Face Detection and Facial Expression Recognition System

Anagha S. Dhavalikar*
**PG Student, Electronics & Telecommunication Department, Vivekanand Institute of Technology, Mumbai, India*
anaghashandilya@yahoo.co.in

Dr. R. K. Kulkarni**
***Professor, Electronics & Telecommunication Department, Vivekanand Institute of Technology, Mumbai, India*
rk1_2002@yahoo.com

Abstract— **A human-computer interaction system for an automatic face recognition or facial expression recognition has attracted increasing attention from researchers in psychology, computer science, linguistics, neuroscience, and related disciplines. In this paper, an Automatic Facial Expression Recognition System (AFERS) has been proposed. The proposed method has three stages: (a) face detection, (b) feature extraction and (c) facial expression recognition. The first phase of face detection involves skin color detection using YCbCr color model, lighting compensation for getting uniformity on face and morphological operations for retaining the required face portion. The output of the first phase is used for extracting facial features like eyes, nose, and mouth using AAM (Active Appearance Model) method. The third stage, automatic facial expression recognition, involves simple Euclidean Distance method. In this method, the Euclidean distance between the feature points of the training images and that of the query image is compared. Based on minimum Euclidean distance, output image expression is decided. True recognition rate for this method is around 90% - 95%. Further modification of this method is done using Artificial Neuro-Fuzzy Inference System (ANFIS). This non-linear recognition system gives recognition rate of around 100% which is acceptable compared to other methods.**

Keywords- **Face detection, Feature extraction, AAM, Expression recognition, Euclidean Distance, ANFIS**

I. INTRODUCTION

Face plays an important role in social communication. Face biometric itself is used in many applications like security, forensic and other commercial applications. Similarly facial expressions are the fastest means of communication, while conveying any type of information. In 1978, Ekman and Frisen [1] reported that, Happy, Sad, Anger, Fear, Disgust and Surprise are the six basic expressions which are readily recognized across very different cultures. A system designed for analyzing facial actions automatically through a human-computer interaction, is called Automatic Facial Expression Recognition System (AFERS). The robust AFER system [2] can be applied in many areas of science such as emotion detection, clinical psychology and pain assessment. There are three major steps in an AFERS;

1. To detect the face from the given input image or video,
2. To extract the facial features like eyes, nose, mouth from the detected face and
3. To classify the facial expressions into different classes like Happy, Angry, Sad, Fear, Disgust, and Surprise.

Face detection is a special case of object detection. In the proposed system, face detection is implemented using skin color detection and segmentation. Also it involves lighting compensation algorithm and morphological operations to retain the face from the input image. To extract the facial features, Active Appearance Model i.e. AAM method is used. Finally, the expressions are recognized as Happy, Sad, Anger, Fear, Disgust, and Surprise, initially by using simple Euclidean Distance method and then by training the Artificial Neuro-Fuzzy Inference System (ANFIS).

In the next part of the paper, section II explains first stage of the proposed system, i.e. face detection using skin color detection, lighting compensation algorithm and morphological operations. Section III gives Feature extraction using AAM method. Section IV deals with expression recognition using Euclidean Distance method and trained ANFIS. Finally Section V gives future scope and Section VI concludes the paper.

II. FACE DETECTION

A. Skin Color Detection

Skin color can be considered as a good feature for detecting human face. Since color allows fast processing and is highly robust to geometric variations of face pattern, skin color has proven to be a useful and robust cue for face detection, localization and tracking. Image content filtering, content aware video compression and image color balancing applications can also benefit from automatic detection of skin in the images.

A different class of color spaces is the orthogonal color spaces used in TV transmission. This includes YUV, YIQ, and YCbCr. YIQ is used in NTSC TV broadcasting while YCbCr is used in JPEG image compression and MPEG video compression. In the YCbCr color space [3], Y component represents the luminance information; Cr component represents the red chrominance information; Cb component represents the blue chrominance information. One advantage of using these color spaces is that most video media are already encoded using these color spaces. Transforming from RGB into any of these spaces is a straight forward linear transformation as given below:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.00 \\ 112.00 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

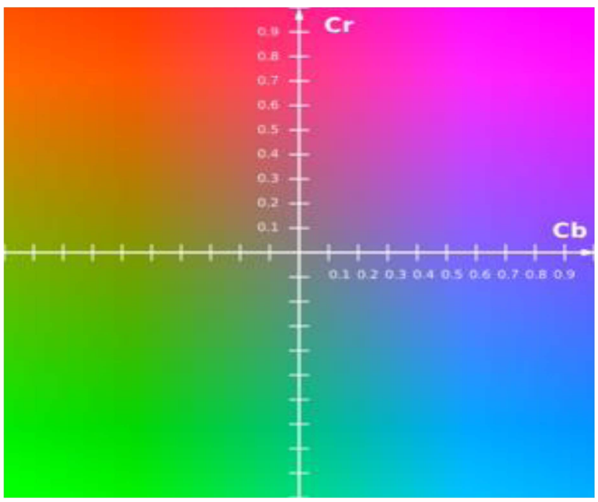


Fig. 1 YCbCr Color model [8]

All these color spaces separate the illumination channel (Y) from two orthogonal chrominance channels (UV, IQ, CbCr). Therefore, unlike RGB, the location of the skin color in the chrominance channels will not be affected by changing the intensity of the illumination. In the chrominance channels the skin color is typically located as a compact cluster with an elliptical shape. This facilitates building skin detectors that are invariant to illumination intensity and that use simple classifiers.

Figure 2 shows the histograms of different color models for two different images. Histogram of an image represents the relative frequency of occurrence of the various gray levels in an image. From the results obtained, it is found that the gray scale distribution for Cb & Cr lies within a certain range of pixel values for any type of skin. This is due to the fact that the Cb & Cr values are independent of illumination of light. Due to this property of YCbCr color space, it is preferred for skin color detection. Since RGB components are subject to the lighting conditions, the face detection may fail if the lighting condition changes. Among many color spaces, this work use YCbCr components since it is one of existing Matlab functions thus would save the computation time.

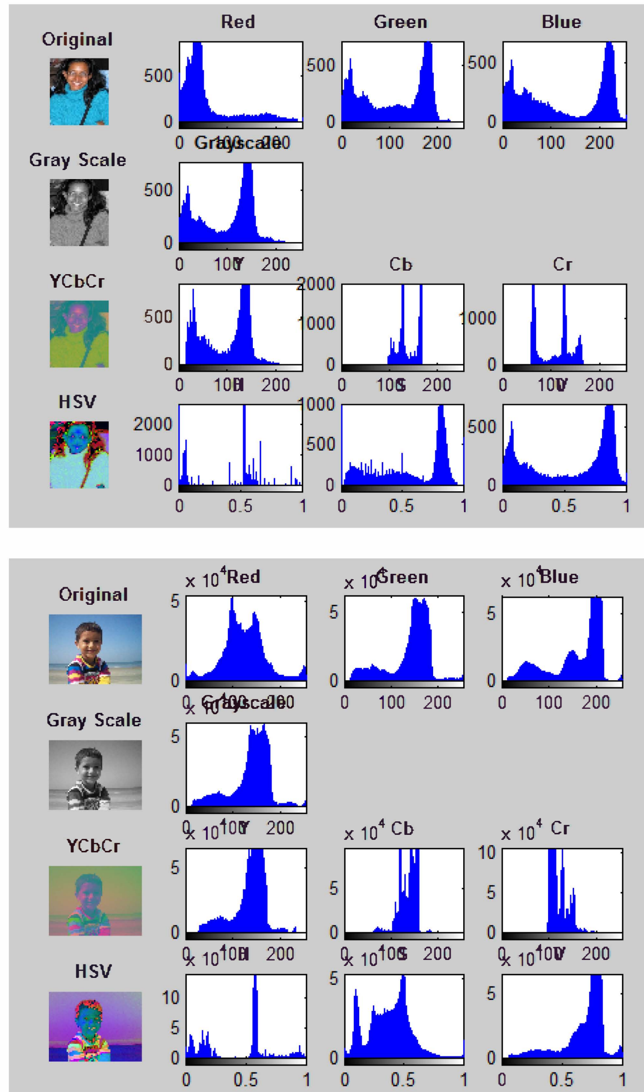


Fig. 2 Histograms of two different images

B. Lighting Compensation

Many times there are large variations in the real colors of skin in an image, since the skin color is often affected by light in the image. So to make color correction in color images, lighting compensation algorithm [4] is used. This method can be described by the following set of equations.

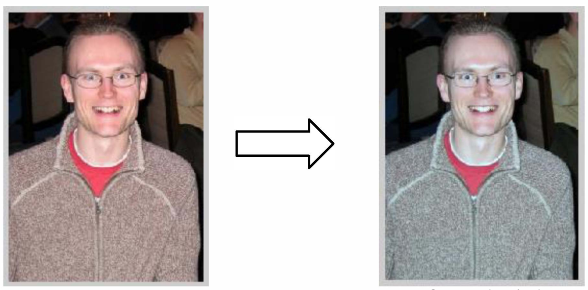
$$R' = R * \left[\frac{K}{R_{average}}\right] \tag{1}$$

$$G' = G * \left[\frac{K}{G_{average}}\right] \tag{2}$$

$$B' = B * \left[\frac{K}{B_{average}}\right] \quad (3)$$

$$K = \frac{(R_{average} + G_{average} + B_{average})}{3} \quad (4)$$

The R, G and B are the amount of stimulus of red, green and blue respectively in the recorded scenery. $R_{average}$, $G_{average}$ & $B_{average}$ are the average of each color channel. After applying this lighting compensation algorithm on the image, the results can be obtained as shown in figure 3.



Original Image Image after Color balance

Fig. 3 Original images and the result of color balance

C. Face Detection using Morpho logical operations

The Morphological operations [4] are used to transform the signals into simpler ones by removing irrelevant information. These operations can reserve essential shape features of an image. Dilation and Erosion are the most basic morphological operations used in image processing. Dilation adds pixels to the boundaries of objects in an image, and erosion removes pixels on object boundaries. Firstly, erosion function is used to get rid of some small pieces, compared with face area, which is unwanted fragment. After that step, dilation operation will help to recover face area. This procedure can be done several times to get good result. Due to the fact that the pixel value variations of other skin-like regions such as hands or legs are smaller than those of face regions because of the presence of features with different brightness, all face region candidates with pixel value variations smaller than a threshold are removed. In order to improve the detection speed and achieve high robustness, the symmetry of the face is checked and all the candidates are removed when the symmetry is verified but no facial features are detected. The effect of Morphological operation is shown in figure 4.

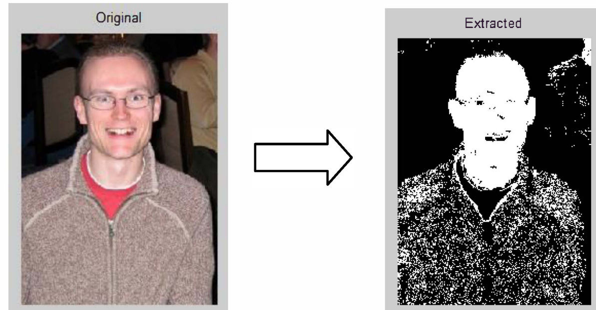


Fig. 4 Result of morphological operation

Thus the pixels on face are filled with white pixels and the other regions in the image are filled with black pixels. The binary image received after a morphological operation is used further for face detection. As shown in figure 5, the face region in the image is detected which is used further for extracting the facial features.

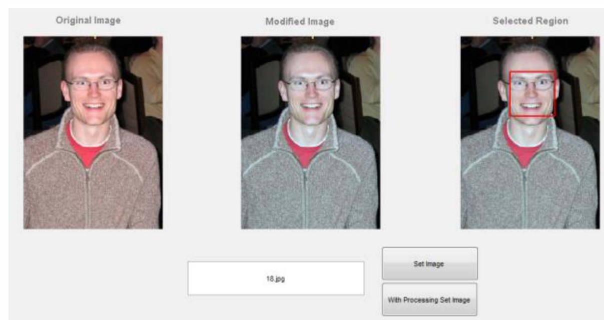


Fig. 5 GUI for face detection

III. FEATURE EXTRACTION USING AAM METHOD

Image-based methods have been applied in many areas of facial computing. One of the most successful recent techniques, which, incorporates both shape and texture information from facial images is the Active Appearance Model (AAM) method. It was developed initially by Cootes and Taylor [5] and has shown strong potential in a variety of facial recognition technologies. As shown in figure 6, the face image is cropped from the detected face. The cropped image is then used for feature extraction using AAM method.

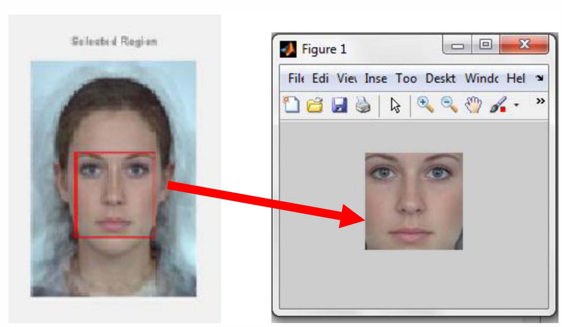


Fig. 6 Cropped images taken for Feature Extraction

The points on the facial features like eyes, eyebrows and mouth are located as shown in figure 7. Initially the Active Shape model is created for a neutral image in the database. It automatically creates a Data File which gives the information about model points located on the detected face. Then the sequence of different expressions like Happy, Sad, Anger, Fear, Disgust, and Surprise is given as a video input starting with the neutral expression. The change in the AAM shape model according to the change in facial expressions measures the distance or the difference between Neutral and other facial expressions.

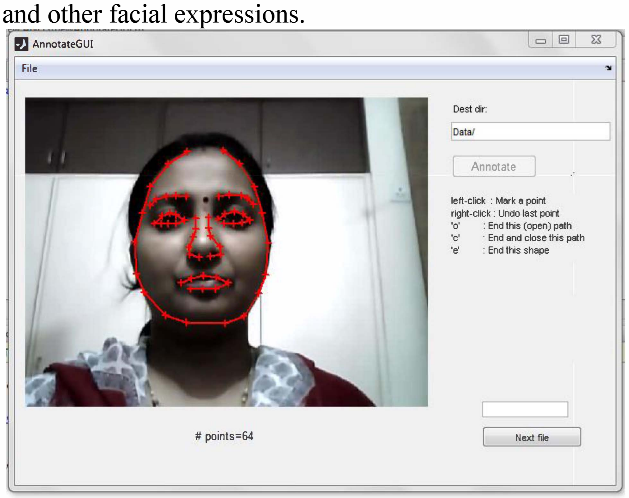


Fig. 7 GUI for locating feature points from the detected face

IV. FACIAL EXPRESSION CLASSIFICATION

A. Expression Recognition using Euclidean Distance method

In this method, the database consists of training data sets and testing data sets of images. For one particular subject, the training and testing data sets consist of images of different expressions like Neutral, Happy, Sad, Anger, Fear, Disgust, and Surprise. Using AAM method, the points on facial features are located for all these images and stored in the form of data file. The data file consists of the relative x-y co-ordinates of those located points. When any test image is given as an input, the system finds the Euclidean Distance between the points on the test image and the points on each training image. If (x1, y1) and (x2, y2) are the co-ordinates of points on training and testing images respectively, then the Euclidean Distance is calculated by the following formula [6]:

$$\text{E. D.} = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \qquad (5)$$

The training image for which, the Euclidean Distance of the test image is minimum, is given as the output image expression. As shown in figures 8 and 9, Figure-1-GUI shows the testing image and Figure-2-GUI shows the training image output. As shown in Figure-1-GUI, red color indicates located points on the test image and blue color indicates the points on different training images. Euclidean Distance for all these points is calculated and where it finds the maximum match, is given as the final output expression as shown in figure-2- GUI. Similarly figure 9 gives the output for Surprise expression. Euclidean distance method is an easy method, useful for static images and requires a large amount of manual work. However, change in the distance from camera affects the recognition rate in this method. True recognition rate for this method is around 90% - 95%.

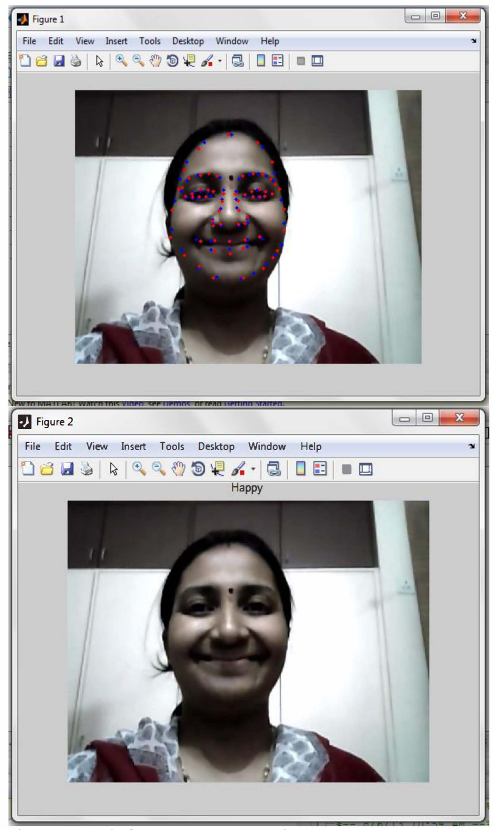


Fig. 8 Result for Happy expression

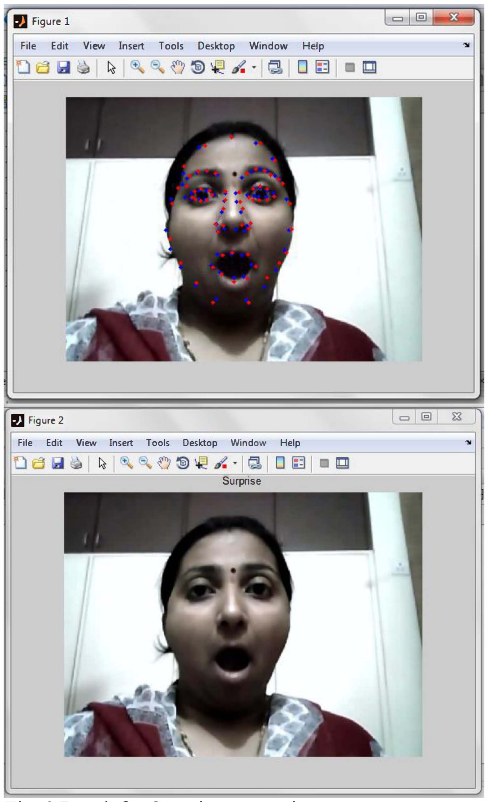
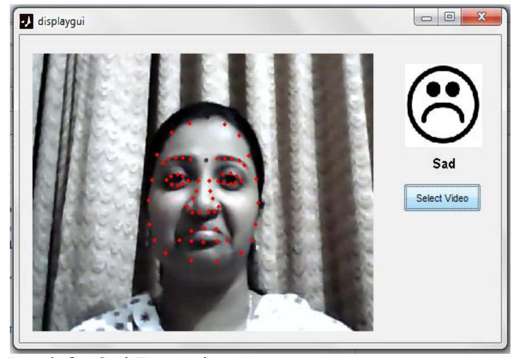


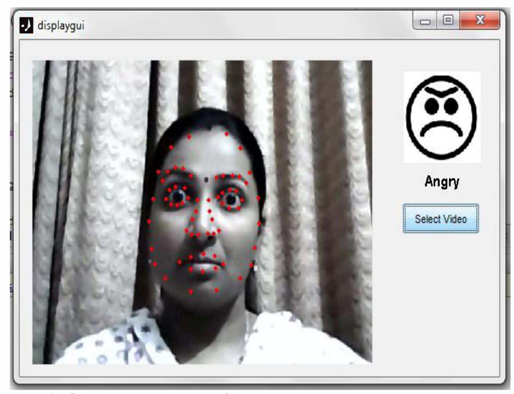
Fig. 9 Result for Surprise expression

B. Expression recognition using ANFIS

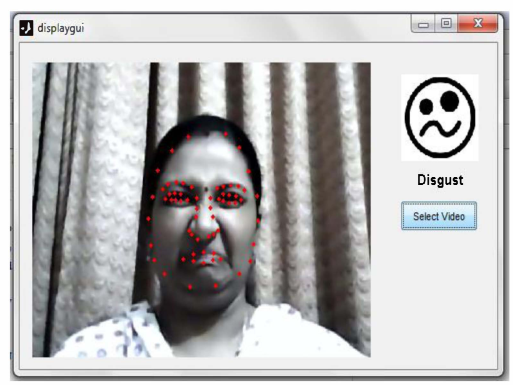
To improve the recognition rate of the system, further modification in the third phase is done using Artificial Neuro-Fuzzy Inference System (ANFIS). In this method, the static images as well as video input can be given for testing the expressions. Here, neuro-fuzzy [7] based automatic facial expression recognition system to recognize the human facial expressions like happy, fear, sad, angry, disgust and surprise has been proposed. Initially a video showing different expressions is framed into different images. Then the sequence of selected images is stored in a database folder. Using AAM method, the features of all the images are located & stored in the form of **.ASF** files. Then a mean shape is created for all the images in data folder. The change in the AAM shape model according to the change in facial expressions measures the distance or the difference between Neutral and other facial expressions. These values are stored in a **.MAT** file & a specific value is assigned for each individual expression for training the ANFIS. These difference values are then given as input to the ANFIS (Artificial Neuro-Fuzzy Inference System). Using the ANFIS tool available in Mat lab, the system is trained for the different images and their video input sequences for different expressions. Using ANFIS the recognition rate can be achieved up to 100 %. Figure 10 shows the different expressions detected by the system.



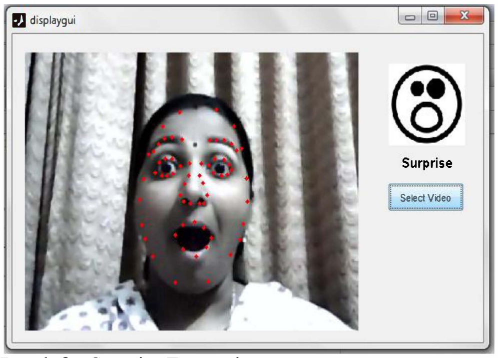
Result for Sad Expression



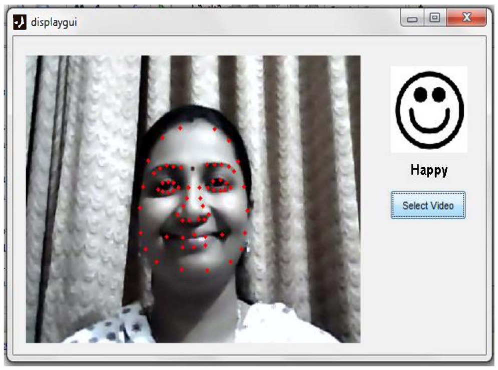
Result for Angry Expression



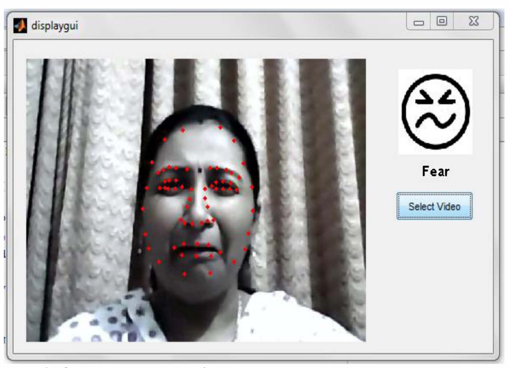
Result for Disgust Expression



Result for Surprise Expression



Result for Happy Expression



Result for Fear Expression

Fig. 10 Results for different facial expressions

V. FUTURE SCOPE

The proposed work can be further extended by increasing the number of different expressions other than the six universal expressions (anger, fear, disgust, joy, surprise, sadness). The classification of other facial expressions may require the extraction and tracing of additional facial points and corresponding features. The system can be improved by using a wider training set so as to cover a wider range of poses and cases of low quality of images.

VI. CONCLUSION

An automated Facial Expression Recognition System has a wide range of applications in psychological research and human-computer interaction applications. The system plays a communicative role in interpersonal relations because they can reveal the affective state, cognitive activity, personality, intention and psychological state of a person. The proposed system consists of three modules. The face detection module is based on image segmentation technique where the given image is converted into a binary image and further used for face detection. Among different color models like RGB, YCbCr, HSI etc., YCbCr is convenient for the robust images. The first stage of face detection has been tried on 105 image samples and acquired correct detection for 95 image samples. Based on the test results, the false face detection occurred under circumstances were the image qualities is low or face size is below 32x32. AAM (Active Appearance Model) method incorporates both shape and texture information from facial images, so it is found to be convenient for feature extraction. For expression recognition, Euclidean distance method is useful for static images and requires a large amount of manual work. The method gives a recognition rate of 90 - 95 %. Due to its ambiguity for real time or robust images, ANFIS (Artificial Neuro-Fuzzy Inference System) has been used as a further improvement. In this system, the static images as well as video can be given as input and tested for the different expressions. Also the system can work accurately for person-independent database. The accuracy of facial expression recognition varies with number of training samples. The system gives the recognition rate close to 100% for large number of training samples. A neuro-fuzzy approach for facial expression recognition is applicable for real time applications such as human emotion analysis, human-computer interaction, surveillance and online conferencing and for entertainments.

References

[1] P. Ekman, T. Huang, T. Sejnowski, J. Hager, "Final Report to NSF of the Planning Workshop on Facial Expression Understanding", 1992.

[2] Ewa Piatkowska "Facial Expression Recognition System", Master's Thesis Technical Report.

[3] Kai-biao ge, jing wen, bin fang, "Adaboost Algorithm based on MB-LBP features with skin Color segmentation for face detection" , Proceedings of the 2011 International Conference

on wavelet analysis and pattern recognition, guilin, 10-13 july, 2011.

[4] Kamarul Hawari, Bin Ghazali, Jie Ma, Rui Xiao, "An Innovative Face Detection based on Skin Color Segmentation", International Journal of Computer Applications (0975-8887), *Volume 34– No.2, November 2011.*

[5] G.J. Edwards, T.F. Cootes, and C.J. Taylor, "Face reccognition using active appearance models", Proceedings of the European Conference on Computer Vision, 1998.

[6] Aliaa A. A. Youssif and Wesam A. A. Asker, "Automatic Facial Expression Recognition System Based on Geometric and Appearance Features", Computer & Information Journal, 2011, Volume 4, Issue 2, Page 115.

[7] V. Gomathi, Dr. K. Ramar, and A. Santhiyaku Jeevakumar, "A Neuro Fuzzy approach for Facial Expression Recognition using LBP Histograms", International Journal of Computer Theory and Engineering, Vol. 2, No. 2 April, 2010.

[8] http://en.wikipedia.org. Google web page