Volume 2, Issue 5, May 2012 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Automatic Expression Recognition Technique Using 2-D DCT and Eigen Vectors Based Neural Network

Shekhar Singh*

Assistant Professor,CSE Department

PIET, samalkha, Panipat, India

shekharsinghsengar@gmail.com

Akshat Jain

Assistant Professor,CSE Department

PIET, samalkha, Panipat, India

Abstract— ***Facial expression recognition has attracted much attention in recent years because of its importance in realizing highly intelligent human-machine interfaces. In this paper, we propose a new facial expression recognition technique that utilizes the 2-D DCT of full size facial images and eigenvectors based feed forward neural network. The neural network is designated to separate group of facial expressions with members "happy", "surprise", "anger", "shock", "disgust" and "sadness". This work is the first trial to use feed forward neural network, eigenvectors and 2-D DCT simultaneously within a single recognition task. To demonstrate the capability of the proposed recognition technique, we use two databases, including a recently constructed one, which contain 2-D front face images of 120 men and 80 women, respectively. Experimental results reveal that the new technique outperforms, on the whole, the simple vector matching and K-means based vector matching techniques and two recently developed methods using fixed size and constructive neural networks. The mean recognition rates of the new technique have been found as high as 98.5% and 95.8% for these two databases, respectively, which are the best results among those obtained for the same databases.***

Keywords— ***Feed forward neural network, 2-DCT, Eigenvectors, Facial expression, Recognition, Facial image, Pattern Recognition, Neural network, Expression classification.***

I. INTRODUCTION

Computer-based recognition of facial expressions has attracted much attention in recent years. The ultimate objective of facial expression recognition has been the realization of intelligent and transparent communications between humans and machines. The facial expression recognition will be a basic and indispensable component of technologies for the creation of human-like robots and machines.

Through detailed investigations of the characteristics of each expression in 2-D DCT based frequency domain, we have discovered that "anger" and "sadness" may be distinguished in better accuracy if only two of them are the subjects of classification. It has also been made clear that the facial expressions may be divided into two groups; with one group having two "easy" members, "smile" and "surprise", and the other group with two "challenging" expressions, "anger" and "sadness", and these two groups may be separated easily. It is these experiment-based insights that have motivated us to use neural network to perform the recognition task. This is the first work in integrating neural network, eigenvectors and 2-D DCT schemes in the facial expression recognition. Experiments using two facial image databases demonstrate that the proposed technique outperforms, as a whole, all the above-mentioned recognition methods for the same databases while attractive computational efficiency.

II. RELATED WORKS

To date, several facial expression recognition methods have been proposed. See for examples, [1]-[8] and the references therein. The facial action coding system (FACS) designated by Ekman [1] is well-known for facial expression description. In this system, the 3-D face is divided into 44 action units, such as nose, mouth, eyes, etc. The movements of muscles of these feature-bearing action units are used to describe any human facial expression of interest. The drawback of this method is that it requires 3-dimensional measurements and may thus be too complex for real-time processing. To alleviate this problem, a modified FACS using only 17 important action units was proposed in [2] for facial expression analysis and synthesis. However, 3-dimensional measurements are still needed. The complexity of the above modified FACS is reduced when compared with the original FACS, but some information useful for facial expression recognition may be lost. In recent years, facial expression recognition based on 2-dimensional digital images has been a focus of research [3] - [8]. In [3], a radial basis function neural network is proposed to recognize human facial expressions. The 2-dimensional discrete cosine transform (2-D DCT) is used to compress the entire face image and the resulting lower frequency 2-D DCT coefficients are used to train a one hidden- layer neural network using BP-based technique or constructive algorithm in [6], [7]. NN-based recognition methods have been found particularly promising [3], [6], [4], since neural network can easily implement complex mapping from the feature space of face images to the facial expression space. However, finding a proper network size has always been a frustrating and discouraging experience for neural network developers. This is dealt with by long and costly trial-and-error recursions. Motivated by these limitations and drawbacks, a recognition technique using constructive neural network has been proposed [7], where recognition rates of 98.5% and 95.8% have been obtained (without rejection) for the training and testing images, respectively. Constructive neural network are capable of systematically determining the proper network size required by the complexity of the given problem, while reducing considerably the computational cost involved in network training when compared with the standard BP-based training techniques [2], [6]. Recently, a recognition technique using 2-D DCT and the K-means algorithm has been proposed, which is generally efficient and provides higher recognition rates [8], but the number of standard vectors may become quite large such that the K-means based clustering and vector matching reduce the computational merits that could be expected.

In the entire above 2-D image based facial expression recognition methods, the confusion matrices reveal that (i) expressions "smile" and "surprise" are relatively easier to recognize and are slightly confused with other expressions, and (ii) "anger" and "sadness" are very often confused because of their similar characteristics, which lowers the overall recognition rate. Through detailed investigations of the characteristics of each expression in 2-D DCT based frequency domain, we have discovered that "anger" and "sadness" may be distinguished in better accuracy if only two of them are the subjects of classification. It has also been made clear that the facial expressions may be divided into two groups; with one group having two "easy" members, "smile" and "surprise", and the other group with two "challenging" expressions, "anger" and "sadness", and these two groups may be separated easily. It is these experiment-based insights that have motivated us to use neural network to perform the recognition task. This is the first work in integrating neural network, decision tree and 2-D DCT schemes in the facial expression recognition. Experiments using two facial image databases demonstrate that the proposed technique outperforms, as a whole, all the above-mentioned recognition methods for the same databases while enjoying attractive computational efficiency.

Figure 1. Facial Expression Color Image

III. THE PROPOSED EXPRESSION RECOGNITION TECHNIQUE

The new technique uses a neural network which is depicted in Fig.1 where each node is implemented by feed forward one-hidden-layer neural network. The neural network is trained such that the two groups are separated with as little confusion as possible. As expected, our experiments indicated that this could be done easily. The neural network trained to divide "happy", "surprise", "anger"," shock", "disgust" and "sadness". However, this is much easier than the separation of all the 6 expressions by a single neural network [2], [6], [7]. The proposed recognition technique consists of two phases: training and testing, which are described separately below.

Features of different images:

The features of facial images used in recognition must not be influenced by the appearance of any individual human. Therefore, pre-processing of the face images is needed in order to extract some information that is required by the recognition task and shared by all the expression images of the same category. One may make difference images by subtracting the neutral images from the expression images. The difference images are then expected to have much less to do with the appearance of the individual whose facial expressions are the subject of recognition. Therefore, the recognition task will become easier due to the use of difference images. It should be noted that the facial expression "neutral" will not be the subject of recognition.

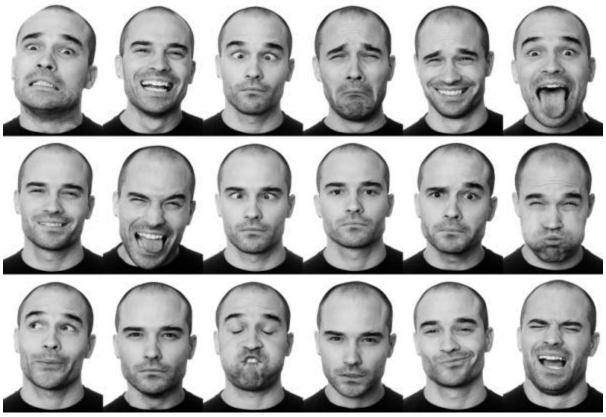

Figure 2. Facial Expression Gray Scale Image of man

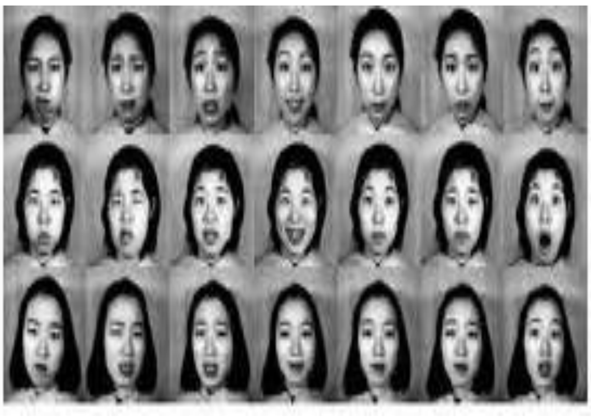

Figure 3. Facial Expression Gray Scale Image of woman

Eigenvector for facial expression:

Facial eigenvectors or, as they are sometimes called, Eigen pictures or Eigen faces, compactly represent whole faces, which is optimal for face reconstruction. This search is realized using cross-correlation in the frequency domain between the entire input image and eigenvectors. This increases detection speed over normal PCA algorithm implementation in the spatial domain. Using PCA Eigen vectors and Eigen features are identified and are then given to neural network algorithm to reduce the computational time. PCA calculates Eigen values with the help of covariance analysis. The Eigen vectors are sorted from high to low according to their corresponding Eigen values. The Eigen vector associated with the largest Eigen value is one that reflects the greatest variance in the image.

Data compression using 2-D DCT:

Obviously, it is very difficult for the classifier to recognize the facial expression from the difference images, as a difference image still has a large number of data. To facilitate the recognition, we need to compress the difference image to reduce data in a proper way, without losing the key features that play important role in the recognition task. The 2-D DCT used frequently in image compression is a powerful tool for this purpose. The 2-D DCT can reduce the number of data significantly by transforming an image into the frequency domain where the lower frequencies present relatively large magnitudes while the higher frequencies indicate much smaller magnitudes. That is to say, the higher frequency components can be ignored without damaging the key characteristics of the original difference image, as far as the facial expression recognition is concerned. The size of the facial expression images is $M \times N$. The 2-D DCT coefficients of a square block with size $L1 \times L2$ of the lower frequencies hold much of the information on the facial expressions, and are arranged as an input vector to the neural network for training or testing purposes. See Fig.2 for the image processing process.

A. Neural Network Training

The dimension of the input vector of the neural network is $M \times N$. One-hidden-layer neural network are considered in this work, which are trained by the program "TRAINGDX" provided in the MATLAB toolbox (TRAINGDX is a BP-based network training function that updates weight and bias values according to gradient descent momentum and an adaptive learning rate). There is only one output "logsig" node in the neural network and the threshold for expression classification is set to 0.6. "logsig" is also the activation function for all the hidden units.

The training parameters, such as the learning rate, number of epochs, etc. are properly selected. The input vector dimension, the number of hidden units, the initial weights, and the training parameters are systematically changed each time neural network is trained in order to achieve higher mean recognition rate of the 4 expressions. The neural network that presents the best recognition rate at each node is saved for testing.

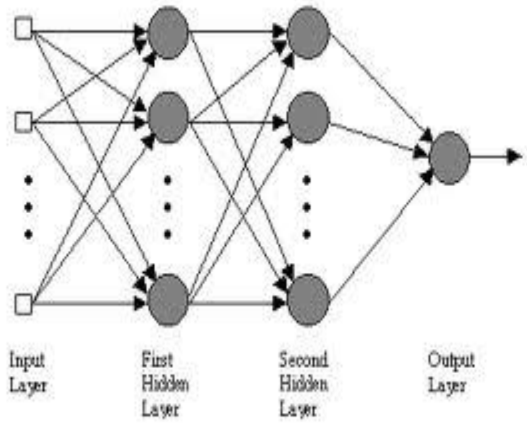

Figure4. Back propagation neural network

B. Testing

After the training is completed, a different portion of the facial expression images is used to test the performance of the trained neural network.

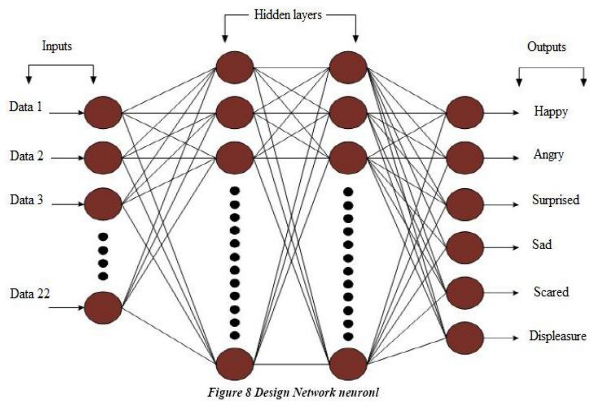

Figure5. Back propagation neural network model for facial expressions recognition

IV. EXPERIMENTAL RESULTS

Here, we first introduce a recently developed database that was constructed by using an efficient projection-based procedure. The database consists of images of 80 women, with each having 7 expression images, i.e., neutral, happy, anger, sadness, shock, disgust and surprise. A digital camera is used to take frontal images of each person. The images are incorporated into the computer where they are converted into gray images of size $M \times N$. Then, horizontal and vertical projections (i.e., summations of the gray-level values of the pixels on the same horizontal or vertical line) for the top two sub blocks are performed. The minimum points of the projection curves will be the candidates for the eye positions. To get stable results, DFT is used to smooth the curves (only 8

DFT coefficients are used in the IDFT). A set of projection curves is given in Fig.3, as an example. Clearly, the eye positions are correctly detected and determined. Next, the mouth is detected using similar projections applied to the bottom block. To obtain reliable mouth positions, compensation of white teeth is introduced before the projections are performed, by setting a proper empirical threshold such that the white teeth are detected and blackened. Based on the eye and mouth positions detected, the image is rotated and scaled if needed, and finally an image of size 256×256 is produced. The image samples shown in Fig.4 (2) belong to this new database.

The proposed technique is applied to database (a) and (b) that have front face images of 120 men and 80 women, respectively. Each individual has 5 facial expression images of size 256×256 ("neutral", "smile", "anger", "sadness", and "surprise"), with four of them ("smile", "anger", "sadness", and "surprise") being the subject of recognition. Sample images from databases (a) and (b) are given in Fig. 4. For each database, the images of the first 60 individuals are used for training and the remaining images are used to test the trained decision tree. Twenty (20) NNs were constructed for each specified pair of 2-D DCT block size and number of hidden units. All the NNs trained present fast convergence and the training process terminated within 500 epochs, with the summed square error (SSE) reaching the pre-specified goal or occasionally saturated. In Fig.5, examples for the SSEs of each node for database (b) are given. Extensive simulations revealed that the SSE goal does not affect the performance of the neural network obtained in terms of training recognition rate if set lower than 0.6. Figs. 6-11 show the maximum and mean recognition rates versus the number of hidden units of NN and the block size of lower frequency 2-D DCT coefficients. To achieve good performance, one needs to set the block size of 2-D DCT larger than 8 and the number of hidden units larger than 5. On the other hand, the recognition rates will not improve when the block size becomes larger than 36 and the net has more than 11 hidden units.

The mean testing recognition rate for database (a) is as high as 98.5%, which presents the highest record among all the previous techniques for the same database. For database (b), the testing mean recognition rate is 95.8%, which is similar to that of [8]. Obviously, the proposed technique presents, on the whole, improved recognition capabilities in comparison with those previous techniques.

Facial Expression	Happy	Surprise	Anger	Sadness	Shock	Disgust	Neutral
Happy	115	1	0	0	2	1	1
Surprise	0	113	1	1	4	1	0
Anger	0	1	116	1	1	1	0
Sadness	0	1	0	115	1	1	2
Shock	0	2	2	0	115	1	0
Disgust	0	1	0	1	1	116	1
Neutral	2	0	0	1	0	0	117

Table1. Confusion matrix of man facial expression recognition

Facial Expression	Happy	Surprise	Anger	Sadness	Shock	Disgust	Neutral
Happy	75	0	0	0	2	1	2
Surprise	0	74	1	0	4	1	0
Anger	0	3	73	0	3	1	0
Sadness	0	0	0	74	2	2	2
Shock	0	4	2	0	72	2	0
Disgust	0	2	1	1	1	74	1
Neutral	2	2	0	2	0	0	74

Table2. Confusion matrix of woman facial expression recognition

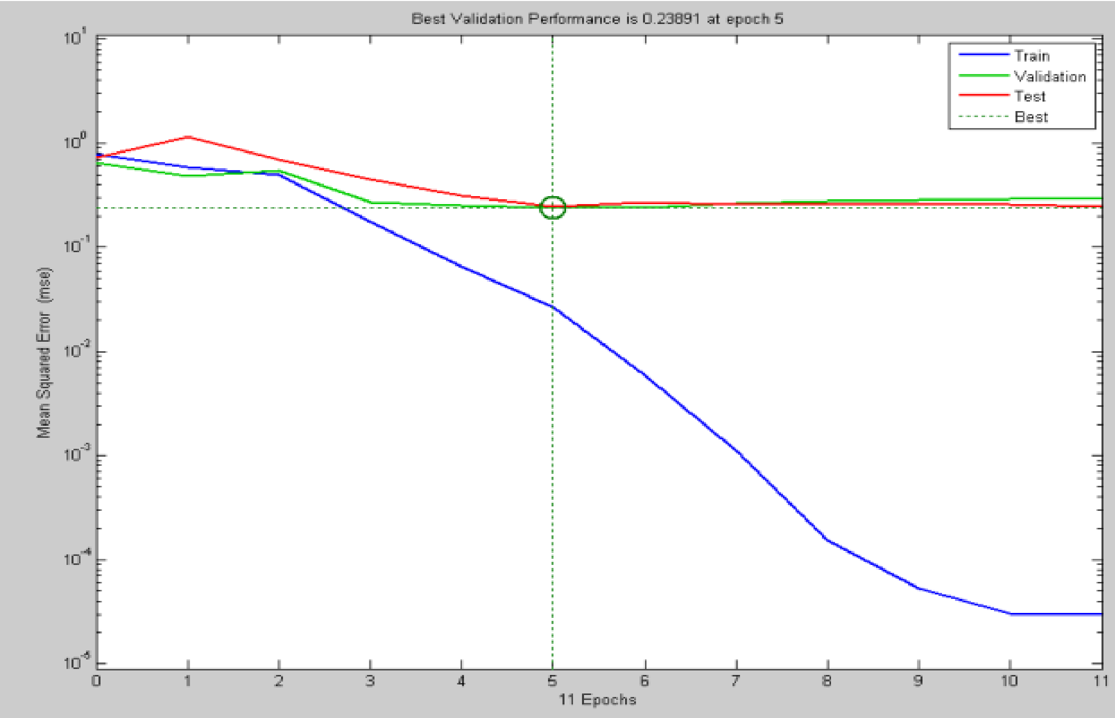

Figure 6: Performance plot

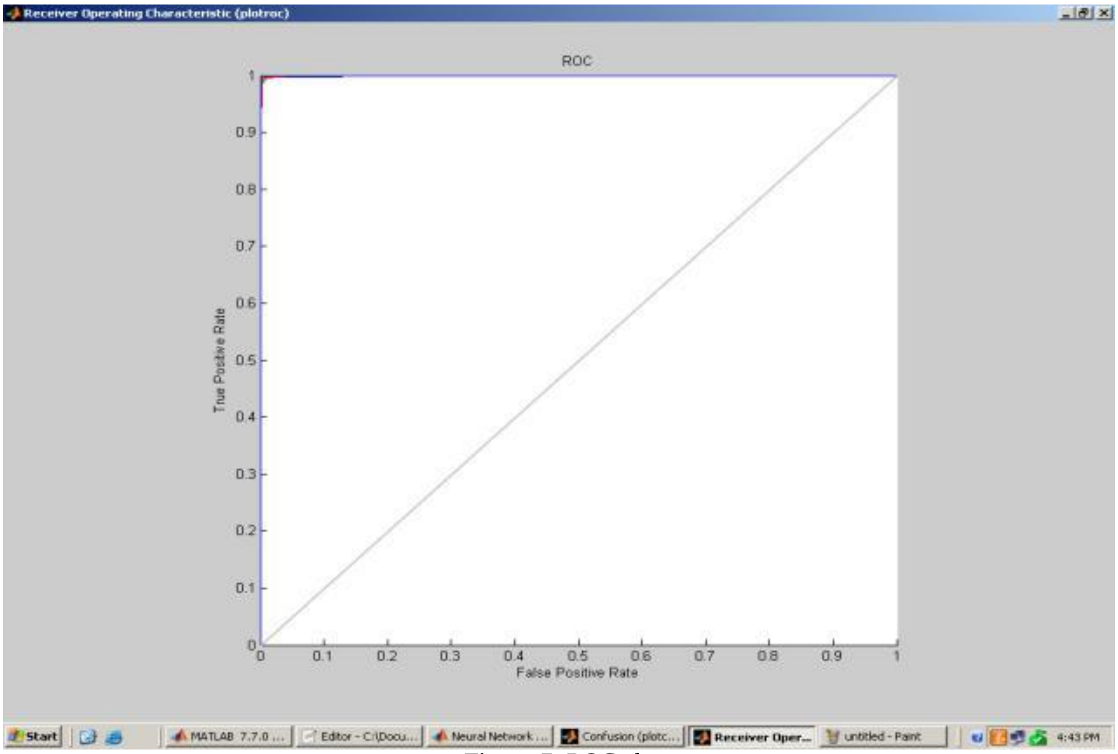

Figure 7: ROC plot

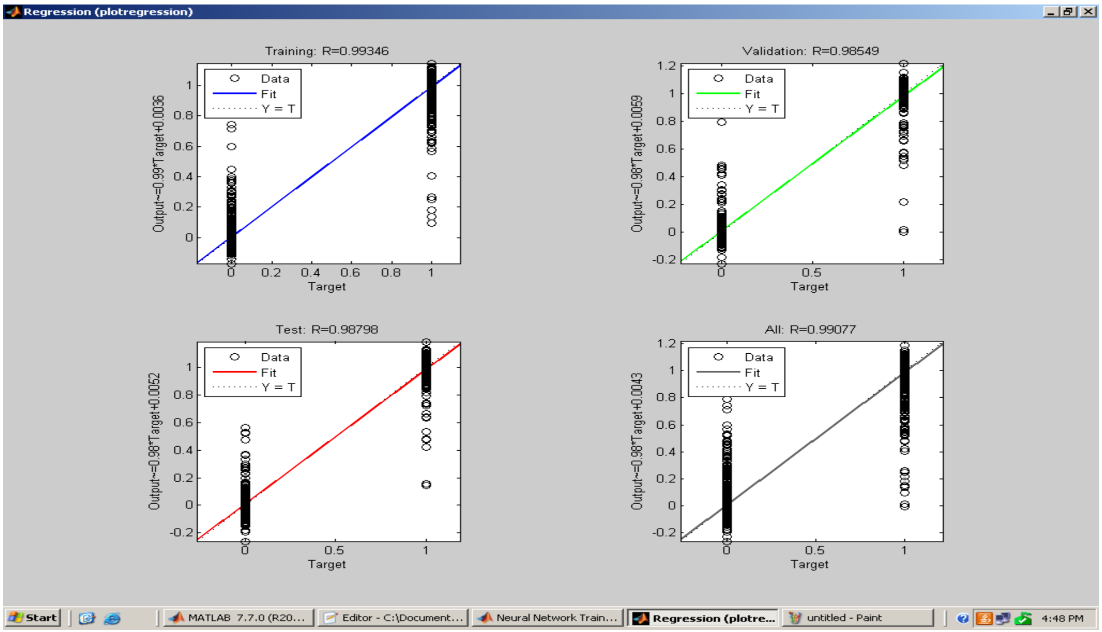

Figure 8: Regression plot

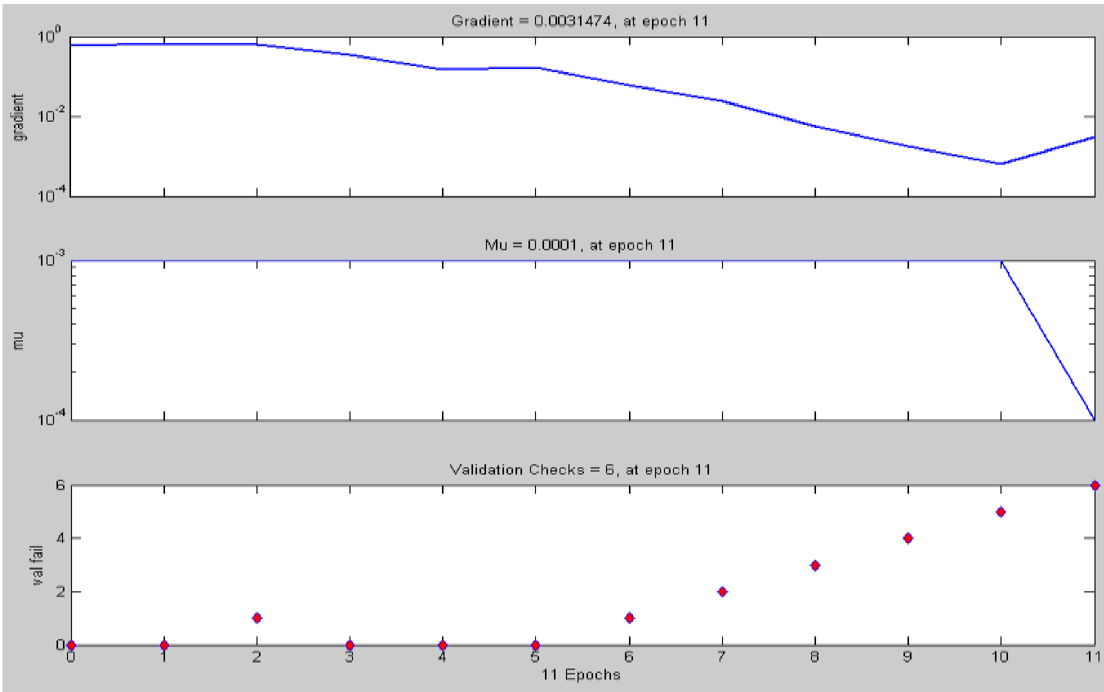

Figure 9: Training State Plot

Training and Testing:

The proposed network was trained with DCT matrix. When the training process is completed for the training data, the last weights of the network were saved to be ready for the testing procedure. The time needed to train the training datasets was approximately 4.60 minutes. The testing process is done for 500 cases. These 500 cases are fed to the proposed network and their output is recorded.

Performance plot: Performance plot show the training errors, validation errors, and test errors appears, as shown in the training process. Training errors, validation errors, and test errors appears, as shown in the following figure 6.

Receiver Operator Characteristic Measure (ROC) Plot: The colored lines in each axis represent the ROC curves. The *ROC curve* is a plot of the true positive rate (sensitivity) versus the false positive rate (1 -specificity) as the threshold is varied. A perfect test would show points in the upper-left corner, with

100% sensitivity and 100% specificity. For this problem, the network performs very well. The results show very good quality in the following figure 7.

Regression plots: This is used to validate the network performance. The following regression plots display the network outputs with respect to targets for training, validation, and test sets. For a perfect fit, the data should fall along a 45 degree line, where the network outputs are equal to the targets. For this problem the fit is reasonably good for all data sets, with R values in each case of 0.93 or above. The results show in the following figure 8.

Training State Plot: Training state plot show the deferent training state in training process and validation check graph. These plots also show the momentum and gradient graph and state in training process. The results show in the following figure 9.

V. CONCLUSIONS

In this paper, a new facial expression recognition technique is proposed which uses 2-D DCT, eigenvector and neural network to separate the facial expressions systematically. The 2-D DCT is applied to the difference images to compress and refine the features useful for the recognition task. The new technique has been applied to two databases of 120 men and 80 women facial expressions. Experimental results have demonstrated the superior effectiveness of the new method.

REFERENCES

[1] P. Ekman and W. Friesen, Facial Action Coding System, Consulting Psychologists Press, 1977.

[2] 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).

[3] 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).

[4] 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).

[5] 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).

[6] 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).

