cohen, Simon. "The Eyes as Window to the Mind." The American journal of psychiatry 174, no. 1 (2017): 1-2.

[39]. Shen, Xiaolu, Xuetao Feng, Jungbae Kim, Hui Zhang, Youngkyoo Hwang, and Ji-yeun Kim. "An extensible framework for facial motion tracking." In Consumer Electronics (ICCE), 2013 IEEE International Conference on, pp. 288-291. IEEE, 2013.

[40]. Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. I-I. IEEE, 2001.

[41]. Bashyal, Shishir, and Ganesh K. Venayagamoorthy. "Recognition of facial expressions using Gabor wavelets and learning vector quantization." Engineering Applications of Artificial Intelligence 21, no. 7 (2008): 1056-1064.

[42]. Chen, Ying, Chunjian Hua, and Ruilin Bai. "Sequentially adaptive active appearance model with regression-based online reference appearance template." Journal of Visual Communication and Image Representation 35 (2016): 198-208.

[43]. Wang, Jun, Lijun Yin, Xiaozhou Wei, and Yi Sun. "3D facial expression recognition based on primitive surface feature distribution." In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, pp. 1399-1406. IEEE, 2006.

[44]. Lorincz, Andras, Laszlo Jeni, Zoltan Szabo, Jeffrey Cohn, and Takeo Kanade. "Emotional expression classification using time-series kernels." In Proceedings of the IEEE Conference on computer vision and pattern recognition workshops, pp. 889-895. 2013.
[45]. Sariyanidi, Evangelos, Hatice Gunes, and Andrea Cavallaro. "Automatic analysis of facial affect: A survey of registration, representation, and recognition." IEEE transactions on pattern analysis and machine intelli- gence 37, no. 6 (2015): 1113-1133.
[46]. Liu, Jingjing, Bo Liu, Shaoting Zhang, Fei Yang, Peng Yang, Dimitris N. Metaxas, and Carol Neidle. "Non-manual grammatical marker recognition based on multi-scale, spatio-temporal analysis of head pose and facial expressions." Image and Vision Computing 32, no. 10 (2014): 671-681.