

# Automatic Classification of Facial Expressions from Video Stream using Decision Tree

Bali Thorat

Vision & Intelligent System Lab  
Department of CSIT  
Dr. B.A. Marathwada  
University, Aurangabad

Ganesh Manza

Department of Planning and  
Statistics  
Dr. B.A. Marathwada  
University, Aurangabad

Pravin Yannawar

Vision & Intelligent System Lab  
Department of CSIT  
Dr. B.A. Marathwada  
University, Aurangabad

## ABSTRACT

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions. This paper presents automatic recognition system for *Happy, Surprise, Disgust, Sad, Anger and Fear* facial expressions contained in video streams using decision tree. The proposed method employs popular and updated 'Viola-Jones' detection method to detect the face, facial components and their classification using decision tree. This research work attempts to recognize fine-grained changes in facial expression and established their relationship with Facial Action Coding System (FACS). The proposed method resulted in average 76.43% correct classification of six basic expressions from video streams with 23.56% expression error rate.

## General Terms

Viola-Jones Algorithm, Sobel Edge Detection, Decision Tree based Classification.

## Keywords

Face Detection, Facial feature point extraction, Facial expression recognition, Feature extraction, Decision Tree.

## 1. INTRODUCTION

A human face is a central point intended for social interaction. It is the main source of information by which people identify others and focuses during a conversation. The impulse to build machines that are as human as possible sparks the research on face and facial expression recognition. Face recognition is one of the basic ability, which is required in order to construct a man-like machine [1]. Human computer interface is developed in such a way that the computer behaves as natural like the human. Thus, human facial features recognition becomes a topic that is paid attention by researchers [2].

For communication, human used two major features i.e. verbal and non-verbal modes for communication which employs words as well as facial expression respectively by human for communication or exchange of information in between. Facial expressions are the facial changes in response to a person's internal emotional states, intentions, or social communications. These facial changes can recognize by humans and give appropriate response to them. Ekman and Friesen [3] has represented 6 basic face expressions (emotions), that are universally recognized as *Happy, Surprise, Disgust, Sadness, Anger and Fear*. Each of these six basic emotions corresponds to a unique facial expression. Each expression was summarized by distinctive clues in the appearance of the facial components *eyebrows, eyes, mouth*, etc. These facial expression clues are further investigated and

encoded into the Facial Action Coding System (FACS) which describes, "All visually distinguishable facial movements." FACS enumerates "Action Units" (AUs) of a face that cause facial movements.

In this research paper, the attempt to automatically analyze and recognize facial motions and facial feature changes from visual information. Automatic expression analysis systems attempt to recognize a set of basic prototypic expressions i.e. *Happy, Surprise, Disgust, Sad, Anger and Fear*. However, such prototypic expressions occur occasionally in daily life. Change in isolated features, especially in the area of the brows or eyelids, is typical of paralinguistic displays; for instance, raising the brows signals greeting. To capture the sensitivity of human motion and paralinguistic communication, automated recognition of fine-grained changes in facial expression is needed [4]. This paper is organized in three phases i.e. face acquisition, feature extraction and facial expression recognition. Facial components were isolated by modified 'Viola-Jones' detection method and its classification using decision tree. The content of the paper is organized in seven sections, Section 1 dealt with Introduction, Section 2 deals with related work, Section 3 deals with Dataset, Section 4 deals with methodology adopted, Section 5 deals with Decision Tree oriented classification, Section 6 deals with detailed experiment and Result, and Section 7 deals with conclusion of work and references.

## 2. RELATED WORK

Ekman and Friesen (1971) [5] has represented 6 basic face expressions (emotions), which were *Happy, Surprise, Disgust, Sad, Anger and Fear*. Till date much of the research work was done on perceiving human face using neural network, templates and example-based techniques. There have been many attempts to automatically recognize AUs, both individually and in combination. Lee, Dileep M.R and Ajit Danti (1997) [6], has proposed LC face model for recognition of human facial expressions. To design an automatic facial expression recognition system Anastasios et al. (2008) [7] proposed a system which was divided into three major phases i.e. face image extraction, facial feature extraction to represent the facial expressions and finally classification phase for expression analysis. Jyh-Yeong (2001) [8], has proposed a methods and technique of Neural Network + Rough Contour Estimation Routine (RCER) where they define radial basis function network and a multilayer perception network. Ganesh Manza et.al (2008) [9], has addressed use of composite action unit normalization for laughing face. For larger database and less computational time Manal Abdullahv (2012) [10], proposed Principal Component Analysis. Behrang Yousef Asr Langeroodi, Kaveh Kia Kojouri has used neural network classifiers for automated facial expression

recognition [11]. Haar + AdaBoost method was proposed by Jacob and was used for detection of AUs corresponding to facial components like eye, eye brow regions etc. (2006) [12]. 2D Gabor filter, Multichannel Gabor filtration system was proposed by Michal for the detection of salient points and the extraction of texture features for image retrieval applications (2005) [13]. Action Unit detection process was done by classifying the features which was calculated from tracked facial points proposed by Valstar et al. their method detects a similar amount of AUs with similar or higher recognition rates (2005) [14].

### 3. DATASET

The proposed methodology was applied on 'VISExp' database. 'VISExp' database was collected in laboratory of Vision and Intelligent System lab of Department Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University. Each expression was recorded in a control environment for the duration of one second. Each user/subject was asked to generate expressions in 'Natural-Expression-Natural' mode without head movement. The subject chosen for data collection were professional theater artist and students from Department of Drama, Dr. Babasaheb Ambedkar Marathwada University Aurangabad. There were 280 samples collected for the work which includes four different subjects and each subject poses for several expressions like {Neutral, Happy, Surprised, Sadness, Disgust, Fear and Anger}. Samples from the database are shown in Figure. 1. Sampling rate for video stream is 26fps; this has generated a total of 7280 frames for expressions {Happy, Surprised, and Sadness, Disgust, Fear and Anger} form respective video streams.

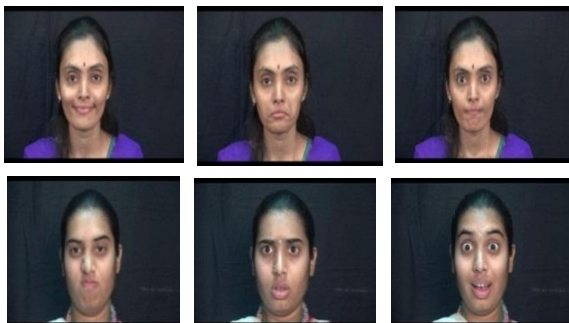


Fig 1: Samples of two expressers containing 6 different Expressions.

### 4. METHODOLOGY

The basic step in recognizing the facial expressions in a given video input is pre-processing of the videos. The video input is captured with the help of standard cameras. Once the input is acquired, it will be pre-processed and saved into 'VISExp' database. The proposed system was designed in light of basic system, which accepts video feed containing expression from 'VISExp' database and passed to pre-processing operation where each video feed was down sampled into discrete 2D image frame with a sampling rate of 26fps. This 2D image vector is further processed for actual extraction of features for expression and thresholds are to be determined so that changes can easily be tracked.

The preprocessing operations were performed on the video containing the face in a frame by frame manner. Each frame was processed and the possible facial components of the face were detected by the preprocessing process. The marked components by the process are eyes, mouth and eyebrows, the temporal variation and deformation was considered as main source of information in isolating expression therefore these

components were treated for feature extraction. The detected components are passed to facial features extractor based on Geometric feature methods.

The general approach to automatic facial expression analysis consists of three steps: face acquisition, facial feature extraction and representation, and facial expression recognition as shown in Figure 2.

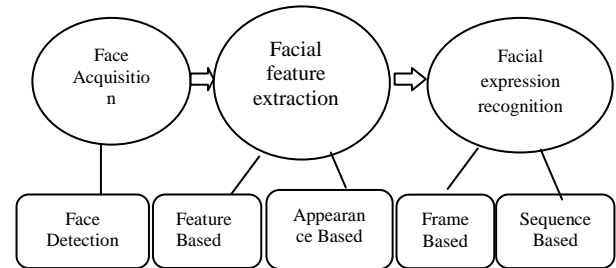


Fig 2: Basic Structure of facial expression recognition.

#### 4.1 Face Acquisition

Face acquisition is a processing stage; its objective is to automatically find the face region for the input video or sequences. It can pass through detector to detect face for each frame of the video sequence, as shown in Figure 3(a). To detect the face image in a scene a real-time face detection scheme proposed in [15] was used, which represents an adapted version of the original 'Viola-Jones' face detector [16]. The 'Viola-Jones' face detector consists of a cascade of classifiers trained by 'AdaBoost'. Each classifier employs integral image filters, which remind of Haar Basis functions and can be computed very fast at any location and scale. The Haar-like features comprising of two edge features, line feature and rectangle feature. For each stage in the cascade, a subset of features is chosen using a feature selection procedure based on 'AdaBoost'. This detects eyes, nose and mouth which were further modified in this research work for adding two more cascaded objects like eyebrows and eyeballs. Detected components from each frame are marked by bounding-box. Then the center of each bounding box associated with face components has been earmarked, as shown in Fig 3(b).



Fig 3: (a) Detection of Face from 2D Frame; (b) Detected faces components.

For perceiving the medial point the coordinates of bounding box are considered. After locating the medial point, the individual object from these bounding boxes was detected. Sobel Edge Detection method is used to detect the edges of these components. The Edge Detection block finds the edges in an input image by approximating the gradient magnitude of the image. The block convolves the input matrix with the Sobel. The block outputs two gradient components of the image, which are the result of this convolution operation. Sobel Edge Detection method is applied on eyebrows, pupil and mouth. Then the projection method is used which locates all nonzero elements of array, and returns the linear indices of those elements in vector used to locate the corner and middle point of these components.

## 4.2 Facial Feature Extraction and Representation

Once the face is located, the next step is to extract and represent the facial changes caused by facial expressions. In facial feature extraction for expression analysis, there are mainly two types of approaches: Geometric feature-based methods and Appearance-based methods. The geometric facial features present the shape and locations of facial components. The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. With appearance-based methods, image filters, are applied to either the whole-face or specific regions in a face image to extract a feature vector. Once the feature vector is generated it is necessary to code the distance among these principal components of the face. The Normalized distances between components are considered and then set the threshold value. The expression is constituted in change in the muscular moment of face from natural state to expression state. Hence it is important to understand the natural neutral state of face. In order to decide the natural neutral threshold value associated with component distance of face the video stream was processed and suitable threshold features were extracted and stored using equation (1).

$$\text{Threshold Value} = \frac{\sum_{l=0}^n Xl}{NOF} \quad (1)$$

Where, NOF stands for Number of Frames contained in video streams. Equation (1) used to calculate the Threshold Value, where Xi is the total geometric feature of respective components divided by Number of frames (NOF).

## 4.3 Facial Expression Recognition

In Facial expression recognition the facial changes can be identified as facial source file and the action units or prototypic emotional expressions. In this paper, classified the recognition approach is frame-based or sequence-based using decision tree. In this recognition process the threshold value of neutral frames was recorded, then this threshold value send to decision tree and then compared with other frames which have different actions i.e. when the expression of subject is 'Surprise' {AU1 + AU2 + AU5 + AU25 + AU27} are active, when expression is 'Happy' {AU6 + AU12 + AU25} are active, when expression is 'Fear' {AU1 + AU2 + AU4 + AU5 + AU7 + AU20 + AU26} are active, when expression is 'Sad' {AU1 + AU4 + AU15 + AU17} are active, when subject gives 'Anger' expression {AU4 + AU5 + AU7 + AU23} are active, similarly when subject gives 'Disgust' expression {AU9 + AU15 + AU16} are active [3].

## 5. DECISION TREE

Once the parameters were obtained the threshold values was set and proceed for creation of Decision tree. As decision trees are powerful and popular tools for classification and prediction [17]. The attractiveness of decision trees is due to the fact that, in contrast to neural networks, decision trees represent rules. Rules can readily be expressed so that humans can understand them.

Decision tree is a classifier in the form of a tree structure as shown in Figure 4, where each node is either:

A leaf node - indicates the value of the target attribute,

A decision node - specifies some test to be carried out on some threshold value.

In decision tree learning, threshold value of neutral expression used to generate a decision tree which gives rules for classifying the data. This builds the tree in the form of top-

down approach with no back tracking which specified how to take decision in order to recognize different Facial Expression [18].

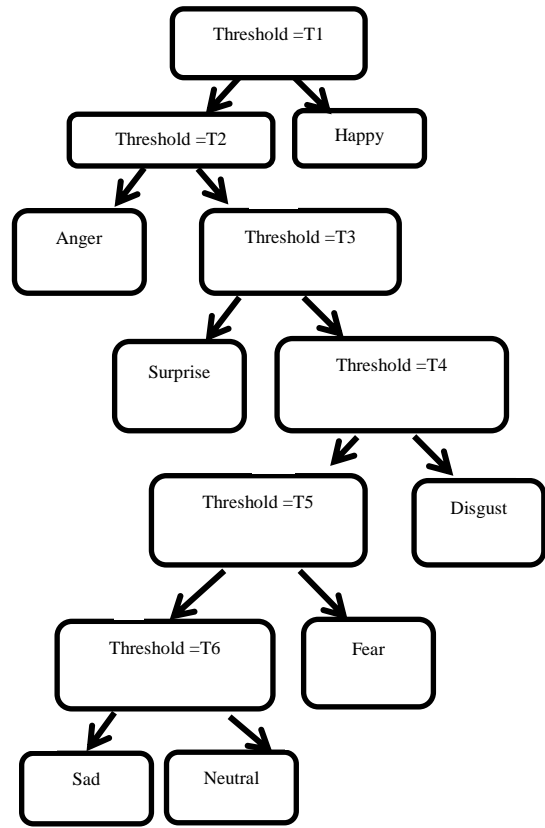


Fig 4: Decision Tree for expression recognition.

Table 1. Action Unit and Respective Threshold

Threshold	Action Units	Range					
		MW	BE	LCP	MH	BSE	LRB
T1	6+12+25	>74	35-45	<=1	>=15	>27	>=34
T2	4+5+7+23	54-83	35-45	<=1	<20	>27	>=34
T3	1+2+5+25+27	54-90	>=45	<=1	>=20	>=30	>=34
T4	9+15+16	<=84	30-45	<=1	>=20	<=27	>=34
T5	1+2+4+5+7+20+26	<=76	30-45	<=1	>=30	>27	<34
T6	1+4+15+17	>=57	30-45	>1	>=20	>27	>=31

MW – Width of mouth. BE – Distance between eyebrows to eye. LCP - Distance between lips Corner Point MH- Height of Mouth. BSE - Distance between eyebrow start points to eye LRB – Distance between left to right brow.

Table 1 shows, threshold and respective action units used for the classification of expression. The decision tree was built in accordance to the values presented in the table 1. The video streams were passed to decision tree for classification based on the extracted feature and the decision tree classify the frame in to respective class.

## 6. EXPERIMENT AND RESULT

Table 2. Expression Classes – Recognition Rate

Data Stream	Neutral	Happy	Surprise	Sad	Anger	Disgust	Fear	Non Classified
Neutral	<b>71.73</b>	0.51	2.84	10.23	0.41	1.82	12.46	5.10
Happy	29.29	<b>53.16</b>	2.76	2.27	2.17	0.89	9.47	2.50
Surprise	33.68	0.31	<b>57.20</b>	5.60	0.00	0.41	2.80	7.21
Sad	48.58	1.26	0.42	<b>47.84</b>	0.00	1.16	0.74	8.56
Anger	37.78	3.23	1.12	7.87	<b>45.08</b>	3.65	1.26	11.66
Disgust	33.33	2.00	0.00	4.52	1.37	<b>48.48</b>	10.30	8.56
Fear	28.19	1.02	2.68	6.76	0.00	1.91	<b>59.44</b>	24.62

The 'VISExp' database of facial expression consists of 6 expressions express by 4 individual users. Above facial expression recognition method were tested on 'VISExp' database and their accuracy was measured. So overall for 4 users create total 280 video samples for basic six expressions which generate 7280 frames with rate of 26 frames/sec, the total volume of database is 7280. Table 2 shows the Expression Classes – Recognition Rate using equation 2.

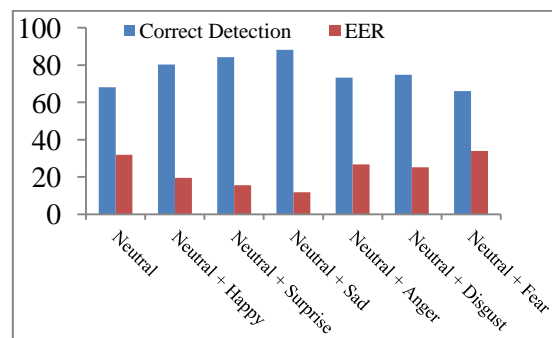
$$\text{Recognition Rate} = \frac{\text{Class of Expression}}{\text{No.of Frames-Unclassified Expression}} * 100 \quad (2)$$

Equation (2) used to calculate recognition rate where class of Expression represents number of frames correctly marked with recognized expressions according to respective class and No. of Frames means total number of frames exists in video stream used for recognition and Non-classified Expression means number frames which are not classified by system.

After applying facial expression recognition method on each user expression total frames of that users are separated into True and false recognition columns i.e. recognized and unrecognized frames and using these values the accuracy was calculated. Table 3 and Graph 1 show the result of facial expression recognition system. Using this system 80.38% result for happy expression, 84.33% result for surprise expression, 73.20% result for anger expression, 74.81% result for disgust expression, 66.06% result for fear expression, 88.17% result for sad expression were observed.

Table 3. Recognition result and error rate of the system

Expression	Correct Detection	Expression Error Rate (EER)
Neutral	68.08	31.92
Neutral + Happy	80.38	19.62
Neutral + Surprise	84.33	15.67
Neutral + Sad	88.17	11.83
Neutral + Anger	73.20	26.80
Neutral + Disgust	74.81	25.19
Neutral + Fear	66.06	33.94
Average	<b>76.43</b>	<b>23.56</b>



Graph 1: Recognition result and error rate of the system

## 7. CONCLUSION

This research work describes modification of cascade object detector mechanism of 'Viola Jones' for detection of face components which inherently contributes in formation of facial expression. The result presented in the paper clearly provides confirmation about using decision tree based approach for expression identification from video stream. The proposed method resulted in average 76.43% correct classification with expression 23.56% error rate. The work may further extended to other complex expressions, synthetic face and synthetic expressions with FACS.

## 8. ACKNOWLEDGMENTS

The Authors gratefully acknowledge support by the Department of Science and Technology (DST) for providing financial assistance from Major Research Project sanctioned under Fast Track Scheme for Young Scientist, vide sanction number SERB/1766/2013/14 and the authorities of Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India, for providing the infrastructure for this research work.

## 9. REFERENCES

- [1] C.B. Tatepamulwar V.P. Pawar H.S. Fadewar, "Techniques for Facial Expression Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 3, March 2014 ISSN: 2277 128X.
- [2] Sutedjo, Aryunto. "Experimental Study on Lip and Smile Detection." Jurnal Ilmu Komputer dan Informasi 4.2 (2012): 59-65.

- [3] Ekman, P, Friesen, "Constants across Cultures in the Face and Emotion", *J. Pers. Psycho. WV*, 1971, vol. 17, no. 2, pp. 124-129.
- [4] Thorat Bali A, Manza Ganesh R, Yannawar Pravin L, "Automatic Detection of Facial Expressions from Video Streams", *NLPDM 2015*, Veer Narmad South Gujarat University, Surat, Gujarat March 3-4 2015.
- [5] Ekman, P, Friesen, "Constants across Cultures in the Face and Emotion", *J. Pers. Psycho. WV*, 1971, vol. 17, no. 2, pp. 124-129.
- [6] De Silva, Liyanage Chandratilake, Facial emotion recognition using multimodal information, vol 1(1997) pp 397-401.
- [7] Koutlas, Anastasios, and Dimitrios I. Fotiadis, "An automatic region based methodology for facial expression recognition". *Systems, Man and Cybernetics*, 2008. *SMC 2008. IEEE International Conference on. IEEE*, 2008.
- [8] Jyh-Yeong Chang and Jia-Lin Chen, "Automated Facial Expression Recognition System Using Neural Networks" *Journal of the Chinese Institute of Engineers*, Vol. 24, No. 3, pp. 345-356 (2001).
- [9] Ganesh R Manza, P L Yannawar, "Mathematical Normalization of Laughing face using Action Units", *Proceedings of International Conference on Advances in Computing*, Anuradha Engineering College, Chikhali, ISBN:978-81-906457-0-6, Feb 21-22, 2008 pp 251-253.
- [10] Manal Abdullah1, Majda Wazzan1 and Sahar Bo-saeed "Optimizing Face Recognition Using PCA" *International Journal of Artificial Intelligence & Applications (IJAIA)*, Vol.3, No.2, March 2012.
- [11] Langeroodi, Behrang Yousef Asr, and Kaveh Kia Kojouri. "Automatic Facial Expression Recognition Using Neural Network."
- [12] Jacob Richard-Whitehill, "Automatic Real-Time Facial Expression Recognition for Signed Language Translation", Department of Computer Science, University of the Western Cape. May 2006.
- [13] Tomasz Andrysiak, Michał Choraś "Image Retrieval Based On Hierarchical Gabor Filters" *Int. J. Appl. Math. Compt. Sci.*, 2005, Vol. 15, No. 4, 471-480.
- [14] M. Valstar, I. Patras, M. Pantic, Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data, in: *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, vol. 3, 2005, pp. 76-84.
- [15] I.R. Fasel, B. Fortenberry and J.R. Movellan, "A generative framework for real time object detection and classification", *Int'l J Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 181-210, 2005.
- [16] P. Viola and M. Jones, "Robust real-time object detection", *Technical Report CRL 20001/01*, Cambridge Research Laboratory, 2001.
- [17] Yegui Xiao, Member, IEEE, L. Ma, K. Khorasani, Member, IEEE, "A New Facial Expression Recognition Technique".
- [18] Shubhrata Gupta, Keshri Verma & Nazil Perveen, "Facial Expression Recognition System Using Facial Characteristic Points and ID3", Electrical Department, NIT Raipur MCA Department, NIT Raipur.