FACIAL EXPRESSION CLASSIFICATION USING STATISTICAL, SPATIAL FEATURES AND NEURAL NETWORK

Nazil Perveen[1], Shubhrata Gupta[2] and Keshri Verma[3]
[1&2]Department of Electrical Engineering, N.I.T Raipur, Raipur, Chhattisgarh, 492010, India
[3]Department of M.C.A, N.I.T Raipur, Raipur, Chhattisgarh, 492010, India
[1]786.nazil@gmail.com, [2]guptashubhrata@rediffmail.com, [3]keshriverma@gmail.com

ABSTRACT

Human facial expression contains extremely abundant information of human's behavior and can further reflect human's corresponding mental state. Facial expression is one of the most powerful, natural, and abrupt means for human beings which have the ability to communicate emotion and regulate inter-personal behavior. This paper provides a novel and hybrid approach for classifying the facial expression efficiently. A novel approach because it evaluates the statistical features namely, kurtosis, skewness, entropy, energy, moment, mean, variance and standard deviation of the whole face and spatial features which are related to the facial actions. Mostly the information about the expressions are concentrated on the facial expression regions such as mouth, eye and eyebrows, so these regions are segmented and templates are being created. Using these templates we calculate the spatial features of the face to classify the expression. And a hybrid approach because both the features are merged and drive through the multi-label Back-propagation neural network classifier. The whole technique is being implemented and tested using JAFFE database in MATLAB environment where the accuracy achieved during classification is 70%.

Keywords: *Back-Propagation Neural Network classifier, Facial Expression Recognition, Spatial Features Statistical Features.*

I. INTRODUCTION

Recognition of facial expression has been an active area of research in literature for the long time. Human facial expression recognition has attracted much attention in recent years because of its importance in realizing highly intelligent human-machine interfaces and it contains extremely abundant information of human behavior which plays a crucial role in inter-personal behavior. The major purpose of facial expression recognition is to introduce a natural way of communication in man-machine interaction. Over the last decade significant effort has been made in developing the methods for facial expression analysis. Facial expression is produced by the activation of facial muscles, which are triggered by the nerve impulses. There are basic seven facial expressions neutral, happy, surprise, fear, sad, angry and disgust. These six basic facial expressions are needed to recognized so that, it will be boon to different research areas. Facial expression recognition has wide variety of application, such as, to develop the friendly man-machine interface to enable the system to have communication analogous to man-machine communication, behavioural science, clinical studies, psychological treatment, video- conferencing and many more.

In this whole research we drive through the different procedure. We total consider 224 images were, 154 images are input for training and 70 images are used for testing. In the initial stage we input the images were we perform pre-processing by extracting region of interest; next we extract the statistical feature of the whole face. In the second stage we create templates and match those templates which help in extracting the spatial features. We merge both these features to increase the efficiency of neural network. In our research work we make use of Back Propagation network to train and test the images.

II. LITERATURE REVIEW

Mehrabian[1] indicated that the verbal part (i.e. spoken words) of a message contributes only 7% of the effect of any message; the vocal part (i.e. voice information) contributes for 38% while facial expression contributed for 55% of the effect of the message. Hence, facial expression play important role in cognition of human emotions and facial expression classification is the base of facial expression recognition and emotion understanding [2]. The ultimate objective of facial expression classification and recognition is being the realization of intelligent and transparent communication between human and machines.

In 1978, Paul Ekman and Wallace V. Freisen implemented Facial Action Coding System (FACS) [3], which is the most widely used method available. They analysed six basic facial expressions, which include surprise, fear, disgust, anger, happiness and sadness.In FACS, there are 46 AUs that account forchanges in facial expression. The combination of these action units results in a large set of possible facial expressions. Over the year 90's different researches have been proposed, for example [4]-[10] and the references there in.

Several techniques had been proposed to devise facial expression classification using neural network. In 2007, Tai and Chung[11] proposed automatic facial expression recognition system using 'Elman neural network' with accuracy in recognition rate is 84.7%, in which they extracted the features using canthi detection technique. In 1999, Chen and Chang [12] proposed facial expression recognition system using 'Radial Basis Function and Multi-Layer Perceptron' with accuracy in recognition rate is 92.1% in which they extracted the facial characteristic points of the 3 organs. In 2004, Ma and Khorasani[13] proposed facial expression recognition system using 'Constructive Fees Forward Neural Networks' with accuracy in recognition rate is 93.75%. In 2011, Chaiyasit, Philmoltares and Saranya[14] proposed facial expression recognition system using 'Multilayer Perceptron with the Back-Propagation Algorithm' with the recognition rate 95.24%, in which they implements graph based facial feature extraction.

III. PROPOSED METHODOLOGY

The proposed methodologyis being explained in Figure 1. In this whole research we first extract the statistical feature for which the image is need to be pre-process. Once the image is being processed the statistical feature extraction is done in which we evaluate certain statistical metrics for example mean, variance, standard deviation, kurtosis, skewness, moment, entropy and energy. After evaluating statistical features, spatial features are being evaluated for which we follow the template matching algorithm using correlation technique. Once the template is being matched the facial points are evaluated which help in calculating the spatial features such as opening and width of eyes, mouth and height of eyebrows. Both these features are merged and set as input to the neural network classification technique which follows the Back-Propagation algorithm to classify the expressions.

3.1. Pre-processing

In order to evaluate statistical feature we need to perform pre-processing. In the initial stage, image is input in order to obtain the region of interest (ROI). The ROI of the face is obtained by simply cropping the area which does not contribute much information in recognizing the facial expressions. As the background details and the hair of the images in JAFFE databases does not contribute much information in recognizing the expressions. The ROI is obtained by cropping the image and reducing the matrix size from 256×256 to 161×111. Some of the examples are described in Table1.

3.2. Evaluating Statistical Features.

Once the region of interest is being obtained from the input image we extract the statistical features. The feature which we are evaluated in this research is as follows.

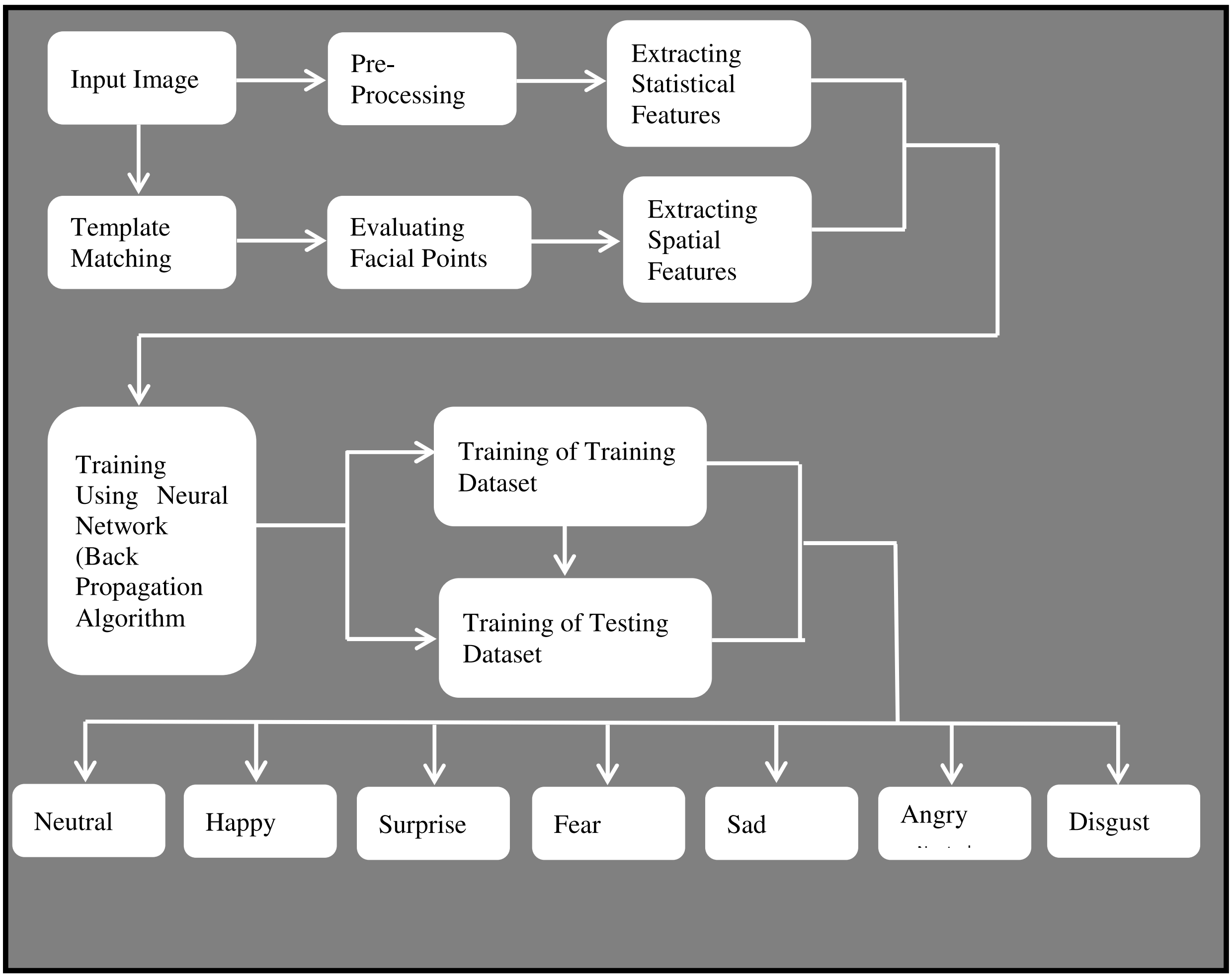


Figure 1. Proposed Methodology for Facial Expression Classification.

Table 1. Deducing ROI from the input face images.

Expression	**Input Image (256×256)**	**ROI images (161×111)**
Happy Face		
Disgust Face		
Surprise Face		
Sad Face		
Angry Face		

Neutral Face		
Fear Face		

3.2.1. Kurtosis

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution, that is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails [17]. Data sets with low kurtosis tend to have top near mean rather than a sharp peak.

$$\text{Kurtosis}=\frac{\sum_{i=1} N(Yi-\bar{Y})^4}{(N-1)s^4} \quad (1)$$

3.2.2. Skewness

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same the left and right of the centre point [18].

$$\text{Skewness}=\frac{\sum_{i=1} N(Yi-\bar{Y})^4}{(N-1)s^4} \quad (2)$$

3.2.3. Moment

Moment is a quantitative measure of the shape of set of data points. The 'second moment', for example, is widely used and measures the 'width' of a set of data points [19].

$$m_k = E(x-\mu)^k \quad (3)$$

Were, k is the order and in order to calculate central moment its value is 2.

3.2.4. Entropy

Entropy is a measure of uncertainty associated with random variables. Entropy of the gray scale images is a statistical measure of randomness that can be used to characterize the texture of the input image [20]. Entropy is defined as

$$\text{Entropy} = -\text{sum}\big(p.* \, log2(p)\big) \quad (4)$$

3.2.5. Energy

Energy is also termed as uniformity in Matlab which is also used to characterize the texture of the input image. Energy defined the properties of gray-level co-occurrence matrix. In Matlab, energy is achieved from 'graycoprops' function. The 'graycoprops' function 4 properties, i.e. 'Contrast', 'Correlation', 'Energy', 'Homogeneity'[21]. We consider here the 2 properties i.e. 'Contrast' and 'Correlation' as the variation in the values are obtained in these two parameters.

Were,

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma i \sigma j} \quad (5)$$

And

$$\text{Contrast} = \sum_{i,j} |i-j|^2 p(i,j) \quad (6)$$

Contrast returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. It is 0 for constant image. Whereas correlation returns a measure of how correlated a pixel is to its neighbour over the whole image. It is not a number for the constant image and 1,-1 for a perfectly positively or negatively correlated image.

3.2.6. Mean

Mean is the sum of the values divided by the number of values. The mean of a set of numbers x1,x2,x3......xn is typically denoted by $\overline{x}$ [22].

3.2.7. Variance

Variance is the measure of the dispersion of a set of data points around their mean value. It is mathematical expectation of the average squared deviations from the mean [23].

$$\text{Variance } (\sigma^2) = \frac{1}{N}\sum_{i=1}^{N}(xi-\mu)^2 \quad (7)$$

3.2.8. Standard Deviation

Standard deviation is a measure of how spread out the data set from the mean, it is denoted by σ [24].

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(xi-\mu)^2} \quad (8)$$

Hence, we consider these 9 features for merging with the spatial features for training and testing in the neural network classifiers.

3.3. Spatial Features

Spatial features are the feature that corresponds to the length and width of the facial action units. In order to evaluate spatial features the template is being created. The height and width of the template is being described in Table2.

Table 2. Size of image and templates

Image/Template	**Height(in pixel)**	**Width(in pixel)**
Input Image	256	256
Eye Template	15	30
Eyebrow Template	15	40
Mouth Template	20	45

The bounding rectangles are being drawn around the specified template according to its size. Once the bounding rectangles is being drawn its top and left coordinates is extracted to calculate the spatial features.

3.3.1. Template Matching

The template matching algorithm implemented we implemented in this project as follows:

Step 1: Send the respective image and its template as input to the template matching procedure

Step 2: Convert the image and template into gray scale by using rgb2gray ().

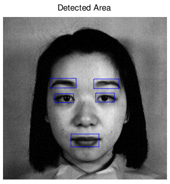
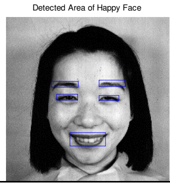
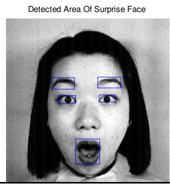
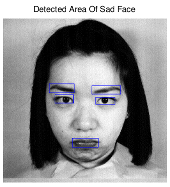
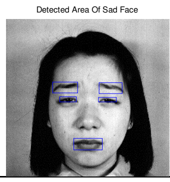
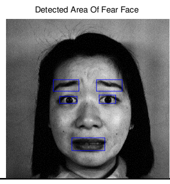
Step 3: Find the convolution of the original image and mean of the template required to be matched

Step 4: Then we find the correlation to get the highest matching of the template in the whole image.

Step5: Now, we find the four values, i.e. maximum of rows, maximum of columns, template height and template width to draw the bounding rectangles.

Table 3 defines the template matching of the different components to be matched of different faces.

Table 3. Matched Templates

Neutral Face	
Happy Face	
Surprise Face	
Angry Face	
Sad Face	
Fear Face	

3.3.2. Extracting Facial Points

There are in total 30 facial points [15] which are also known as facial characteristic points. Table 4 describes some of the facial point evaluation.In this way, we calculate 30 facial points, were,

lle, lre, llb, lrb, lmo:- left of left eye, right eye, left eyebrow, right eyebrow, mouth.
wle, wre, wrb, wlb:- width of left eye, right eye, left eyebrow,right eyebrow.
tle, tre, trb, tlb, tmo :- top of left eye, right eye, left eyebrow,right eyebrow, mouth.
hle, hre, hlb, hrb, hmo:- height of left eye, right eye, left eyebrow, right eyebrow, mouth

Table 4. Evaluation of the Facial Points

Region	**Facial point**	**X coordinate**	**Y coordinate**
Left eye	1	lle + wle	tle + hle*4/5
Right eye	2	lre	tre + hre/2
Left eyebrow	17	llb + wlb/2	tlb + hlb/3
Right eyebrow	18	lrb + wrb/2	trb + hrb/2
Mouth	23	lmo	tmo + hmo/2

3.3.3. Computing Spatial Features.

Once the 30 facial points are calculated the spatial features are being evaluated [16] as follows:

Openness of eyes:

((fp7_y-fp5_y) + (fp8_y-fp6_y))/2 (9)

Width of eyes:

((fp1_x-fp3_x) + (fp4_x-fp2_x))/2 (10)

Height of eyebrows:

((fp19_y-fp1_y) + (fp20_y-fp2_y))/2 (11)

Opening of mouth:

(fp26_y - fp25_y) (12)

Width of mouth:

(fp24_y – fp23_y) (13)

Were,

fp1_x, fp2_x, fp3_x, fp4_x, fp7_y, fp5_y, fp8_y, fp6_y, are the x, y coordinate position of the facial points detected around the eye template. Similarly the facial points fp1_y, fp2_y, fp19_y, fp20_y are the x, y coordinate position detected around the eyebrow template. Facial points fp23_y, fp24_y, fp25_yand fp26_y are the y coordinates of mouth template. Afterall these facial points are calculated the spatial features openness of eyes, width of eyes, opening of mouth, width of mouth and height of eyebrows are being calculated. These 5 features are merged with the statistical features for training and testing in neural network classifier.

3.4. Neural Network Classifier

A classification tasks usually involves separating data into training and testing sets. Each instance in the training set contains one class label and several attributes. The goal of classifier is to produce a model which predicts label of the test data given only the test attributes. There are different neural network classification techniques which are categorized into two type feedback and feed forward networks.

Back-propagation is a multilayer forward networks. In forward networks there is no feedback, hence only, a forward flow of information is present. There are various nets that come under the feed forward type of nets among all the most important type of network is the Back-Propagation networks. Figure 2.show an example of Back-Propagation network.

3.4.1. Training

There are generally four steps in the training process:

a. Assemble the training data.
b. Create the network object.
c. Train the Network.
d. Simulate the network response to new inputs.

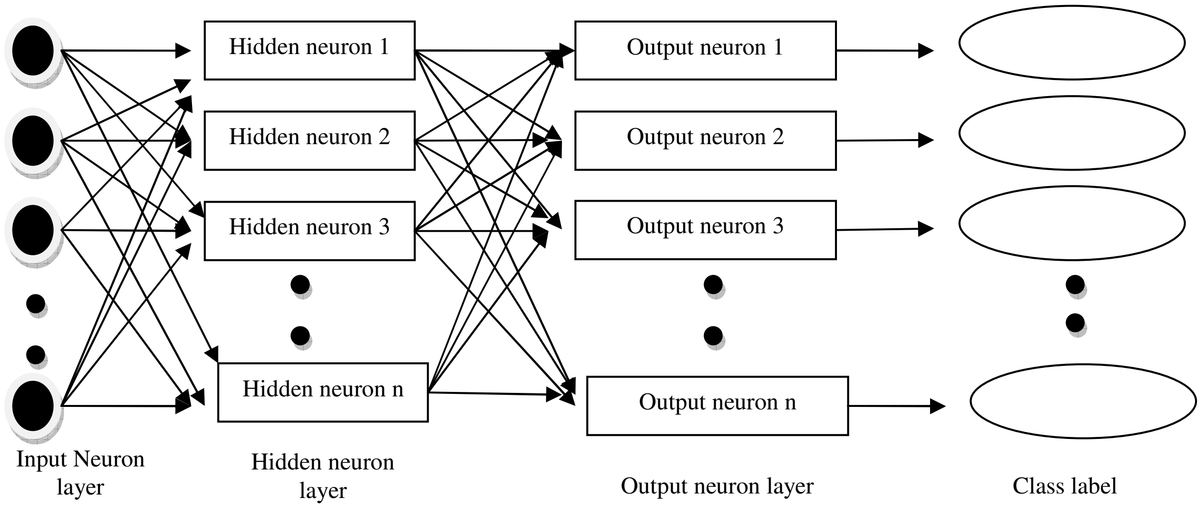


Figure 2. Architecture of Back Propagation Neural Network.

We name the training data set as 'train_data' and we simulate the network with the dataset named 'train_target'. Since, the code is implemented in matlab the back-propagation network [25] is created as follows:

net = newff(minmax(train_data),[100,7], {'tansig','purelin'},'trainlm'); (14)

Were equation (14) contains,
newff- create feedforward back-propagation network.
Minmax(train_data)- gives the number of neuron in the input layer, in our case it is '6', because of the six features
50- are the hidden neurons in the hidden layer.
7- are the output neurons.
'tansig'- transfer function of the hidden layer.
'purelin'- transfer function of the output layer.
'trainlm'- is the network training function that updates weight and bias values

3.4.2. Training Function

There are different types of training function among which 'trainlm' is the fastest back-propagation algorithm in the neural network toolbox. This training function, update weight and bias values according to 'Levenberg-Marquardt' optimization. The only drawback of this training function is that, it requires more memory than any other algorithm.

In our proposed technique we input all the merge features specified from equations (1)-(13) in equation (14) to obtain the classification results.

3.4.3. Epochs

Epoch is the step in training process. In our dataset number of epochs are 100.

3.4.4. Learning Rate

Learning rate is used to adjust the weights and biases of the network in order to move the network output closer to the targets. In our training learning rate is 0.05

3.4.5. Training Result

There are total 3 figures Figure 3, Figure 4, Figure 5.

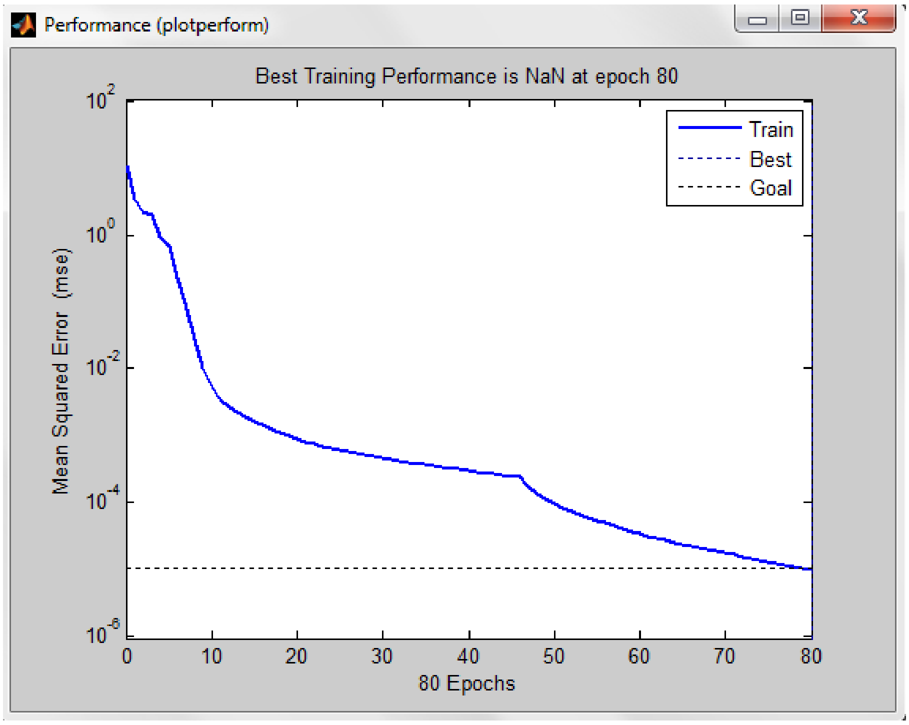


Figure 3.Performance plot

Figure 3, Performance plot is mapped between mean squared error and number of epochs that leads the train data to the best training performance. Figure 4, is the training state determines the position of gradient, mu and validation check when epoch is 80 at which network is completely trained. Figure 5, is the plot that tells the linear regression of targets relative to outputs. A straight linear line tells that the output data is exactly same as target data.

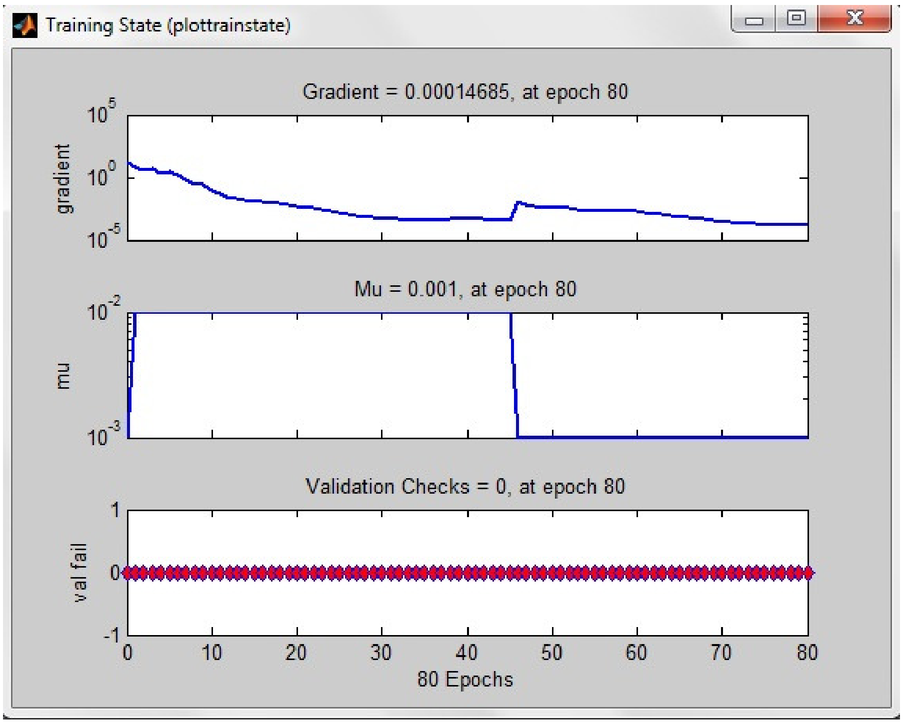


Figure 4. Training State Plot

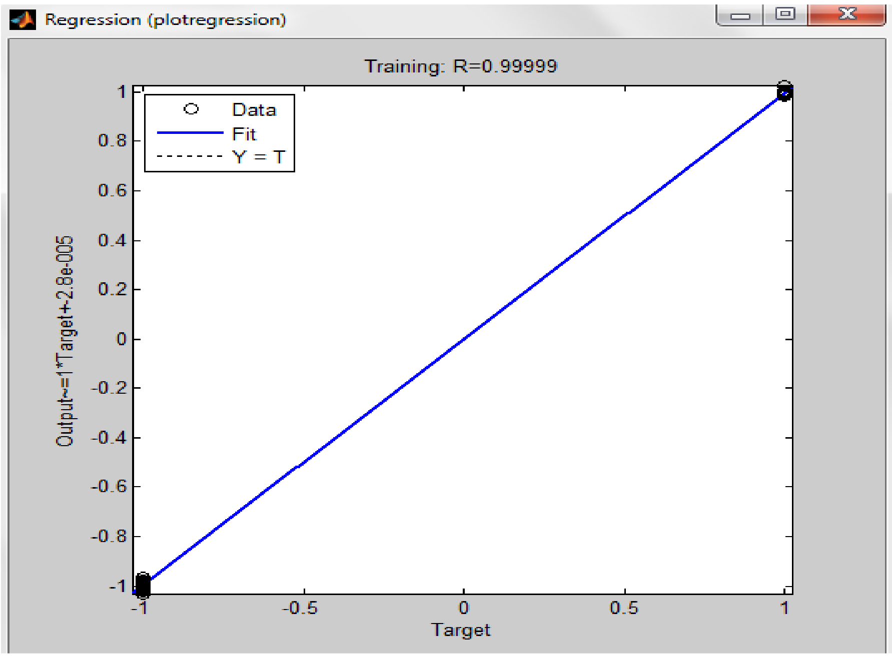


Figure 5. Regression plot

3.5. Testing

In testing phase we input the images to test whether it classify the face into respected class label or not. There are in total seven class labels hence there are seven output neurons each for particular expressions. Table5 shows the confusion matrix that are results obtained after the testing phase for classification. In total 10 images for each expression is input into the testing phase.

IV. Results

The results which we obtained from are plotted in figure 6 which describe that best classification rate is when all the 10 expressions are classified correctly, but as we gained 70% accurate results the correct classification rate describes the results we obtained.

V. Conclusions

Extensive efforts have been made over the past two decades in academia, industry, and government to discover more robust methods for classify the expressions of assessing truthfulness, deception, and credibility during human interactions. In this paper we proposed very simple techniques to evaluate features namely statistical and spatial features for training and testing in the neural network classifier. The total number of images provided for training is 154, i.e. 22 for each expression and for testing are 70, i.e. 10 for each expressions. The confusion matrix show the number of face out of 10 which is classified. In total 49 faces out of 70 is classified correctly. Hence 70% accurate classification is achieved using this research technique

Table 5. Confusion Matrix

Expression	Neutral	Happy	Surprise	Fear	Sad	Angry	Disgust
Neutral	8	-	-	-	1	-	1
Happy	-	5	-	-	2	1	2
Surprise	-	-	9	-	-	1	-
Fear	-	-	-	6	2	2	-
Sad	2	-	-	-	6	2	-
Angry	-	-	-	-	-	8	2
Disgust	1	-	-	-	2	-	7

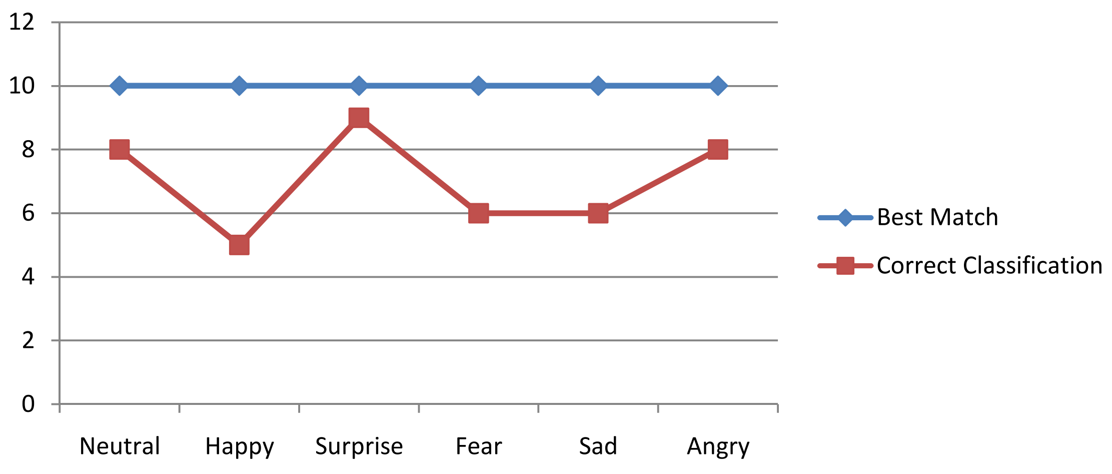


Figure 6.Classfication chart between the number of test face and the result obtained

VI. FUTURE SCOPE

The proposed work is on-going project hence there are different path to explore it, as we can use different features which can improve its accuracy from 70%. We can try some other network to increase its accuracy other than back-propagation network. We can also apply it for different database other than JAFFE databases.

REFERENCES

[1] Yuwen Wu, Hong Liu, HongbinZha, "Modeling facial expression space for recognition", National Natural Science Foundation of China (NSFC). Project No: 60175025, P.R.China.

[2] A. Mehrabian, "Communication without words," Psychology today, volume 2, no.4, pp.53-56, 1968.

[3] P. Ekman and Wallace V. Friesen, "Facial Action Coding System", consulting psychologist press, 1978.

[4] F. Kawakami, H. Yamada, S. Morishima and H. Harashima, "Construction and Psychological Evaluation of 3-D Emotion Space," Biomedical Fuzzy and Human. Sciences, vol.1, no.1, pp.33–42 (1995). 2427.

[5] M. Rosenblum, Y. Yacoob, and L. S. Davis, "Human expression recognition from motion using a radial basis function network architecture," IEEE Trans. on Neural Networks, vol.7, no.5, pp.1121-1138(Sept.1996).

[6] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," IEEE Trans. Pattern Analysis & Machine Intelligence, vol.22, no.12, pp.1424-1445(Dec. 2000).

[7] Y. S. Gao, M. K. H. Leung, S. C. Hui, and M. W. Tananda, "Facial expression recognition from line-based caricature," IEEE Trans. System, Man, & Cybernetics (Part A), vol.33, no.3, pp.407-412(May, 2003).

[8] Y. Xiao, N. P. Chandrasiri, Y. Tadokoro, and M. Oda, "Recognition of facial expressions using 2-D DCT and neural network," Electronics and Communications in Japan, Part 3, vo.82, no.7, pp.1-11(July, 1999).

[9] L. Ma, K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," IEEE Trans. System, Man, and Cybernetics (Part B), vol.34, no.4, pp.1588-1595 (2003).

[10] L. Ma, Y. Xiao, K. Khorasani, R. Ward, "A new facial expression recognition technique using 2-D DCT and K-means algorithms," IEEE.

[11] S.C.Tai and K.C.Chung, " Automatic Facial Expression Recognition using neural network," IEEE 2007

[12] Jyh-Yeong Chang and Jia-Lin Chen,"Facial Expression Recognition System Using Neural Networks", 1999 IEEE.

[13] L. Ma and K. Khorasani," Facial Expression Recognition Using Constructive Feedforward and Neural Networks", IEEE transactions on systems, man and cybernetics- part B: Cybernetics, vol.34, No. 3, June 2004.

[14] Chaiyasit, Philmoltares and Saranya," Facial Expression recognition using graph based feature and artificial neural network",

[15] "Extracting Facial Characteristic Points from expressionless face", from the book

[16] Jyh-Yeong Chang and Jia-Lin Chen," A Facial Expression Recognition System Using Neural Network", IEEE 1999.

[17] http://www.mathworld.wolfram.com/Kurtosis.html

[18] http://www.mathworld.wolfram.com/Skewness.html

[19] http://en.wikipedia.org/wiki/Moment-(mathematics)

[20] http://en.wikipedia.org/wiki/Entropy-(information_theory)

[21] http://www.mathworks.in/help/toolbox/images/ref/graycoprops.html

[22] http://www.purplemath.com/modules/meanmode.html

[23] http://www.mathworld.wolfram.com/StandardDeviation,html

[24] http://www.investopedia.com/terms/v/variance.asp

[25] http://www.dali.feld.cvut.cz/ucebna/matlab/toolbox/nnet/newff.html

Authors

Nazil Perveen born in Bilaspuron 04-dec-1987.Pursuing M.tech (computer technology) from National Institute of technology, Raipur, is India and B.E. in Computer Science and engineering from Guru Ghasidas University, Bilaspur, India with Honors (Gold Medalist) in 2009. Her research area basically belongs to Automatic Facial Expression Recognition and area related to its implementation

S. Gupta (**LMISTE, LMNIQR, MIEEE, MIET** and waiting for fellow membership of IE (I)) got her BE (Electrical) from GEC Raipur in 1988, M.Tech. (IPS) from VRCE Nagpur in 1998 and Ph.D (Power-Quality) from NIT Raipur in 2009. Her fields of interest are **power system, power quality and power electronics.** She has +20 yrs of teaching experiences in various subjects of EE. Presently she is working as Asso. Prof. in N I T Raipur.

Kesari Verma has completed her M. Tech degree from Karnataka University. She obtained her Ph.D. degree from Pt. Ravishankar Shukla University in Novel approach to predictive data modeling and pattern mining in temporal databases., currently working in National Institute of Technology, Raipur. She has 12 Year of teaching experience. She is member of CSI and life member of ISTE. She has 11 year of teaching experience. She is working in CGCOST sponsored project Iris Recognition system.