

FCA: A Proposed Method for an Automatic Facial Expression Recognition System using ANN

Jyoti Mahajan

Assistant Professor, Computer Engg. Dept.
Government College of Engineering &
Technology, Jammu.

Rohini Mahajan

Lecturer, Computer Engg. Dept.
Government College of Engineering &
Technology, Jammu.

ABSTRACT

Facial Expression plays an important role in human communication, allowing people to express themselves without the use of any verbal means but still understanding each other's mood. Thus the interfaces must have the ability to detect any kind of change in the behavior of the user and to communicate based on the information available through interaction rather than the commands given by user. Thus, facial expression recognition is a challenging problem in computer vision. The project retrieves real-time images from a webcam and converts them to gray scale images. In facial feature expression recognition system we calculate 18 feature values from 16 feature points extracted from facial images. Then, these pre-defined feature vectors extracted from the images are sent to multilayer perceptron network for training and classification using back propagation. Using the result, the software will be able to develop human computer interaction and to judge the emotions of the user. The main aim is to work upon five different emotions - neutral, happy, sad, surprised and fear. Using only two-thirds of the total features, our approach achieves a classification rate (CR) which is higher than the CR obtained using all features. The system also outperforms several existing methods, evaluated on the combination of existing and self-generated databases. \

Keywords

Feature vectors, covariance analysis, perceptron, morphological, CascadeObjectDetector, holistic.

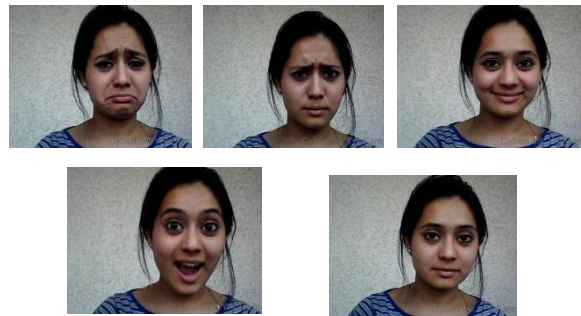
1. INTRODUCTION

Human facial expression is controlled by neurons and their neural network and also muscles of the skin that enables the user to convey the behavior and for inter human communication. To recognize human expression is a tedious task and also an important one for different fields of studies research and games. There are five basic facial expressions that project the human psychology: fear, happiness, neutral, sadness and surprise. Examples of these facial expressions are shown in Figure 1.

Sad

Angry

Happy



Surprise

Neutral

Figure 1: Five basic emotions of humans

In this paper, a novel approach is proposed to recognize facial expressions from static images, via extraction and selection of salient features points and finding the corresponding feature vectors. So the technique is to use filters to extract all the feature points and corresponding feature vectors. Along with that the analytic approach to form computer interactive software is followed. In this facial expression recognition, 18 feature values can be obtained from the 16 feature points marked on the facial features like eyelids, eyebrows and lips which are extracted from the facial images captured by a color web camera [2].

Taking into consideration of well-description of AUs and easiness for image processing, the feature points were obtained and corresponding feature values were decided as shown in Figure 2.

The feature value extraction part consists of the following 4 steps:

1. Extraction of face region,
2. Extraction of facial organs,
3. Extraction of feature points and
4. Calculation of feature values

As the feature points to calculate the feature values, 3 points on both eyebrows, 4 points on both eyes and mouth, totally 18 points were selected.

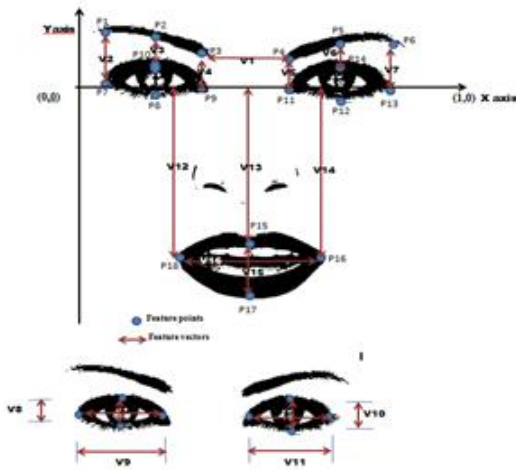


Figure 2: Estimation of feature points and feature values

Then, the degrees of the basic five expressions are recognized from the feature values and hence the values are given as input to the neural nets for training of the nets and finally classification of the emotion of a new user is done using back propagation on a multi-layer perceptron. This network has 18 input values and five output values. The aim is to design an interface that has human machine interaction with an input channel in which the computer recognizes and classifies the human emotion by capturing real time image using a webcam and with an output channel in which computer presents the emotions to human.

2. LITERATURE REVIEWED

In feature extraction technique two types of filters are used. Fixed Filters are used to extract primitive features such as edges. Adaptive Filters are trained to extract more complex facial features for classification as with adaptive filters the specifications are not known and change with time. Normally the face detection is done in 2 ways [1] Holistic Approach is the face is determined as a whole unit. After the face is detected a template based model is used and Analytic Approach is the one in which only some important facial features are detected. After the face is detected, then the featured-based methods will be used to track the facial features. There are several approaches taken in the literature for learning classifiers for emotion recognition [1]. The Static approach where classifier classifies video frame and matches to one of the facial expression categories based on the tracking results of that frame using Bayesian classifier and Dynamic Approach where classifiers take the temporal pattern in displaying facial expression using Hidden Markov model (HMM) for facial expression recognition

Existing technique for facial expression recognition can be broadly classified as [4]. Geometric-based approaches are techniques in which facial image is actually converted into an image with points on the on the various facial organs be it eyes, eyebrows, nose and lips or the shape of facial organs [6]. Classification is done by analyzing the distances between these feature points or the relative sizes of the facial components. Pantic *et al.* [6] proposed a method for detecting facial actions by analyzing the contours of facial components, including the eyes and the mouth. A classifier is then used to recognize the individual facial muscle action units (AUs). However, these methods require obtaining correct location of the feature points especially when the image is of low quality or with a complex background. Appearance-based methods process the entire image by applying a set of filters to extract

facial features. Zhen *et al.* [7] used Gabor wavelets to represent appearance changes as a set of multi-scale and multi-orientation coefficients. They proposed a ratio-image based feature that is independent of the face albedos. Their method can cope with different people and illumination conditions. Feng [7] used Local Binary Patterns (LBP) to extract facial texture features and combined different local histograms to recover the shape of the face. Facial expression recognition can be improved by combining appearance and geometric features using Hybrid-based approaches. Zhang and Ji [8] proposed a multi-feature technique that is based on the detection of facial points, nasolabial folds, and edges in the forehead area. In their method, facial features are extracted by associating each AU with a set of movements, and then classified using a Bayesian network model.

Various techniques being followed includes Co-Variance Analysis and Feature vectors and feature value technique by Ekman and Friesen which are geometric based techniques and Region extraction and Morphological Image Processing and Contouring which are appearance based techniques. Co variance analysis was reduced to two main steps where the first step includes finding the emotive facial features which includes the eyes and the mouth and second step involves finding the matrix which is the covariance matrix for the difference between the neutral emotion and the other four emotions. After the extraction of the eyes and the mouth region which are most prominent evidences of the relative emotion we obtain and compare the covariance of the neutral emotion and any other emotion. The covariance matrix is formed which contain the following feature vector values. $f_k = [r \ c \ I(r,c) \ G(r,c)]^T$

Where r = the pixel row, c = the pixel column, $I(r, c)$ = the intensity of the gray scale image at (r, c) and $G(r, c)$ = the gradient magnitude (computed by a pair of Gaussian derivative masks) at (r, c) . Ekman and Friesen developed the Facial Action Coding System (FACS) to code facial expressions where movements on the face are described by a set of action units (AUs). Each AU has some related muscular basis. This system of coding facial expressions is done manually by following a set of prescribed rules. The inputs are still images of facial expressions. When AUs occur in combination they may be additive, in which the combination does not change the appearance of the constituent AUs, or non-additive, in which the appearance of the constituents does change. [11]. Region extraction is extraction of the region of interest by calculating the f_x and f_y gradients using a Sobel filter, followed by calculating the gradient image. Eyes are always located in the upper half of the face, and the mouth located in the lower half and hence the algorithm is executed to exactly locate the position of the eyes and mouth on the facial image. A kernel of 60x150 columns is run vertically till one half of the face, to obtain the width of the eyes. Similarly, we extract the mouth region by projecting its gradients on the axes for the lower half of the face. [10]. This may result in having mouth region and eye region overlapping. Morphological Image Processing and Contouring is the technique in which to extract facial features and facial image is segmented and forms the Region of Interest which is then resized to larger size to make facial components more prominent. SUSAN edge detection operator and then places a mask around the pixels and calculates the number of pixels in mask with the same brightness and refers that mask as SUSAN. Then we trace the boundaries of segments and draw bounding box by taking into account x, y coordinates and height and width of each segment. Image is partitioned into two regions i.e. upper and lower portion on the basis of center

of cropped face assuming the fact that eyes and eyebrows are present in the upper part of face and mouth and nose is present in the lower part. A portion is an upper part if x and y values are less than centx and centy where centx and centy are x- and y coordinates of center of cropped image. Eyes and eyebrows are present in this area. For left eye and eyebrow portion certain threshold for values of x and y is considered for eliminating outer segments. For right eye and eyebrow also specific threshold value is chosen for eliminating outer segments. An area is, of the lower portion if, its value is greater than centx and centy where centx and centy are x- and y coordinates of center of an image. Nose and mouth are present in this area. For nose and mouth area segments, x lies in the specific range and y also lies in certain range is considered as region of interest for eliminating outer segments. The number of segments for each portion is stored. If number of segments is > 2 then following procedure for combining the segments is called. Segments are checked in vertical direction. If there is overlapping then the segments are combined.

Again if segments are >2 then distance is obtained and the segments which are closer are combined. This process is repeated until we get two segments for each part and in all total six segments. Bounding box location of feature segments obtained in the above step are used to calculate the height and width of left eyebrow, height and width of left eye, height and width of right eyebrow, height and width of right eye, height and width of nose and height and width of mouth. Distance between center of left eye and eyebrow, right eye and eyebrow and mouth and nose is also calculated. Thus total 15 parameters are obtained and considered as feature vector

3. WORKING OF THE METHOD OF THE AUTOMATIC FACIAL RECOGNITION SYSTEM

Facial expression recognition process is a part of facial image analysis. A. Mehrabian [(1968) has mentioned that the verbal part of a message contributes only 7% of its meaning, the vocal part contributes 38% while facial expression gives 55% to the effect of that message. Hence the facial part does the major contribution in human communication. There are five basic types of facial expressions. They are neutral, fear, sadness, surprise and happiness. From the review of papers on facial expression, it is clear that most of the facial expression recognition systems (FERS) were based on the Facial Action Coding System (FACS), Y.L. Tian *et al.* (2001), Y. Tong *et al.* (2007), M. Pantic *et al.* (2000). In this system, the changes in the facial expression are described with FACS in terms of 44 different action units (AUs), each of which is related to the facial muscle movements. 44 AUs can give up-to 7000 different combinations, with wide variations due to age, size and ethnicity. M. Pantic *et al.* (2000) gave detail survey on facial expression recognition in their paper. Most of the research works on facial expression recognition (FER) are grounded on still images [5]. The proposed methodology followed by the system:

1. Read input image from database and localize face using morphological image processing operations
2. Crop the face image.
3. Extract features using our algorithm and find Facial Feature vectors.
4. Train neural network.
5. Recognize expression

Initially the web-cam takes the image. This image must be a frontal image with proper illumination of light, frontal

orientation of face, no facial hair, moustache, goggles, spectacles, etc. The image processing includes binarisation of the image or converting the image into grey scale image, removal of noise from the image, detecting the grey levels of the various pixels of the image and finally segmentation.

4. PROPOSED FACIAL CODING ALGORITHM (FCA) FOR THE FACIAL EXPRESSION RECOGNITION

- Initially the frontal facial region is selected which forms the actually region of interest.
- Then the facial image is divided into two parts as the upper portion and lower portion of the face. The upper portion of the face includes eyes and eyebrows in [portion(X)] while the lower portion includes the nose and the mouth [portion (Y)].
- The image is further segmented. Now portion(X) of the image is bilaterally divided. This forms portion(X) left and portion(X) right where former includes the left eye and left eyebrow while the latter includes the right eye and right eyebrow.
- Now each portion is individually taken to form the bounding box and the Region of Interest (ROI) for the various facial organs. This is done using the CascadeObjectDetector.
- It also includes finding the ROI using in-plane rotation or rotation along left or right axis (out-of-plane rotation) or both.
- After finding the ROI we locate the facial organs using the same technique. Then we locate feature points on each of the facial organs. Each eyebrow has 3 points and each eye has 4 points.
- The points on the left eyebrow: P1, P2 AND P3. The points on the right eyebrow: P4, P5 AND P6. The points on left eye: P7, P8, P9 and P10. The points on right eye: P11, P12, P13 and P14.
- Similarly, feature points for the portion(Y) of the face are obtained. This portion(Y) includes nose and mouth. Mouth region has four points P15, P16, P17 and P18.
- In totality there are 18 feature points which are obtained from various facial organs and based on these facial points there are 16 facial vectors obtained and calculated based on the Euclidean distance (root mean square value). This is best method for obtaining the distance between two points.
- The distances obtained are as follows:
 - V1- distance vector between point P3 and P4
 - V2- distance vector between point P1 and P7
 - V3- distance vector between point P2 and P10
 - V4- distance vector between point P3 and P9
 - V5- distance vector between point P3 and P4
 - V6- distance vector between point P4 and P11
 - V7- distance vector between point P6 and P13
 - V8- distance vector between point P8 and P10
 - V9- distance vector between point P7 and P9
 - V10- distance vector between point P12 and P14
 - V11- distance vector between point P11 and P13
 - V12- distance vector between point P18 and X-axis
 - V13- distance vector between point P16 and X-axis
 - V14- distance vector between point P15 and X-axis
 - V15- distance vector between point P15 and P17
 - V16- distance vector between point P16 and P18
- These distance vectors are given as input to the input layer of the network and for classification technique feed-forward back propagation algorithm is used.

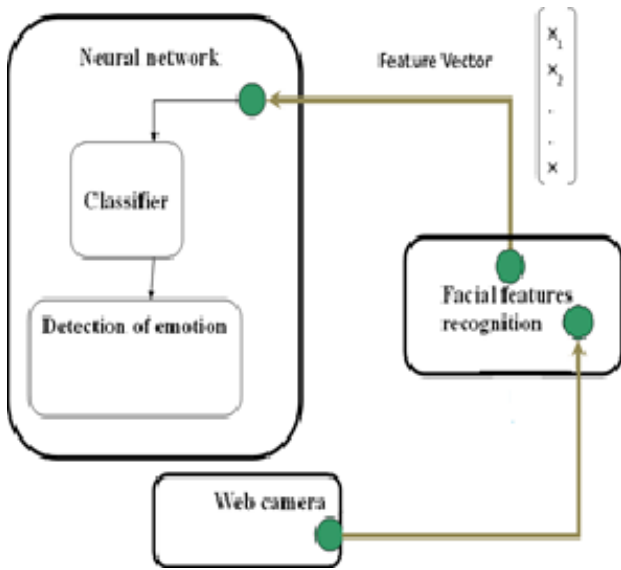


Figure 3: Proposed method

The neural network used in classification method is a feed forward multi-layer perceptron which follows back propagation algorithm using sigmoid function and three hidden layers.

Other techniques being used are SOM. SOM, known also as Kohonen's self-organizing map, which is a clustering technique that can be used to automatically provide insight into the nature of data without supervision. Then the transformation is done using the unsupervised neural network into a supervised LVQ (Learning Vector Quantum) neural network, which is another clustering technique. When handling noise and multiple inputs of data, back propagation performs better than SOM. LVQ is excellent for classification, but when handling noise is a little bit worse than back propagation. This is special reason that makes back-propagation better than the numerous other neural network models.

The technique actually depends on the three main factors:

- No. of Iterations
- No. of local maxima and local minima
- Error difference rate

The designed application would be able to judge the efficiency on all the parameters but the most efficient one is the error difference rate. This parameter would just help to provide a stopping condition or a stopping parameter to our algorithm.

5. IMPLEMENTATION AND RESULTS

The data is selected from 9 persons, including 5 females and 4 males. We did not put any constraint on their age. This test is done mostly with short clips, thus it was more likely to get only some particular expressions such as neutral, joy, and sometimes surprise.



Figure 4: Analysis of the expressions , cropping of the image and formation of bounding box to obtain the region of interest

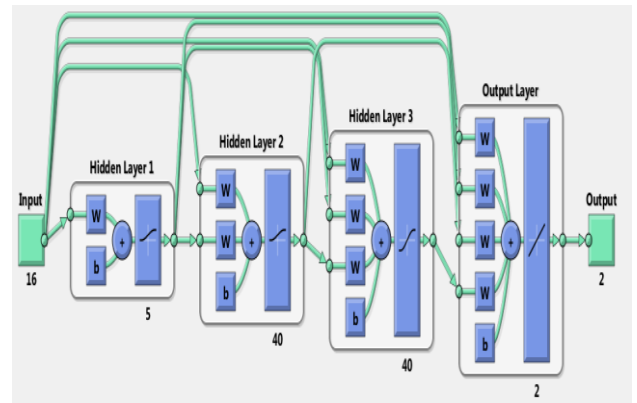


Figure 5: Training of the network for the recognition of the expression

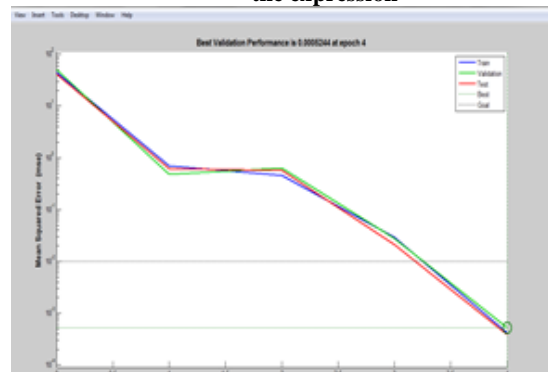


Figure 6: Performance Plots

Implementation of the method is done on 45 images and according to the results its concluded that it is 68.8% accurate.

Table 1: Study of the emotions

Emotion/ Gender	Total	Accurately classified	Percentage
Angry	9	6	66.7%
Happiness	9	7	77.8%
Neutral	9	8	88.9%
Sad	9	4	44.4%
Surprise	9	6	66.7%

6. CONCLUSION

68.8% accuracy (approx.) is obtained as a few misses in the proposed method are encountered. The misses usually included regions with a similar skin likelihood values and

those regions that are definitely skin regions but they correspond to other parts of the body such as the neck, arms, and legs. In males the accuracy is about 70% while in females the accuracy is about 68%. So, the given technique is capable of finding the expressions along with the gender of the person.

7. FUTURE SCOPE

The further improvement in this method can be done by using the improved technique for the thresholding and edge detection in the image. Also as of now the software is working only for the images clicked by the webcam and now on any other standardized data set so the future work would be to make the program work for all the images not just the standardized images or the images clicked by webcam.

8. REFERENCES

- [1] Yafei Sun Under Supervision Of Dr. Michael S. Lew and Dr. Nicu Sebe August 2003 Neural Networks For Emotion Classification
- [2] H. Shimada T. Kunihiro, D. Yang, H. Yoshikawa, 1997 Design of Affective Interface for Realizing Human-Machine Empathy
- [3] Hamit Soyel and Hasan Demirel May 2008 Facial Expression Recognition Using 3d Facial Feature Distances
- [4] P. Li, S. L. Phung, A. Bouzerdom, And F. H. C. Tivive 2010 IEEE Feature Selection For Facial Expression Recognition
- [5] Mohammad Shahidul Islam¹, Surapong Auwatanamongkol “A Novel Feature Extraction Technique for Facial Expression Recognition” *IJCS International Journal of Computer Science Issues, Vol 10, Issue 1, No 3, January 2013*
- [6] A. Mehrabian, “Communication without words”, *Psychology today, vol. 2, no.4, pp. 53-56, 1968.*
- [7] Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen, Jian-Cheng Huang, — “A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA”, in *International Journal of Information Technology Vol. 11 No. 11, 2005*
- [8] Aleksic.P.S. And A. K. Katsaggelos. “Automatic Facial Expression Recognition Using Facial Animation Parameters and Multi-Stream HMMs”, *6th European Workshop on Image Analysis for Multimedia Interactive Services*, Montreux, Switzerland, 2005.
- [9] Amit Pillay & Ravi Mattani in 2010 Facial Feature Detection
- [10] Mike Lucia, Jatan Pandya, Tom Zajdel December 2, 2010 “Emotion Detection”
- [11] Ying-li Tian¹ Takeo Kanade¹ and Jeffrey F. Cohn in 2000 Recognizing Action Units for Facial Expression Analysis
- [12] Hua Gu Guangda Su Cheng Du, “Feature Points Extraction from Faces Image and Vision Computing”, pp.154-158, NZ, Palmerston North November 2003.

Jyoti Mahajan obtained his B.E. Degree in Computer Science & Engineering from Bangalore University, M.S. (Software Systems) from BITS Pilani and Ph.D. in Computer Science from University of Jammu. He has published 10 papers in International and National Journals of repute and presented 9 papers in National and International Conferences. He is working as Assistant Professor in the Department of Computer Engineering at Government College of Engineering & Technology, Jammu, India. He has about 14 years teaching experience. His Research interests include Data Structure, Software Engineering like Software Systems, Neural Networks., Image Processing.

Rohini Mahajan obtained his B.E. Degree in Computer Science & Engineering from Government college of Engineering and Technology, Jammu, M.TECH from Shri Mata Vaishno Devi University, Katra. She is working as Lecturer in the Department of Computer Engineering at Government College of Engineering & Technology, Jammu, India. Her Research interests include Artificial Intelligence, Neural Networks, Image Processing.