Emotion Recognition from Videos Using Facial Expressions

P. Tamil Selvi, P. Vyshnavi, R. Jagadish,
Shravan Srikumar and S. Veni

Abstract In recent days, automatic emotion detection is a field of interest and is used in fields such as e-learning, robotic applications, human–computer interaction (HCI), surveillance, ATM monitoring, mood-based playlists/YouTube videos, psychological studies, medical fields like supporting blind and dumb people, for treating autism in children, entertainment, animation, etc., The proposed work describes detection of human emotions from a real-time video or image with the help of classification technique. The major part of human communication constitutes of facial expression, which is around 55% of the total communicated information. The basic facial expressions that are considered by the psychologists are: happiness, sadness, anger, fear, surprise, disgust, and neutral. The proposed work aims to classify a given video into one of the above emotions using efficient facial features extraction techniques and SVM classifier. The author's contribution is to increase the efficiency in emotion recognition by implementing the above mentioned superior feature extraction and classification methods.

Keywords Appearance model · Emotion recognition · Feature extraction · Gabor filter · MATLAB · Occlusions · SVM classifier

1 Introduction

The proposed work utilizes human facial expressions as the features to detect different human emotions. The "Enterprise05" video database is chosen for the purpose which consists of 44 subjects who utter five sentences each for six different emotions, which are happiness, sadness, anger, fear, surprise, and disgust. The

P. Tamil Selvi (✉) · P. Vyshnavi · R. Jagadish · S. Srikumar · S. Veni
Department of Electronics and Communication Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore 641112, India
e-mail: tamiltheone67@gmail.com

S. Veni
e-mail: s_veni@cb.amrita.edu

S.S. Dash et al. (eds.), *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, Advances in Intelligent Systems and Computing 517,
DOI 10.1007/978-981-10-3174-8_47

videos from the database are converted into frames (i.e., images). Preprocessing is an important process before feature extraction for better results. As part of preprocessing, the images are resized and normalized for further process. The preprocessed images' faces are separated from its background using Viola–Jones face detection algorithm. Two types of features are calculated in this work, which are geometric-based features and texture-based features. The Gabor filter technique is used for texture-based feature extraction from the images and the 63 fiducial point detection method is used for geometric-based feature extraction. The extracted features are given as the input to the classifier. The classification is done by SVM classifier algorithm for exact emotion recognition.

2 Related Work

The process of locating the region of interest and extracting the features play a major role in determining the efficiency of the classifier. The choice of the features to be extracted also affects the efficiency of the classifier. The features that can be chosen are geometric features and texture features. Priya Sisodia et al. [1] have used more than one features increases the efficiency of the classifier. However, the experiment was carried out with texture-based features using Gabor filters. The extracted texture features are given to a SVM classifier. They have concluded that Gabor filter not only gives a better performance under noise and intensity differences in the image but also gives efficiency when compared all other traditional methods.

Li Zhang and Ming Jiang [2] have used facial action coding system for emotion recognition in humanoid robots. A combination of two artificial neural networks is used to locate the action units (AUs) and SVM-based classifiers are used to detect the emotions shown in real time. The two artificial neural networks detect upper and lower facial AUs. Six AUs were extracted and given as inputs to the SVM classifier which detects the emotion of a person.

Happy and Aurobinda Routray [3] have stated other methods like PCA (principle component analysis) method followed by LDA (linear discriminant analysis) technique for feature extraction. And also have implemented Harris detection method for corner point detection, which is also one of the efficient methods for extracting features other than the method employed in this work, i.e., 63 fiducial facial point detection. Harris corner point detection and 63 fiducial facial point detection methods are explained in Sect. 6. This proposed work proves that the 63 fiducial point method is more efficient than the Harris corner detection method.

3 Methodology

The author implements the recognition process by first deducing the frames or images by sampling the videos that are needed to be analyzed. The number of frames is dependent on the duration of the video and it is also notable that high frequency of the sampling increased the efficiency of recognition but in turn increases the complexity of processing. Later, these images are to be individually taken for detecting the faces in each image. The face detection algorithm adopted in this proposed work is Viola–Jones algorithm. The emotions in the face are evidently concentrated in the mouth and the eye region. Thus, mouth and eye detections are also done using Viola–Jones algorithm itself. The detected mouth and eye images from the face are used to extract the texture features from them.

This paper imposes a double degree feature extraction process, i.e., two types of features are extracted from the images namely, texture feature extraction and geometric feature extraction. The Gabor filters are used for extracting texture based features from the image and 63 fiducial facial point detection methods for geometric-based feature extraction. The all extracted features are tabled together in a.csv file (excel file). Further, these features are combined and carried as a single excel sheet to the classifier to train the system so as to predict the test images properly. The classifier used in here in SVM classifier (Fig. 1).

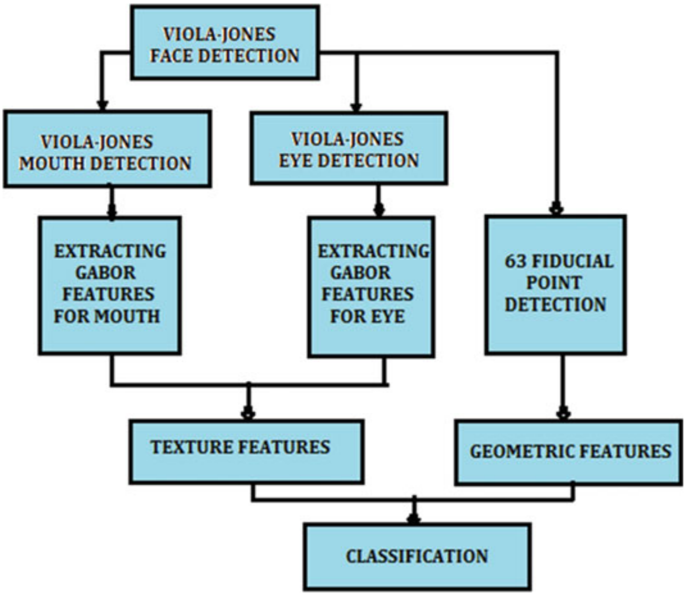

VIOLA-JONES FACE DETECTION
VIOLA-JONES MOUTH DETECTION
VIOLA-JONES EYE DETECTION
EXTRACTING GABOR FEATURES FOR MOUTH
EXTRACTING GABOR FEATURES FOR EYE
63 FIDUCIAL POINT DETECTION
TEXTURE FEATURES
GEOMETRIC FEATURES
CLASSIFICATION

Fig. 1 Block diagram of proposed system for recognition

4 Face Detection

Viola–Jones algorithm was proposed by Paul Viola and Michael Jones in the year 2001. Although this algorithm is used for face detection purposes it can also be used as a classifier where other objects can also be detected once after training. Viola–Jones is a very robust framework designed to work in a real-time system. It has a very high detection rate and a very low false rate making it an ideal framework for face detection in image processing applications. Viola–Jones algorithm is used to detect human faces from an image for the proposed method. This step helps in eliminating the face from its background. But one setback in this system is that the face needs to be upfront with no tilt in the face and facing the camera directly.

The Viola–Jones algorithm has four main steps:

- Haar Feature Selection
- Integral Image Creation
- Adaboost Training
- Cascading Classifiers

The basic concept behind this framework is the addition of pixel intensities in predefined windows and comparison of these added intensities with the adjacent windows added intensities. Also the image needs to be converted to gray scale before proceeding to the Viola–Jones algorithm.

4.1 Haar Feature Selection

Even though all humans have distinctive identities there are a few features which can be considered as common to all humans, for example the region is a bit lighter than its adjacent eye regions. The Haar features are just a set of fixed windows which consist of black and white rectangular regions, which are run through the entire image. Then the difference between the sums of pixel intensities of the image region under the white region is subtracted from that of the black region. This difference is used to find if a facial feature is present over that window or not by comparing threshold values provided (Figs. 2 and 3).

For example, for detecting the eyes, we employ the fact that the upper cheeks are lighter than the eyes and hence we construct the feature window as shown in Figs. 4 and 5.

Similarly, the mouth region is darker than the upper and lower lip regions, so mouth can be detected. The basic formula used in this algorithm is

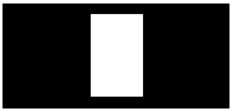


Fig. 2 Nose detection using the Haar feature

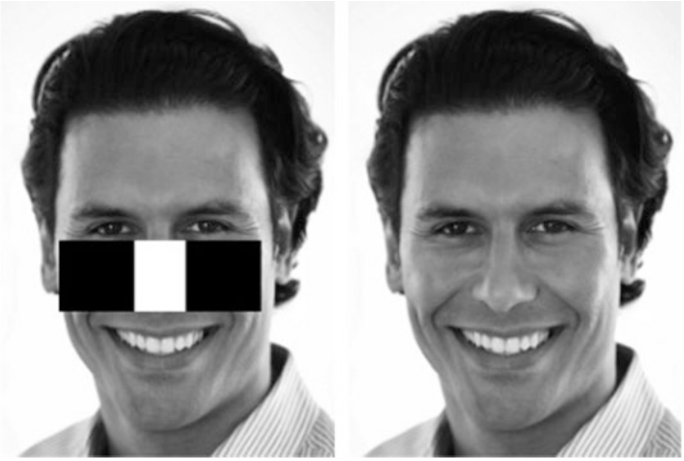


Fig. 3 Nose Haar feature window over the image

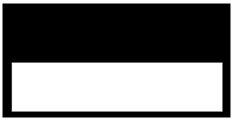


Fig. 4 Eyes detection using the Haar feature

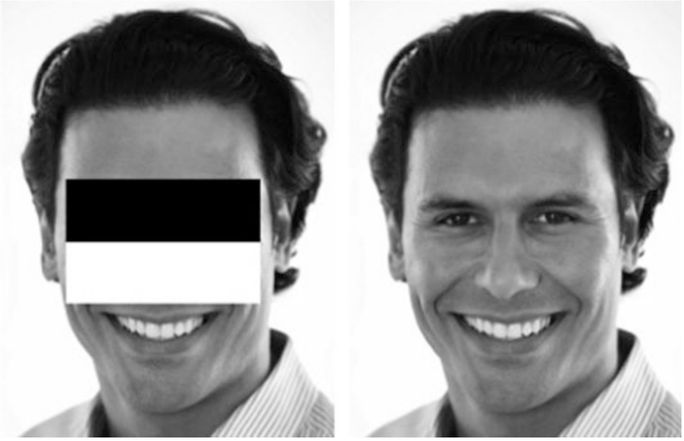


Fig. 5 Eyes Haar feature window over the image

$$\text{Value } V = \Sigma(\text{pixels in black area}) - \Sigma(\text{pixels in white area}) \tag{1}$$

Size of white or black regions in a Haar window is also considered while evaluating ‘V’. Weights with $\omega_i > 0$ can be used for adjusting the contributing sums

$$V = \omega 1 \cdot \mathrm{SW1} + \omega 2 \cdot \mathrm{SW2} - \omega 3 \cdot \mathrm{SB} \tag{2}$$

Weights ω_i need to be specified when defining a Haar wavelet.

4.2 Internal Image Creation

An area table that is the summation of values in a rectangular subset of a grid is a data structure and algorithm done in a quick and efficient manner. The value at any point (x, y) in the summed area table is just the sum of all the pixels above and to the left of (x, y), inclusive as given below.

$${}^{I}\sum{}_{(x,y)=}\sum_{\substack{x' \leq x \\ y, \leq y}} {}^{i(x',y')} \tag{3}$$

Once the Haar feature is selected, they are convoluted with the given face image and the results are stored as many subimages and given to Adaboost training for training the system.

4.3 Adaboost Training (Adaptive Training)

Adaboost basically means ‘adaptive boosting’ where a number of weak algorithm in a linear manner in order to form a strong algorithm. The best part is that it can be used to detect objects other than faces too. It just depends on the classification algorithm that you give. Adaboost training is done by combining the results of the “weak learners” into a weighted sum that is given as the output of the boosted classifier.

4.4 Cascade Classifiers

The value found after the application of the Haar feature windows is given to the cascade classifiers. The classifier then classifies the image as a face or non-face. In the case of eyes and mouth, it classifies the convoluted images as eyes or not eyes

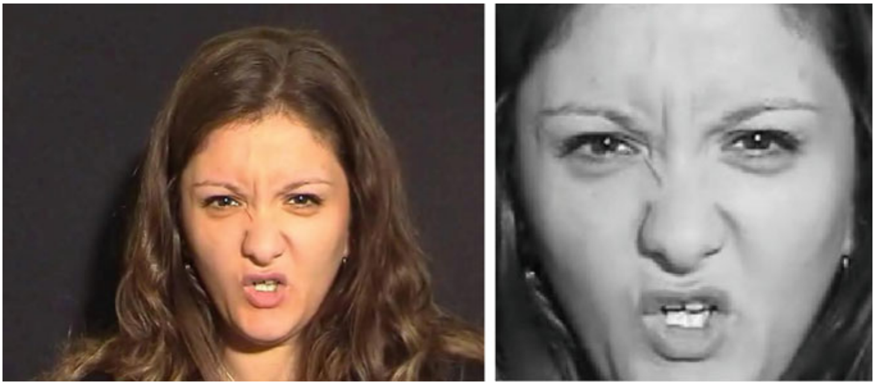


Fig. 6 Input and output of face detection

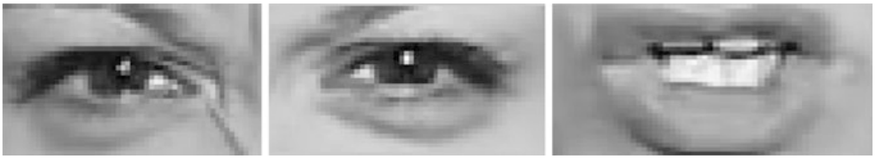


Fig. 7 Output of left eye, right eye, mouth detection

and nose or not nose depending on the "value" calculated in the above step. The classified region is then cropped out from the original image (Figs. 6 and 7).

5 Texture-Based Feature Extraction

The technique used for geometric-based feature extraction is Gabor filter method. The Gabor filter is usually the multiplication-convolution of a sinusoid wave with Gaussian function. Here, as Gabor filter helps in extracting useful data from the images.

$$G(x,y) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{(x^2+y^2)}{2}} \mathrm{e}^{-i(2\pi f(x\cos(\theta)+y\sin(\theta)))} \tag{4}$$

where f represents the frequency of the texture, that is, requires θ is the orientation that can be varied for accessing multiple directions, σ can be varies to change the size of the region of the image that is analyzed.

This advantage of differently available orientations and frequencies makes this method efficient for texture feature extraction. The images with faces along different

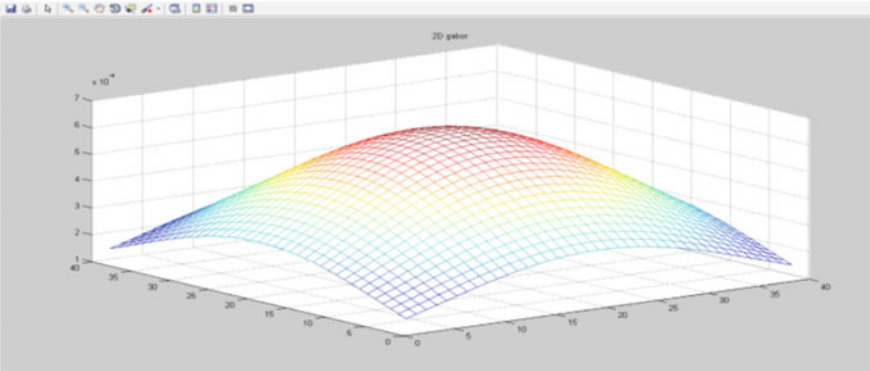


Fig. 8 2D-Gabor Filter Bank Magnitude plot

directions and masked or occlusion face features can also be extracted efficiently (Fig. 8).

6 Geometric-Based Feature Extraction

6.1 Harris Corner Detection

Harris corner detector is a mathematical operator used in computer vision systems and image processing in order to extract some very useful features. The basic working of this system can be described as the comparison of intensity values in a given small window of an image. First, a small window is to be selected with respect to the size of the image. Every window has a window function with the 'x' axis and 'y' axis parameters as its input. The window function can either be a Gaussian function or a constant function, but in either case the function value must be positive only over the window and must have a zero value over all other values of 'x' and 'y'. Now a comparison of intensities is done in the given window using the mathematical Eq. (5).

$$E(u,v) = \sum w(x,y)[I(x+u,y+v) - I(x,y)]^2 \tag{5}$$

We can understand from the above equation that the difference between the shifted intensity and the intensity of the current pixel will be very less or tending to zero for a window that is over a region with no corner point or an edge, but will be very high otherwise. So we need to only consider the regions over which the window gives very high values of $E(u, v)$. Now the Eigen values of M are taken into consideration in order to differentiate a corner point from an edge.

$$\lambda_1 \approx 0 \quad \text{and} \quad \lambda_2 \approx 0 \Rightarrow \text{no features of interest.} \tag{6}$$

$$\lambda_1 \approx 0 \quad \text{and} \quad \lambda_2 = \text{large positive number} \Rightarrow \text{edge} \tag{7}$$

$$\lambda_1 \quad \text{and} \quad \lambda_2 = \text{large positive values} \Rightarrow \text{corner} \tag{8}$$

Equation (6) helps to find non-face images. Equation (7) draws all edges in the given image. Equation (8) marks all corner points in the processed image. However, not all non-faces, edges, and corner are detected efficiently by this method. Hence, this paper employs the high-efficient corner point detection named 63 fiducial facial point detection.

6.2 63 Fiducial Facial Point Detection

The technique is used for face recognition, pose estimation, and also to trace the corner points on the face. This method is said to be efficient because it helps in spotting maximum number of corner points on face which are countered on areas of the face that are salient for showing distinct features for extraction. Thus, the technique makes this system real time in nature. This method begins with the generation of initial graphs for all training images, one graph for each orientation, here 13 different orientations are used for every face in the database. The set of 13 oriented faces for a single image is called a facial bunch graph (FBG). Once the system has plotted graphs for different pose of face in training set, graphs for new test images need not be generated separately as they can be simply drawn automatically by elastic graph matching technique, which helps in adjusting the existing train image graphs elastically to match the test image characteristics.

The graphs are constructed like: At first fiducial points are marked on the given image depending on the image profile. The number of points marked for use as features in this project is 63 and these are numbered automatically for further use as shown in Fig. 10. Then the relation between the points on the face is realized to construct a general shape model or graph of the face which is shown as the red line in Fig. 9. In the graphs shown below, the red line drawn on the face is actually the geometric appearance when all the facial landmarks/points are connected. There are 63 such points that are detected and marked as shown in Fig. 11. Once these 63 points are marked on the given face image their coordinates are noted. But, only optimal number of points are itself sufficient for classification. For the purpose, only few important facial landmarks are taken into consideration like eye corner points, mouth corner points, eyebrow corner points, and the tip of the nose. The distance between these corner points is measured and these measurements are the extracted features from the images, Likewise six feature distance values are taken as features. The distance between any of these corner points are calculated using Euclidian distance formula.

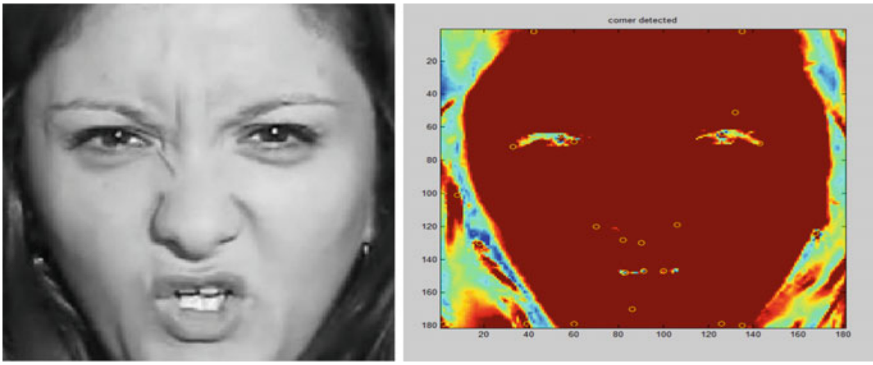


Fig. 9 Results for Harris corner detection

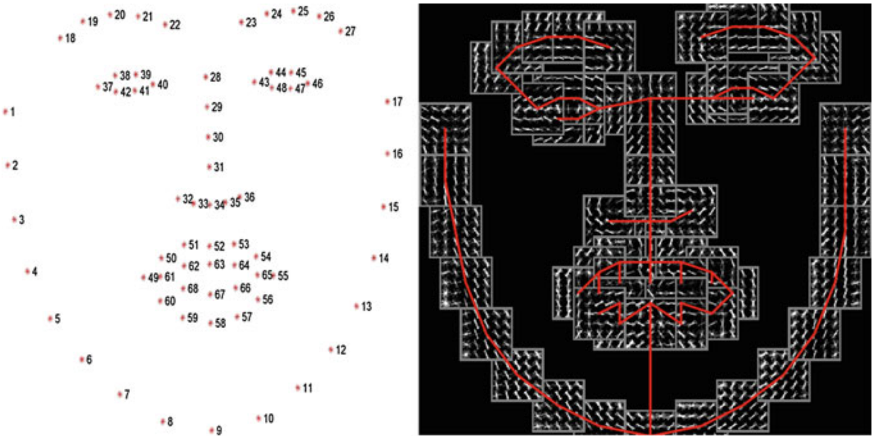


Fig. 10 Fiducial point graph model fitting

7 Emotion Classifications

The classifier used for the proposed work is SVM classifiers. SVM classifier is a nonprobabilistic linear classifier. Along with “kernel trick” it can even perform non-linear classification efficiently. Mapping a problem into a space with a high or large dimensions makes it more likely that the problem will become linearly separable. Classification involves the training of the SVM classifier based on the extracted features from the training samples and then testing the classifier using test samples. This task involves six emotions, i.e., it requires a multiclass SVM classifier. In SVM (choosing one-against-all approach), they are plotted as points in a graph and decision boundaries between different classes are obtained. From the whole database, 80% of the images are taken as training images and 20% of the

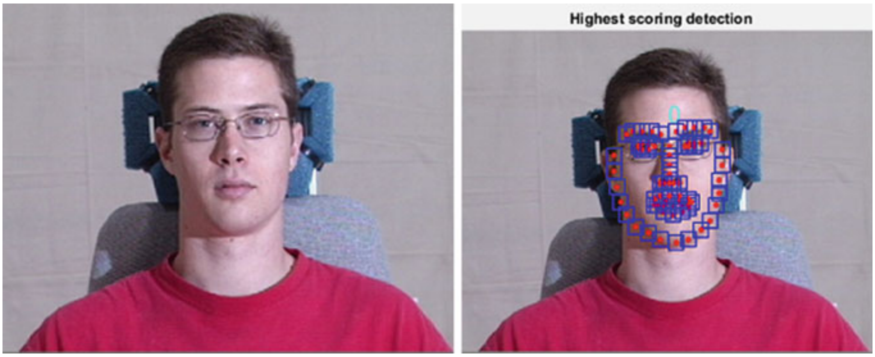
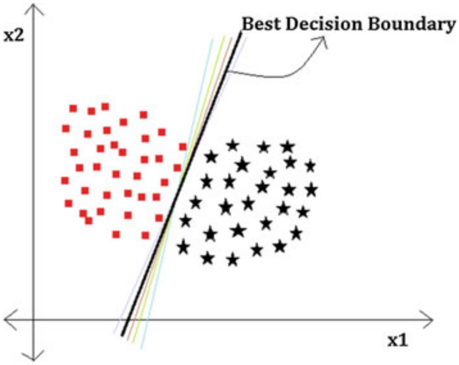


Fig. 11 Fiducial point method results

Fig. 12 Best decision boundary decision



images are taken as testing. These testing and training images are given to the classifier. The features are first marked in feature space hence each sample is represented by a point in the feature space. SVM classifier builds a decision boundary between two classes in the feature space. Later, the new test data plotted is classified depending upon its location among the boundaries of the graph. During the training phase, the equation of the best decision boundary hyperplane is found. The decision hyperplane is constructed by maximizing the margin between two class feature points (Fig. 12).

8 Conclusion

The Gabor and fiducial points combine feature extraction-classification and make the paper more efficient than its previous methods. Both Gabor and fiducial points are promising techniques in texture feature and geometric feature, respectively.

Thus, this high profile method improves the overall efficiency of the output. Also, this paper can be implemented easily in numerous applications as they possess less complexity in the processing phase making it simple for the classifier. The proposed work produces accuracy of 87% after classification.

Acknowledgment We would like to thank our class adviser Mr. Gandhiraj R for guiding us through-out the project. We also thank the university and lab staff for providing us with the software and all other necessities. We would like to thank our friends and college department for supporting us for completing the project successfully.

References

1. Priya Sisodia, Akhilesh Verma, Sachin Kansal, "Human Facial Expression Recognition using Gabor Filter Bank with Minimum Number of Feature Vectors," *International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 5 – No. 9, July 2013.*
2. Li Zhang and Ming Jiang, "Intelligent Facial Action and Emotion Recognition for Humanoid Robots," *International Joint Conference on Neural Networks (IJCNN) July 6–11, 2014.*
3. S L Happy and Aurobinda Routray, "Automatic Facial Expression Recognition Using Features of Salient Facial Patches," *IEEE transactions on affective computing, vol. 6, no. 1, January-March 2015.*
4. Thushara S and S Veni, "A Multimodal Emotion Recognition System from Video," *ICCPCT Conference, 2016.*
5. Ligang Zhang and Dian Tjondronegoro, "Facial Expression Recognition Using Facial Movement Features," IEEE *transactions on affective computing, Vol. 2, no. 4, October–December 2011.*
6. K. Sreenivasa Rao and Shashidhar G. Koolagudi, "Recognition of emotions from video using acoustic and facialfeatures," *Springer-Verlag London 2013.*
7. https://en.wikipedia.org/wiki/Viola%E2%80%93Jones_object_detection_framework.
8. https://en.wikipedia.org/wiki/Gabor_filter.