International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015 56
ISSN 2229-5518

3D Human-Computer Intelligent Interaction by Facial Expression analysis using Facial Action Coding System-FACS

K.S.Jayareka[1], S.RajaMohamed[2]

Abstract— Facial expressions are the natural way to show emotions of human creature. A knowledge based emotion recognition approach for human and computer intelligent interaction based on tree structure can be developed. Decision Tree (DTA) and Sequential Tree Approaches (STA) are examined in the Facial Expression analysis process. From which, computer is trained to recognize and reproduce the six basic emotional states such as Surprise, Anger, Fear, Happy, Sad and Disgust using Facial Action Coding System (FACS). It is based on the enumeration of all Action Units of a face that causes minute facial movements. Facial Expression Recognition Process (FERP) involves both Facial Parameter Extraction and Facial Parameter Analysis. With the FERP, Human-Computer Intelligent Interaction (HCII) aims at providing natural ways for computers to talk, encourage, console, support, share and calm humans during their different states of emotions. In Neural Network, Feed Forward Back Propagation and hierarchical based clustering algorithm is used on each node of these trees to improve the efficiency and robustness of Facial Expression Analysis.

Index Terms— Clustering Algorithm, Decision tree, Facial expression, FACS, HCII, Neural Network and Sequential tree.

1 INTRODUCTION:

Facial Expression analysis is the emerging research work where the computer is trained to recognize, reproduce or respond to human expressions and emotions. Emotions can be expressed through facial expression, gestures and voice. The most expressive way to show emotion is through facial expression. FACS is Facial Action Coding System is a system to taxonomize human facial movements by their appearance on the face, based on a system originally developed by a Swedish anatomist named Carl-Herman. It was later adopted by Paul Ekman and Wallace V. Friesen, and published in 1978 [2]. FERP is the way of communicating with the machine. Application area is like behavioral sciences, security, animation, decision making, human computer interaction such as communication with robot etc.

1.1 Existing System

The existing systems to analysis the facial expression analysis include Dictionary based approaches where Real image are not considered only 6 basic emotions are experimented [1]. In Dynamic Appearance Descriptor Approach to Facial Actions Temporal Modeling they used UNBC-McMaster pain database, the SAL database, the GEMEP-FERA, Cohn-Kanade, SEMAINE databases where Pose Variation is not available [5].

- *Author name is K.S.Jayareka currently pursuing masters degree program in computer science engineering in Kalaignar Karunanidhi Institute of Technology, Anna University, Coimbatore, TamilNadu, India, PH-8489271823. E-mail:jayarekasubramanian@gmail.com*
- *Co-Author name is Prof.S.RajaMohamed currently working as Associate Professor in computer science engineering Dept in Kalaignar Karunanidhi Institute of Technology, Anna University, Coimbatore, TamilNadu, India, PH-7200474754. E-mail:rajadelip@gmail.com*

Autism Respond to Facial Expressions in Virtual Reality Environments for Adolescents, Non verbal communication is used with ASD victim and the results often shows Contempt and disgust , fear and surprise were often confusing[6][10]. In Human Aging Facial Expression Analysis the Lifespan FER can deliver accuracy higher that human perception but drawbacks are manual labeling of facial fiducial points and the automatic detection is not possible [7]. For Deformable 3D based face model there exist High recognition rate when compared with other Isomap Methods, realistic emotional expression is detected with high recognition rate but the deformation is made using manually and not automatically done [8].

1.2 Proposed Work:

A. Contribution:

- We learned to discover 3D modeled face from the human face.
- We learned how to do the facial parameter extraction and facial parameter analysis.
- We also discussed how the facial features are measured using Euclidean distance between two or more white pixels.
- How does connected component analysis works in a group of pixels near the Facial features.
- We also incorporated the Sequential tree approach and Decision tree approach.
- The Result of this observation is calculated on the weight basis that can be done using Neural

Network and hierarchical clustering algorithm.

B. Outline:

The content of this paper deals with the following sections. Section 2 discusses the 3D conversion of Human face. Then in Section 3 we present the Facial Parameter Extraction and Facial Parameter Analysis the Interpolation technique and the Mean distance calculation. Decision tree approach and Sequential tree approach is analyzed in the Section 4 and in Section 5 Experimental Result is checked in the Feed Forward Neural Network and hierarchical based Clustering Algorithm. Fig 1 Represents the Architecture diagram of the 3D HCI by Facial Expression analysis using FACS

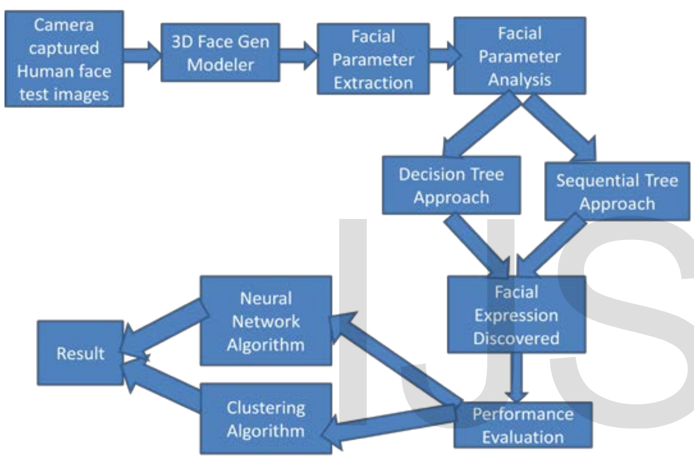


Fig 1 Represents the Architecture diagram of the 3D HCI by Facial Expression analysis using FACS

2 FACIAL EXPRESSION ANALYSIS IN 3D ANIMATED FACE OF HUMAN:

The picture of a 3D human face is demonstrated here, the human face has main features as eyes, eyebrows, nose, mouth containing lips and teeth but so many attributes like hair, skin texture, moles, pimple and other different cases like acid attacked face, wounded face, scar, split mouth and splint eyes or any other defects are may also present in a normal human face. In such cases the face expression cannot be analyzed clearly or it may lead to miscalculation of facial parameter extraction. For example a aged face contains so many wrinkles in the skin and the skin texture is also not like a young face thus difficulty occurs in analyzing such varied face. For male gender they naturally have beard and mustache in their face which automatically hides the facial feature thus it leads to erroneous results. In order to overcome such difficulties and errors, a 3D modeled face is taken as the test image for our observation. 3D modeled face is obtained by taking the face picture in three views that is front view, left side view and right side view. The three viewed pictures of human face is fed into the face generator modeler tool [10]which is available in the internet and the feature points are marked as per instructions, finally 3D modeled face is obtained as shown in the figure 1. Next is the Facial Expressions of the 3D modeled face, this can achieved by considering the FACS Action units by P.Ekman and W.V.Friesen in 1971 [2]. FACS is based on the enumeration of all Action Units of a face that cause facial movements.

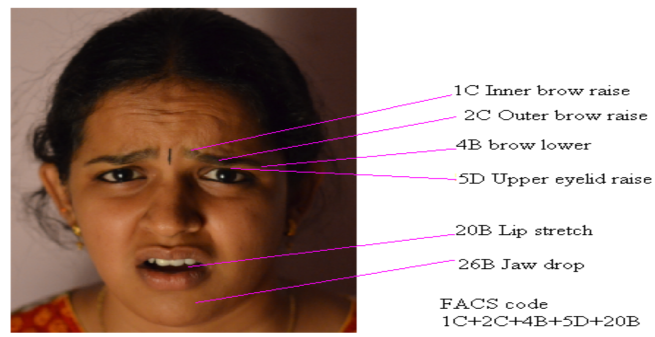


Fig 2 Represents FACS Action Unit of a Human Face with Anger Expression

The combination of these Action Units results in a large set of possible facial expression. With this FACS Action Units the expressions can be made in the same Face gen tool [4]. Fig 2 Represents FACS Action Unit of a Human Face with Anger. Expression Remaining expressions are the mixture of these six basic expressions in human.

The first step to be performed in expression recognition is localization of facial feature points whose position and deformation are the characteristics of each basic expression. Each emotion has its own characteristic features, which distinguishes from others. By extensively going through the datasets some of the inferences are, the emotion *'Surprise'* has a wide open eyes with mouth opened. Eyebrow found quiet relaxed for the emotion *'Happy'*. Therefore, an accurate extraction of contours of these kinds of facial features would enable us to automatically recognize these expressions [3]. Fig 3 Represents Detail Design of the Facial Expression Analysis.

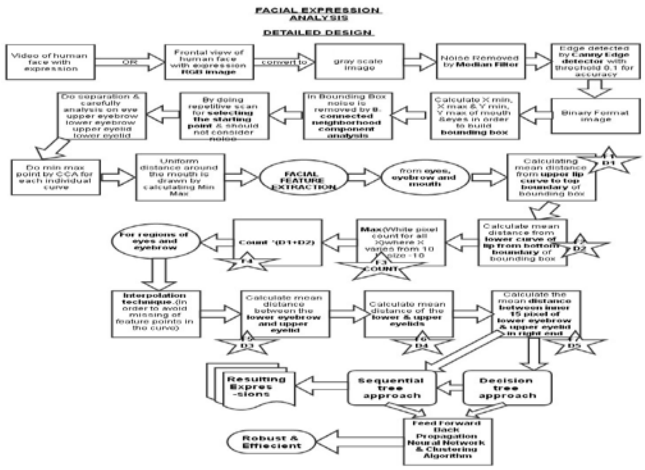


Fig 3 Represents Detail Design of the Facial Expression Analysis

3 Facial Feature Extraction

The eyes and mouth are the main features we are going to extract in this approach where there are various approaches exist in the research field.

3.1 Feature extractions need sequence of steps.

Those steps are called as pre-processing steps as follows

- Conversion of image to appropriate format
- Filtering and Edge detection
- Connected component analysis
- Bounding box generation and Centering

3.1.1 Conversion of image to appropriate format

The input image is first converted into the appropriate gray scale image in order to find the contour as shown in fig 4. Contour detection is possible when the input image is converted into grayscale or y-cb-cr images. Figure 2 represents the grayscale image of the 3D human face.

Fig 4 Grayscale image of the 3D human face.

3.1.2 Filtering and Edge detection In filtering and edge detection process the gray scale converted image is filtered in order to remove the noise that exist in any normal image as shown in the fig 5. We used Median filter in order to remove impulse noise. Edge detection is possible by using canny edge detector or sobel edge detector for getting the proper contour of edges thus features of the human face can be easily extracted.

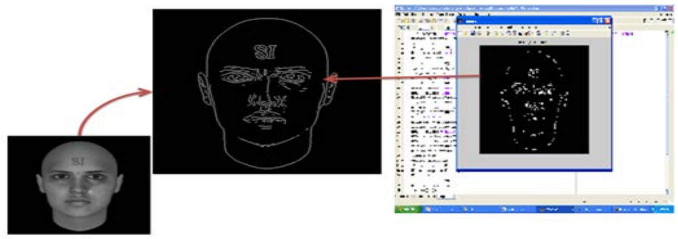


Fig 5 Filtered and Edge Detected Image

Fig 6 Six expressions emoted by a 3D human mouth region Disgust, Anger, Fear, sad, Surprise, Happy.

Fig 7 curves in regions of eyes and mouth

3.1.3 Connected Component Analysis

From the previous result connected component analysis is done on each pixel starting from the first pixel of the gray scale image and if any pixel is found in the gray scale image the connected component analysis is performed on each next 10 pixels. In case if no next pixel is found then the pixel is determined as noise. While considering the connected component analysis a special care is taken on the four separated segmented parts or curves i.e. upper eyebrow, lower eyebrow, upper eyelid, lower eyelid individually instead of taking whole part of the eyes as shown in the fig 7.

3.1.4 Bounding Box Generation and Centering

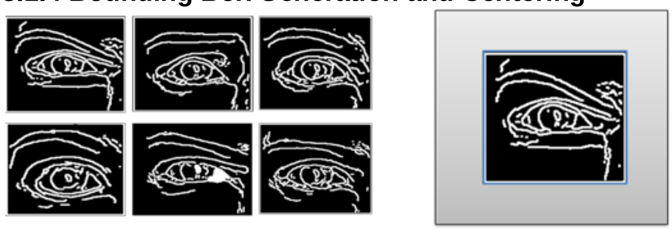


Fig 8 Bounding box and centering.

To bring uniformity in the mouth region we draw a bounding box. Bounding box is created by the max and min points of the mouth region by calculating Xmin an Ymin and X min max and Y max as shown in the fig.8. Fixed constant size bounding box is generated in order to have consistency on the mouth regions of various input images.

ISSN 2229-5518

The centre obtained from the drawn bounding box is directly mapped to the centre of the feature of eyes and mouth and the difference in the distance is applied to all the other pixels in the same regions of the bounding box as shown in the below figure.

3.2 Facial Features Extraction

The feature set consists of features extracted from the regions of eyes, eyebrows and mouth.

3.2.1 Feature set extraction from regions of Mouth

Table 1: Feature set extracted from regions of mouth

Feature	Parameter	Description
F1	D1	Mean of the distances of the upper curve of lip from top boundary
F2	D2	Mean of distances of the lower curve of lip from bottom boundary
F3	COUNT	Max[count of white pixels for all X] where X varies from 10 to size(image)-10
F4	F4	COUNT* (D1 + D2)

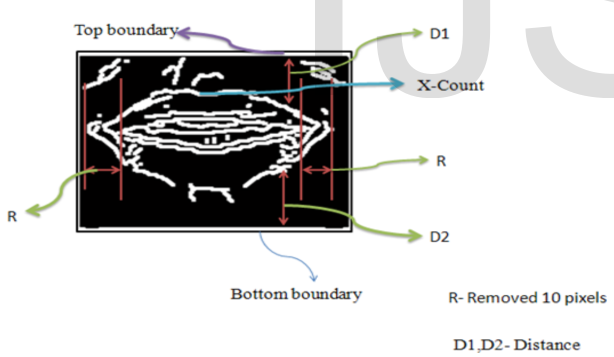


Fig 8 Feature set extracted from regions of mouth depicting table 1.

Feature Extraction is efficiently performed by the removing 10 pixels on either side of the contour as in fig in order to avoid miscalculation of the count value and eliminating noise components at the ends. Consistent checking of noise occurrence is examined to get the actual feature and noise elimination. Consistency of noise is checked by non continuous pixels. Feature 1 deal with the measure of distance from the top boundary to the upper curve. Feature 2 deals with the measure of distance from the bottom boundary to the lower curve. The count value indicates the maximum value of number of white pixels for each of the X - coordinate. Feature 4 is obtained through aggregation of F1, F2, and F3.

3.2.2 Feature set extracted from regions of eyes and eyebrows

Before the feature set extraction of eyes and eyebrows, interpolation of all the four curves should be performed.

3.2.2.1 Interpolation

The extraction of features is based on the coordinate values of the image. So, it must be made sure that there are no missing points in each curve, which will affect in feature extraction. In order to achieve this, interpolation technique is performed. The image consists of four lines (Upper and lower eyebrow, Upper and lower eyelid), which are processed one at a time before interpolation. The process of interpolation involves finding the missing values of 'y' for a given 'x'.

Table 2: Feature set extracted from regions of eyes

Feature	Parameter	Description
F5	D3	Mean of distances between the curves of lower eyebrow & upper eyelid
F6	D4	Mean of distances between the curves of lower & upper eyelids
F7	D5	Mean of distances between inner 15 pixels of lower eyebrow & upper eyelid taken from the right end

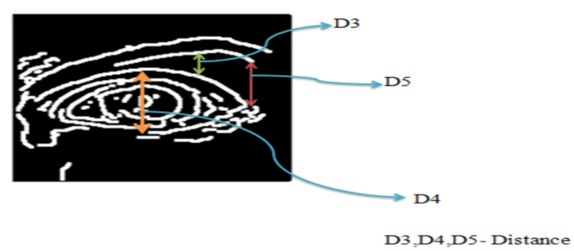


Fig 9 Feature set extracted from regions of eyes depicting table 2.

Feature 5 is the mean of the distances of all the points between lower curve of eyebrow and the upper curve of eyelid. D4 accounts to the mean of the distances of 7 pixels on either sides of the center of the eyelids as shown in fig 9. D5 is clearly shown in fig 9 that takes only a few pixels of innermost curves. This is because the eyebrow movements for various emotions are changing prominently at these pixels. The features are calculated for the expressive and the normal expression. The input to the algorithm is the deviation calculated between these expressions. The deviation for all the features are observed and named the same.

ISSN 2229-5518

3.3 Decision tree approach

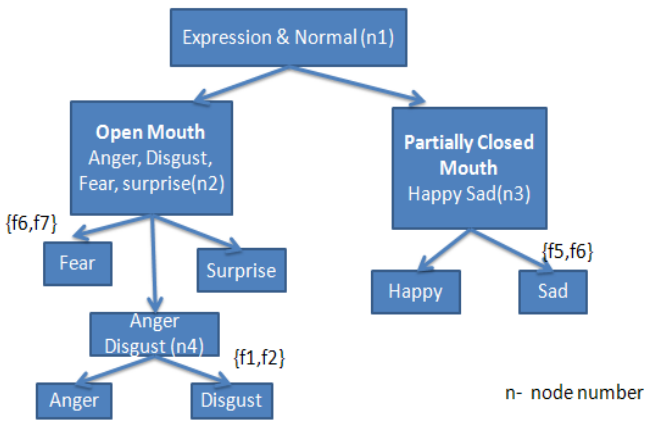


Fig 10 Decision tree approach

The feature set containing the deviation parameters are given as input to the decision tree. On extensive observation it was found that emotions *surprise*, *fear*, and *disgust* do have their mouth regions widened. The other emotions get classified to closed mouth expressions using the features F1, F2, F3, and F4. Next level finds the eyebrows clearly distinguishing the *happy* and *sad* expressions. With the features F6 F7 fear and surprise vary in the distance between the eyes and eyebrows as shown in fig 10. Still *anger* and *disgust* come under the same cluster, which gets classified through the features F1 F2. Here the upper curve of lip of anger expression does have a uniform distance from the top boundary, where as *disgust* has no uniformity in the distance.

3.4 Sequential tree approach

The Sequential tree approach focuses on expulsion of one expression at a time from the whole lot of expressions. A class of features that is prevalent in distinguishing the emotion is considered. Since eyeballs are so prominent in *surprise* it is eliminated in the first level. The eyebrows get raised higher than normal in the expressions *fear* & *surprise* but since *surprise* is classified in the first level, *Fear* can be easily classified. From the remaining expressions eyeballs are found quiet big for the expression *happy* when compared to other expressions *sad, anger* and *disgust* as shown in fig.8. Expression sad is classified in next level since it is the only closed mouth expression. The same set of features (F1 & F2) used in decision tree approach classifies *anger* and *disgust* as shown in fig 11.

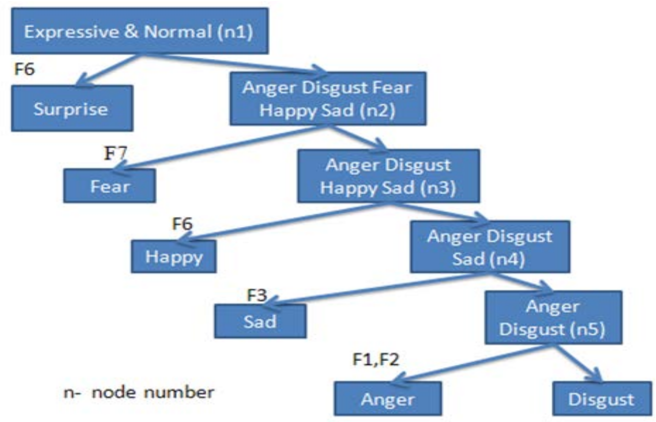
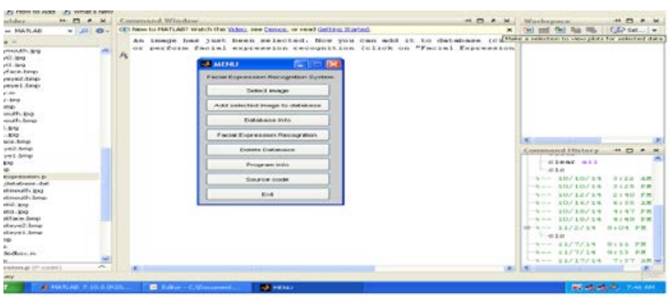


Fig 11 Sequential tree approach

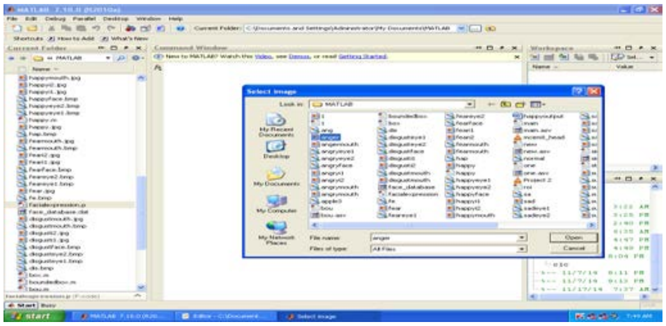
4 EXPERIMENTAL RESULTS

We have evaluated our technique on various images containing the frontal view of the human face using the image processing toolbox and the neural network toolbox of the MATLAB 13a samples of each emotion were taken. The most distinguishing feature points, which determine the subtle variations among the different emotions, were selected. After refining the feature set (F3), the classification into open and closed mouth categories was greater than 92%. Separate feature sets were selected for better accuracy for the next level of the tree. The overall classification rate was greater than 85% for the decision tree approach and 89 % for sequential tree approach. For emotion recognition each node in the decision and sequential trees was considered to be a classifier like neural networks or clustering algorithm. In decision tree approach clustering algorithm is used to group the samples based on the six emotions whereas in sequential tree approach, one expression is pulled out of the whole set based on the best possible features.

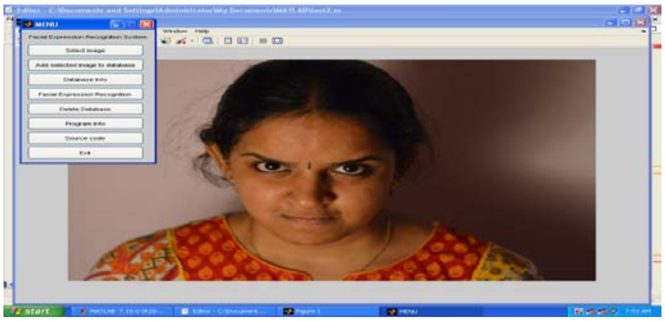
Fig 12 Sample input and outputs



A) Represents input image selection

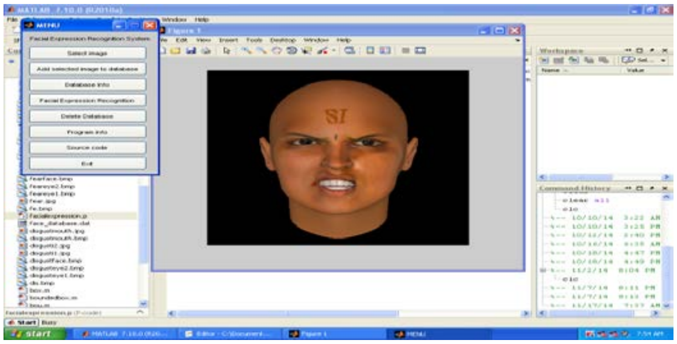


B) Represents searching the Expression

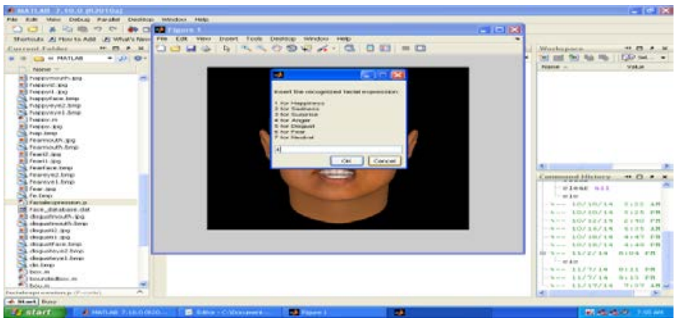


C) Represents Anger image of Human face as the searched image

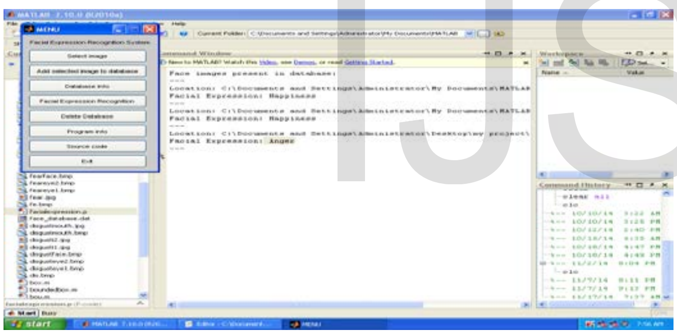
ISSN 2229-5518



D) Represents 3D Face from human input



E) Represents the processing of the expression by FACS and Tree approaches



F) Represents the output showing the expression as anger.

To get robust results neural network approach is used. Since there are different feature sets at different levels for the classification of emotion, so a neural network was trained at every node in the tree. We selected Feed Forward Back Propagation Neural network due to its simplicity and efficiency. The number of hidden layers in the network and the number of neurons in each layer was chosen by trial and error method based on the performance function until it reaches the specified goal.

Failure analysis:

The feature set extracted for the different approaches is considered to be best distinguishable features between various expressions. This obviously reduces the load on the Neural Network, but cannot be considered as the standalone criteria for classification. When considering the two approaches the sequential tree approach has a cutting edge over the decision tree approach, because when considering expressions such as fear and surprise the eyeballs are very similar but in the second approach the surprise is already eliminated. The classification process is much improved in approach two, but due to the difference in expressions generated by wide range of people, it leads to the misclassification to a small extent. The other example is anger and disgust classification where the eyes and eyebrows are very similar.

5 Comparison with existing system

3D Human-Computer Intelligent Interaction using Facial Expression and Analysis in Image Processing has used the 3D modeled faces which are highly accurate in calculating the Facial Features when compared to normal image of a human face. Pose variation for facial expression analysis is rare in existing system which can be included in 3D model. In order to increase the recognition rate few features from other sources like voice, gestures can be extracted

6 Conclusion and Future works

Computer intelligence in order to recognize and to reproduce emotions can be performed by facial expression, gestures and voice recognition. In this system facial expression from a 3D human model is analyzed. Future work is to analyze different pose variation and voice expression analysis and performance evaluation.

References

[1] Sima Taheri, Quiang Qiu, Rama Chellappa "Structure Preserving Sparse Decomposition for Facial Expression Analysis"IEEE Transactions on Image Processing, 2014.

[2] P.Ekman and W.Friesen, "The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Santa Clara,CA,USA:Consulting Psychologists press",1978.

[3] P.Nagabhushan, R.Pradeep Kumar, G.Siva Rama Krishnan, C.Lakshimi Priya "Knowledge Based Emotion Recognition Approaches for Human Computer Intelligent Interaction Using Tree Structures "International Conference on Cognition and Recognition 2010.

[4] FaceGen Modeller: A Tool to generate random faces.

[5] Bihan Jiang, Michel Valstar, Brais Martinez, MajaPantic "A Dynamic Appearance Descriptor Approach to Facial Actions Temporal Modeling"IEEE Transactions on Cybernetics, Vol.44, NO.2, FEBRUARY 2014

[6] EsubalewBekele, ZhiZheng, Amy Swanson, Julie Crittendon, Zachary Warren, and NilanjanSarkar, "Understanding How Adolescents with Autism Respond to Facial Expressions in Virtual Reality Environments" IEEE Transactions on Visualization and Computer Graphics, Vol.19, No.4, APRIL 2013

[7] GuodongGuo, RuiGuo, Xin Li, "Facial Expression Recognition Influenced by Human Aging"IEEE Transactions on Affective Computing, Vol.4, No3, JULY-SEPTEMBER 2013

[8] Yun Tie, Ling Guan, "A Deformable 3D Facial Expression Model For Dynamic Human Emotion State Recognition", IEEE Transactions on Circuits and system for Video Technology, Vol.23, No.1, JANUARY 2013.

[9] Jacob Whitehill, Zewelanji Serpell, et al. "The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions" IEEE Transaction on Affective Computing, Volume 5, No.1 January -March 2014

[10] GuodongGuo, RuiGuo, Xin Li, "Understanding How Adolescents with Autism Respond to Facial Expressions in Virtual Reality Environments"
IEEE Transactions on Visualization and Computer Graphics volume 19, No 4, April 2013