A New Approach for Classification of Optimal Facial Expression Using Neural Network

Shital L. Pachpole[1], PROF. D U Adokar[2] ,DR. S. R. Suralkar[3]

[1,2,3]*Dept. of Electronics and Telecommunication, SSBT's COET, Jalgaon (Maharashtra),*

Abstract: Facial emotion plays a vital role for human interactive communication and also used in numerous real applications. Facial expression identification from frontal still images has in recent times become a hopeful investigation area. Their applications include human-computer interface, human emotion examination robot control, driver state surveillance and medical fields. This project aims to develop expression classification system of different principal facial expression by using Extended Cohn-Kanade database. Here, 2D-DCT (Two-Dimensional Discrete Cosine Transforms) is employed in addition to image statistics, texture and entropy parameters with a view to design a hybrid feature vector (FV). The single hidden layer feed-forward neural network has been used as a classifier in order to classify different expressions from frontal facial images. The learning algorithms such as scaled conjugate gradient (SCG) have been used. The contribution of this work is in the design of an optimal feature vector and also in the development of a simple neural network classifier containing a single hidden layer with carefully chosen number of neurons based on the experiments. For given databases, the proposed neural network classifiers deliver the maximum facial expression classification accuracy along with the least number of connection weights and biases.

Keywords: Facial Expressions, 2D DCT, FV, SCG, Neural network, Classifier

I. INTRODUCTION

Emotion is also known as expression or mood, and has always been used for showing human's feeling. The facial features and expressions are one of the groups of windows through which humans express emotions, and are critical to daily communication. Research in social psychology has shown that conveying messages in meaningful conversations can be dominated by facial expressions, and not by spoken words. The complex world of face emotional expression contains important information, and it can then be said to play a communicative role; it can be controlled to some extent and be intentionally used to acquire knowledge from the others and know their emotional state.

In this paper, a novel method is presented to classify human expressions from a sequence of input images that we have considered. Thus, each facial image is represented by a unique feature vector which corresponds to a typical emotion. Then obtained features of images are given as input to neural network for classification purpose. Transform domain features DCT (Discrete cosine transform) with image statistics are obtained. The algorithms used for training neural network is Scaled conjugate gradient algorithm..The parameters considered for classification are Confusion matrix and Cross entropy error. It is shown from the experimental results that the obtained accuracy is very good.

II. REVIEW OF RELATED WORK

A person's face is considered as the mirror of the mind. Facial expressions and the changes in facial expressions provide important information about affective state of the person, his temperament and personality, psychopathological diagnostic information, information related to stress levels, truthfulness etc. [3]. With growing terrorist activities all over the world, detection of potential troublemakers continues to be a major problem. Body language and facial expressions are the best ways to know the personality of a person and the response of a person in various situations. The facial expressions tell us about concealed emotions which can be used to verify if the information provided verbally is true. These expressions representing the emotional state of a person can serve an important role in the field of terrorism control and forensics.

Over the last few years, active research is being done to correlate movements of the face with emotional states. Darwin [7] first published *"The Expression of the emotions in Man and Animals" in* which he stated the three basic principles related to expressions and gestures in humans, viz., "(a) Principle of Serviceable associated habits, (b) Principle of Antithesis, and (c) Principle of actions due to constitution of nervous system, independently from the will, and independently to a certain extent habit" [7].

Ekman and Friesen [8] developed the Facial Action Coding System (FACS) to measure the facial behavior. In FACS, they used Action Units (AUs) based on the muscular activity that produces momentary changes in the facial expression. The system further classified an expression by correctly identifying the action unit or combination of action units related to a particular expression. A FACS database was created to determine the category or categories in which to fit each facial behavior. This database is available in the form of a FACS manual. Based on the action units, the researcher has to interpret the actual emotion.

Padgett [9], Hara and Kobayashi [10, 11], Zhang [12] and Zhao [13] used neural network approach for expression classification. They classified images into six or seven emotional categories. Padgett et al., [9] trained neural networks from the data of 11 subjects and tested with the data from one subject. The training and testing dataset was interchanged and new networks were trained and tested. A classification accuracy of 86% was achieved in this study. Hara and Kobayashi [10, 11] also used neural networks approach. The training dataset consisted of from data of 15 subjects (90 images) and these networks were tested using data from another 15 subjects. The classification accuracy achieved was 85 %. Zhang et al., [14] used the JAFFE data base which consists of 10 Japanese female subjects. Although an accuracy of 90.1% was achieved; same data was used for training and testing. A 100 % recognition rate was achieved by Zhao et al., [13] who used the Ekman and Friesen database [15], but they used the same data for training and testing. Khan et al [16] used thermal methods to quantify the facial expressions. He could achieve an accuracy of 56 %.

III. PROPOSED METHODOLOGY

The research problem undertaken is to develop a neural network based classifier system for classification of principal emotions (expressions) from human facial images of extended cohn-kande dataset. All these images are frontal still gray images of the human faces, which will be used to develop an optimal neural network based classifier system with a view to correctly classify the expressed emotion by a person.

An image feature extraction and generation scheme has been suggested with a view to achieve the best classification accuracy. Features based on 2D DCT are augmented with features with regard to image statistics, entropy, texture, homogeneity, etc., so that every image is represented by a unique feature vector. Based on the annotations available and the visual inspection of the frontal still facial image of a person, the exact emotion is classified, which is the target emotion. The facial image database is transformed to a knowledgebase to be used by a neural network, where a feature vector is followed by the expressed emotion .

Obviously, the performance of the neural network is always the best on the training dataset, because it is this dataset only which has been presented iteratively to the neural network in order to achieve learning and estimate the optimal values of the connection weights and biases of the neural network. Therefore, examination of the performance of the neural network on validation as well as the testing dataset is more important as neural network never sees these datasets during the process of learning. Cross-validation dataset is used for termination of the neural network training and as training proceeds, at the end of every iteration, it is ensured that both the error on training dataset and that on validation dataset go on decreasing with respect to the number of iterations (epochs). Though, the magnitude of these errors might be significantly different. It is generally noticed that the error on the training dataset is always much lower than that on validation dataset.

The magnitude of the error gradient and the number of validation checks are used to terminate the training. The number of validation checks represents the number of successive iterations that the validation performance fails to decrease. If this number reaches 6 (the default value), the training will stop.

All input features are normalized with zero mean and unit variance, so that the neural network based classifier models can run faster and better. Fig. 1 shows some sample facial images of Cohn-Kande database. In the following figure the face expression is as following order ANGRY DISGUST HAPPY SAD SURPRISE.

Fig.1. Samples of images from the cohn-kande database along with expressions (emotions)

For every setting of the number of neurons in the hidden layer, ten different trials of the neural network are run and during every trial, the results of the neural network are found to be drastically different most of the time. This is because of randomness in initialization of connection weights and biases and randomness in data partitioning into training, CV (cros-validation) and testing subsets at the beginning of every trial run. In addition, as this problem represents learning from data, the problem always has multiple non-optimal solutions (multiple local minima) and one optimal solution (global minimum). When training of the neural network proceeds, the learning algorithm often might get trapped to any one of the local minima and during every trial it can go on finding different local minima and thus further worsens the matter of finding a global optimal solution. Because of these complex situations, it is impossible to maintain exact reproducibility in the results produced by the same configuration of the neural network. For training of the neural network Scaled Conjugate Gradient backpropagation (SCG) algorithm is used.

Because of the fact that input feature values are bipolar, 'tansig' activation function is used for the hidden as well as output layer neurons. Cross entropy error criterion is more appropriate for classification. The crux of the best design is the simplicity and the minimum time and space complexity of the neural network based classifier. The network with minimum number of free parameters (connection weights and biases) should be employed.

The performance of the neural network based classifier is recorded for all ten different trials (runs) with respect to cross-entropy error on training dataset, CV dataset and Testing dataset, Average Classification Accuracy on Training dataset, Average Classification Accuracy on CV dataset, Average Classification Accuracy on Testing dataset and Overall Average Classification Accuracy. Finally, the best performance measures are highlighted with an emphasis to the Average Classification Accuracy in comparison with cross-entropy error, because we are solving a classification problem. As overall average classification accuracy reflects the average classification accuracy on training, validation and testing datasets; the neural network configuration with the highest average overall classification accuracy is chosen.

IV. EXPERIMENTAL RESULT

For 70 neurons in the hidden layer, the neural network is re-trained ten times with different random initialization of connection weights including biases and random partitioning of the data into training, cross-validation and testing datasets (Training: 90%, Cross-validation: 5 % and Testing 5 %) Among ten different trials, during Trial 5, the maximum average overall classification accuracy is observed. The overall classification accuracy observed is 98.1%. The confusion matrix shows the overall classification accuracy in bellow figure 2.

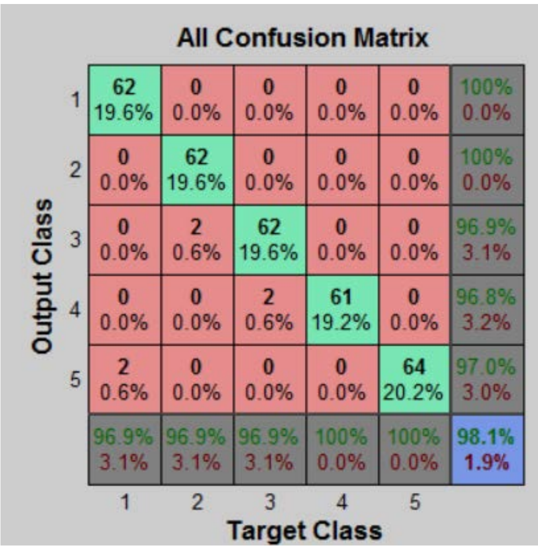

Fig. 2. All confusion matrix

The Fig. 2 shows all confusion matrix , in which green square shows correct classified samples while red square shows misclassified samples

The receiver operating characteristics in fig.3 shows the true positive rate versus false positive rate for training, testing and validation dataset.

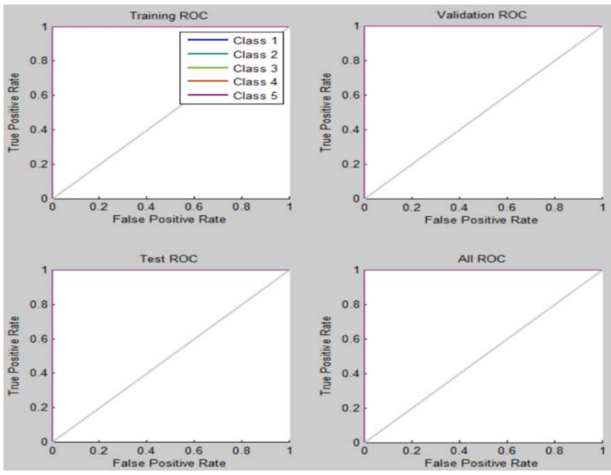

Fig. 3. Receiver Operating Characteristics

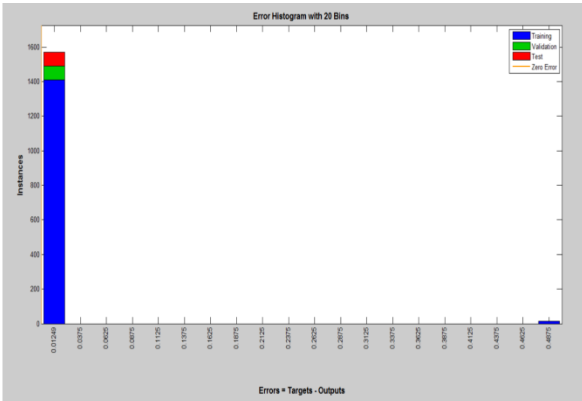

Fig. 4. Error Histogram with 20Bins

In above fig 3 we can see the error histogram for 20 bins i.e. 20 intervals.

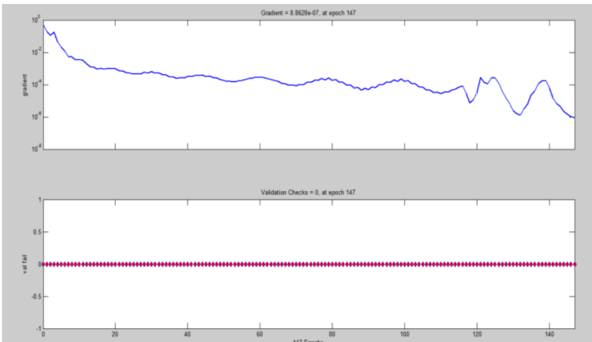

Fig. 5. Training State.

In above fig.5 validation checks are zero at epoch 147.

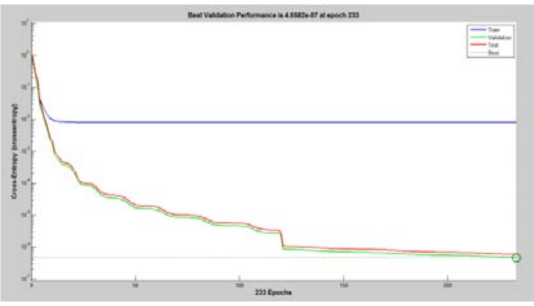

Fig. 6. Plot showing Cross-Entropy Error versus number of Epochs for training, cross validation and testing data

In above graph shown in Fig. 6, we can see that the best validation performance is 4.6582e-07, which is obtained at epoch number 233. So at this instant, the training of the Neural Network is terminated.

4.2 Design of Neural Network Based Classifier.

For the development of a classifier based on neural network, feature vectors, $\mathbf{FV_{DCT}}$ has been employed. The feature vector, $\mathbf{FV_{DCT}}$ based on 2D – DCT for Cohn-Kanade Database is comprised of 75 features, such that $\mathbf{FV_{DCT}}$ = [DCT1, DCT2, DCT3, …,DCT64, Average, SD, Entropy, moment2, moment3, median, variance, contrast, correlation, energy, homogeneity].

First, input features based on 2D – DCT have been applied to the single hidden layer feed-forward neural network. In case of $\mathbf{FV_{DCT}}$, the maximum average overall classification accuracy is obtained for a neural network with a single hidden layer containing 70 neurons and trained with scaled conjugate gradient backpropagation algorithm. The maximum average overall classification accuracy is observed as only 98.1 % .

For 70 neurons in the hidden layer, the neural network is re-trained ten times with different random initialization of connection weights including biases and random partitioning of the data into training, cross-validation and testing datasets (Training: 90%, Cross-validation: 5 % and Testing 5 %). Among ten different trials, during Trial 1, the maximum average overall classification accuracy is observed.

It is seen from the Table no. 1 that 62 emotions are correctly classified as surprise, two emotions are incorrectly classified as angry, resulting into the classification accuracy for surprise as 62/64, i.e., 96.9%. It is also observed that 62 emotions are correctly classified as sad and two emotions are incorrectly classified as happy. Thus, out of total 64 emotions, only 62 emotions are correctly classified as sad, resulting into the classification accuracy for sad of 62/64 i.e., 96.9%. It is also noticed that 62 emotions are correctly classified as happy and two emotions are incorrectly classified as disgust, thus, leading to the classification accuracy for happy as 62/64, i.e., 96.9%.

Table 1 No. of correct classifications by the trained Neural Network

No. Of Recognized Expressions Expressions (Output)	**No. Of Desired Expressions (Target)**				
	Surprise	**Sad**	**Happy**	**Disgust**	**Anger**
Surprise	62	0	0	0	0
Sad	0	62	0	0	0
Happy	0	2	62	0	0
Disgust	0	0	2	61	0
Anger	2	0	0	0	64
Correct Recognition	62/64	62/64	62/64	61/61	64/64
Classification Accuracy	96.9%	96.9 %	96.9%	100%	100% +
Overall Average Classification Accuracy	98.1%				

In addition, it is also observed that all 61 emotions are correctly classified as disgust, thus resulting into the classification accuracy with respect to disgust as 100 %. Furthermore, 64 emotions are correctly recognized as angry and the classification accuracy for angry 64/64, i.e., 100%.

V. CONCLUSION

The novelty of the paper has been not only in the selection and generation of the most suitable input features from images resulting into the optimal feature vector in the anticipation of the best classification performance but also in the design of the optimal neural network classifier. Maximum average overall classification accuracy 98.1% is achieved for hybrid input features based on DCT, i.e., FV_{DCT} (Total number of features = 77) using a neural network with 70 neurons in the hidden layer and trained with Scaled Conjugate Gradient algorithm . This is the best ever reported maximum average overall classification accuracy on the Cohn-Kanade database. In view of this, it is recommended to use the suggested optimal feature vector comprising of input features based on uniform DCT combined with image statistics and other features such as texture, entropy, etc. for training a neural network with a single hidden layer comprising of only 70 neurons

REFERENCES

[1] Ekman P., Huang T. S., Sejnowski T. J., Hager J. C., “Final Report to NSF of the Planning Workshop on Facial Expression Understanding” ,1992.

[2] Alvino C., Kohler C., Barrett F., Gur R. E., Gur R. C.,Verma R., “Computerized measurement of facial expression of emotions in schizophrenia”, Journal of Neuroscience Methods , 2007, 163, pp. 350-361.

[3] Wang P., Barrett F., Martin E., Milonova M., Gur R. E., Gur R. C., Kohler C., Verma R., “Automated video-based facial expression analysis of neuropsychiatric disorders” Journal of Neuroscience Methods, 2008, 168, pp. 224-238.

[4] Dinges D. F., Venkataraman S., McGlinchey E. L., Metaxas D. N., “Monitoring of facial stress during space flight: Optical computer recognition combining discriminative and generative methods”, ActaAstronautica , 2007, 60, pp. 341 – 350.

[5] Padgett C., Cottrell G.W., “Representing Face Images for Emotion Classification,” Proceedings of Conference on Advances in Neural Information Processing Systems, 1996, pp. 894-900.

[6] Z. Zhang, “Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron”, International Journal of Pattern Recognition and Artificial Intelligence, 1999, 13, pp. 893-911.

[7] G. Guo and C.R. Dyer, “Learning from Examples in the Small Sample Case: Face Expression Recognition,” IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 35, no. 3, June 2005, pp. 477-488.

[8] Z. Wenming, Z. Xiaoyan, Z. Cairong, and Z. Li, “Facial Expression Recognition Using Kernel Canonical Correlation Analysis (KCCA),” IEEE Trans. Neural Networks, vol. 17, no. 1, Jan. 2006, pp. 233-238.

[9] Zhengdong, S. Bin, F. Xiang, and Z. Yu-Jin, “Automatic Coefficient Selection in Weighted Maximum Margin Criterion”,Proc. 19th Int’l Conf. Pattern Recognition, 2008, pp. 1-4.

[10] Shan, S. Gong and P.W. McOwan, “Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study,” Image and Vision Computing, vol. 27, 2009, pp. 803-816.

[11] Xun Wang, Xingang Liu, Lingyun Lu, ZhixinShen, "A New Facial Expression Recognition Method Based on Geometric Alignment and LBP Features," 2014 IEEE 17th International Conference on Computational Science and Engineering (CSE), 19-21 Dec.2014, pp. 1734-1737.

AUTHOR'S PROFILE

[1] **Shital Pachpole** has received her Bachelor of Engineering degree in Electronics and Telecommunication engineering from Government College of Engineering Jalgaon, Year 2011. At present she is pursuing Master of Engineering in Digital Electronics.

[2] **Prof. D U Adokar** has received his Master of Engineering degree in Electronics and Telecommunication engineering and working as an associate professor in SSBT’s COET Jalgaon.

[3] **Dr. S R Suralkar** has received his doctorate and working as Head of Department of Electronics and Telecommunication Engineering in SSBT’s COET Jalgaon.