International Journal Of Computer Science And Applications Vol. 1, No. 2, August 2008
ISSN 0974-1003

Performance Comparison of Facial Feature Extraction Techniques In Designing Human Emotion Recognition System Using Optimal PCA Neural Network

G.U.Kharat
Prof. & Dean (Research & Development), Anuradha Engineering College, Chikhli (M.S.), India.
gukharat@rediffmail.com

Dr. S.V.Dudul
Prof & Head Department of Applied Electronics.SGB Amravati University Amravati. (M.S.), India.
dudulsv@rediffmail.com

ABSTRACT

This research aims at developing "Humanoid Robots" that can carry out intellectual conversation with human beings. The first step in this direction is to recognize human emotions by a computer using neural network. In this paper all six universally recognized basic emotions namely angry, disgust, fear, happy, sad and surprise along with neutral one are recognized. Various feature extraction techniques such as Discrete Cosine Transform (DCT), Fast Fourier Transform (FFT), Singular Value Decomposition (SVD) are used to extract the useful features for emotion recognition from facial expressions. Principle Component Analysis (PCA) is used for emotion recognition using the extracted facial features and the performance of various feature extraction technique is compared. Authors achieved 100% recognition accuracy on training and test dataset.

Keywords: Discrete Cosine Transform (DCT), Fast Fourier Transform (FFT), Singular Value Decomposition (SVD), Principle Component Analysis (PCA), Machine Intelligence.

1. INTRODUCTION

It is highly expected that computers and robots will be used more for betterment of our daily life. Information – Computerized Society expect a harmonious interaction or heart to heart communication between computers and / or robots and human beings. For its realization it seems to be necessary that computers and robots will be implemented with artificial mind that enables them to communicate with human beings through exchanging not only logical information but also emotional one. The first step to realize humanoid robot is to recognize human emotions. Mehrabian [1] indicates that the verbal part (i.e. spoken words) of a message contributes only for a 7% of the effect of the message, the vocal part (i.e. voice information) contributes for 38% while facial expressions of the speaker contributes for 55% of the effect of the spoken message. Hence in order to develop "Active Human Interface" that realizes heart to heart communication between intelligent machine and human beings we are implementing machine recognition of human emotions from facial expressions.

Affective computing addresses issues relating to emotion in computing and has been pioneered by the work of Picard at MIT [2]. Picard describes how "Affective interaction can have maximum impact when emotion recognition is available to both man and machine" and goes on to say if one party can not recognize or understand emotion then interaction is impaired [3]. The problem of recognizing facial expressions had attracted the attention of computer- vision community [4-11]. Bassili [12] suggested that motion in the image of the face would allow emotions to be identified even with minimal information about the spatial arrangement of features.

FACS is developed by Ekman and Frison [18] using action potentials for emotion recognition. Essa and Petland [13] and Essa [14] proposed FACS+ model extending Facial Action Coding System (FACS) model to allow combine spatial and temporal modeling of facial expressions. Optical flow computations for recognizing and analyzing facial expressions are used by [5, 7, 9, 11 and 15 to 28]. Anthropometric facial points are used for feature extraction to recognize emotions [9, 11 and 29].

This paper provides the simplest approach of using DCT, FFT and SVD for extraction of facial features and their performance comparison with optimally designed PCA.

2. FACIAL EXPRESSION DATABASE

Facial expression database in six universally recognized basic emotions and neutral one is collected from Japanese female database. Ten expressers posed 3 to 4 examples of each of the six emotions along with neutral one for a total of 219 images of facial expressions. This data was prepared when expresser look into the semi reflective plastic sheet towards camera. Hairs were tied away o expose all expressive zones of the face. Tungsten lights were positioned to create even illumination on the face. The box enclosed the region between camera and plastic sheet to reduce back reflections. The images were printed in monochrome and digitized using flatbed scanner. Sample images are shown in figure 1. Total 210 images are selected for our experiment.

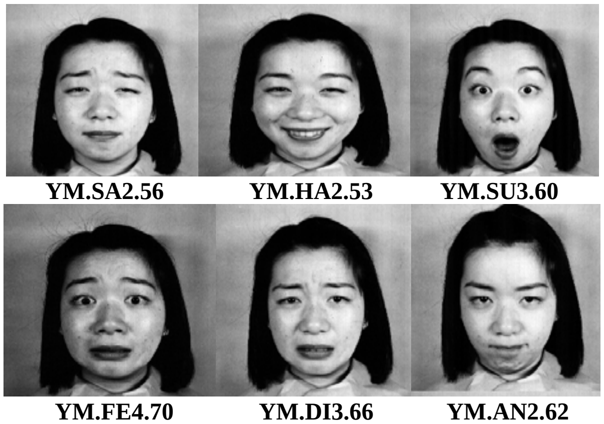

Figure 1. Images of Japanese females in various emotions.

3. COMPUTER SIMULATION EXPERIMENT

3.1 Feature Extraction Using DCT

The authors have developed a program in MATLAB to obtain DCT and statistical parameters namely energy, entropy, variance, standard deviation, contrast, homogeneity and correlation of an image. An optimal feature vector is obtained containing the features extracted by DCT and physical parameters of each image. Thus dataset for all 210 images is prepared to feed to Neural Network for emotion recognition.

3.2 Feature Extraction Using FFT

A program in MATLAB is developed to obtain FFT and statistical parameters namely energy, entropy, variance, standard deviation, contrast, homogeneity and correlation of an image. An optimal feature vector is obtained containing the features extracted by FFT and statistical parameters of each image. Thus dataset for all 210 images is prepared to feed to Neural Network for emotion recognition.

3.3 Feature Extraction Using SVD

Similarly a program in MATLAB is developed to obtain SVD and statistical parameters namely energy, entropy, variance, standard deviation, contrast, homogeneity and correlation of an image. An optimal feature vector is obtained containing the features extracted by SVD and statistical parameters of each image. Thus dataset for all 210 images is prepared to feed to Neural Network for emotion recognition.

4 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis networks (PCAs) combine unsupervised and supervised learning in the same topology. Principal component analysis is an unsupervised linear procedure that finds a set of uncorrelated features, principal components, from the input. A MLP is used to perform the nonlinear classification from these components. This panel allows you to select the number of principal components to extract from the input, the type of PCA to perform, and the number of hidden layers in the MLP.

PCA is a data reduction method, which condenses the input data down to a few principal components. As with any data reduction method, there is the possibility of losing important input information. The number of principal components selected will be a compromise between training efficiency (few PCA components) and accurate results (a large number of PCA components). It is not possible to provide a general formula for selecting an appropriate number of principal components for a given application.

Both learning rule choices are normalized implementations of the Hebbian learning rule. Straight Hebbian learning must be utilized with care, since it may become unstable if the inputs are not properly normalized. For this reason, it is not given as an option within this panel. The two more robust choices are Oja's and Sanger's implementations of the Hebbian principle. Between the two, Sanger's is preferred for PCA because it naturally orders the PCA components by magnitude.

5 PCA FOR EMOTION RECOGNITION

The generalized procedure for emotion recognition from facial expressions using different feature extraction techniques is shown in figure 2. We have used DCT, FFT and SVD for feature extractions and PCA for emotion recognition.

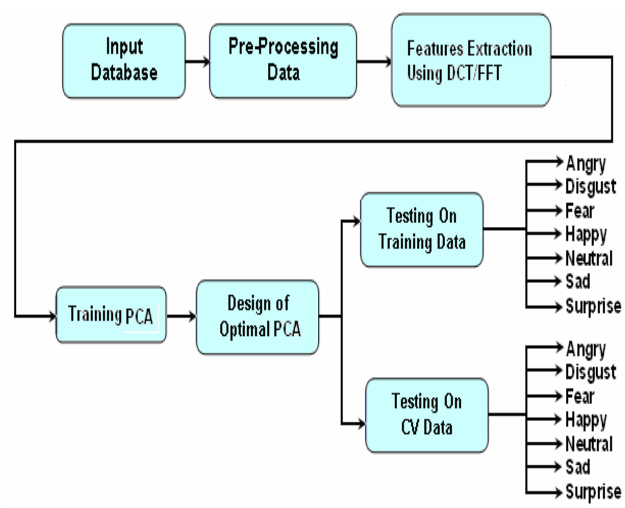

Figure 2. Procedure for Emotion Recognition.

i) Human Emotion Recognition Using DCT

The randomized data is fed to the PCA neural network and is trained for different hidden layers. It is observed that PCA with single hidden layer gives better performance. The number of Processing Elements (PEs) in the hidden layer is varied. The network is trained for 20 principal components (PC) and average minimum MSE on CV data is obtained when 19 PEs are used in hidden layer as indicated in figure 3. Rigorous experimentation is done by varying number of exemplars for training and cross validation data using 10 PEs in hidden layer. The network is again trained three times. Average Minimum MSE on Trained and CV data set along with average classification accuracy is calculated and is shown in figure 4 & 5. It is observed that best results obtained when 15% exemplars are used for cross validation (CV) data and 85% for training.

With 15% CV and 85% train data various transfer functions are used for training the network and average minimum MSE on training and CV data is measured (Figure 6). Average classification accuracy for different transfer function is plotted in figure 7. It is observed that tanh is most suitable transfer function. Using tanh transfer function the PCA neural network is trained using learning rules namely Sangers full, Ojas full, Momentum, Conjugate-Gradient (CG), Quick Propagation (QP), Delta Bar Delta (DBD). Minimum MSE on training and CV data set is measured and is indicated in figure 8. Finally network is tested on training and testing dataset (figure 9). It is observed that Sangers full and momentum are most suitable learning rules for our neural network.

From above experimentation selected parameters for PCA neural network are given below.

PCA NN (71-19-7), Number of epochs = 5100
Unsupervised learning control epochs = 100
Supervised learning control epochs = 5000
Number of principal components = 20
Exemplars for cross validation = 15%
Exemplars for training = 85%
Termination is at 500 epochs without improvement.

Layers of PCA	Transfer Function	Learning Rule	Step size	Momentum
Input layer	Tanh	Sangers full	-	-
HL#1	Tanh	Momentum	0.1	0.7
Output layer	Tanh	Momentum	0.01	0.7

Time elapsed per epochs per exemplar = 0.0129 mSec.
Number of free parameters (P) of PCA = 1508
Number of exemplars in training dataset (N) = 178
(N/P) ratio = 0.118.
Finally designed PCA is tested on training and CV dataset and results are shown in table 5 to 8.

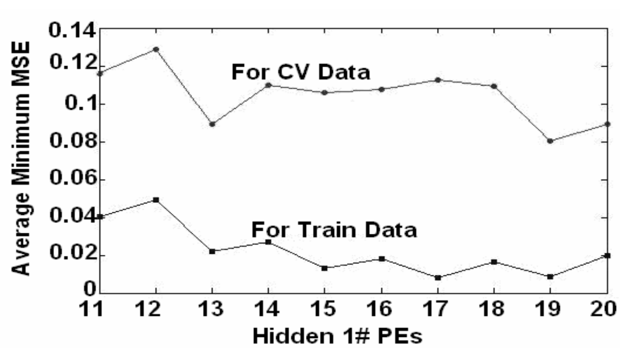

Figure 3 Graph showing variation of average minimum MSE on Training and CV dataset with number of processing elements in the hidden layer.

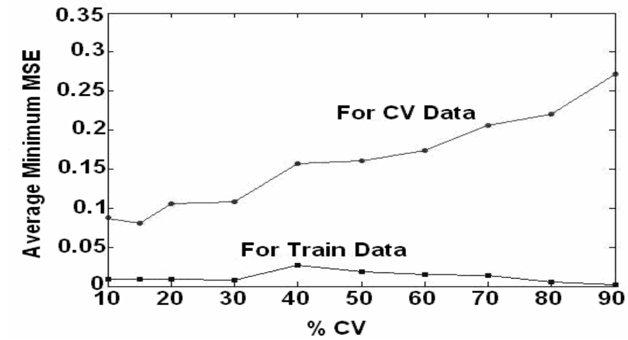

Figure 4. Graph showing variation of average minimum MSE on Training and CV dataset with % of CV data.

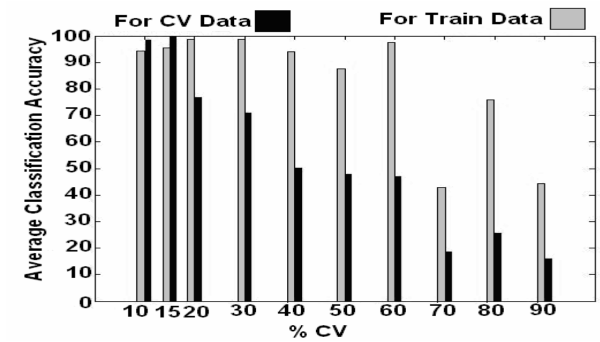

Figure 5. Graph showing variation of % average classification accuracy with % of CV data.

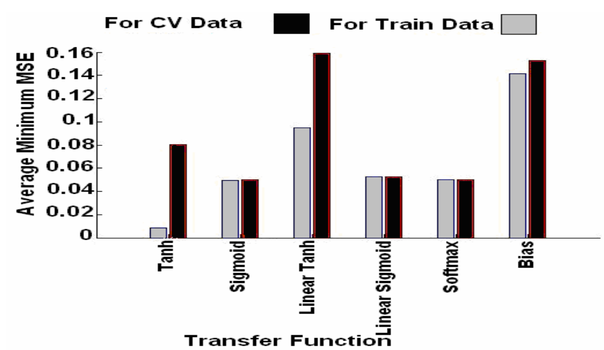

Figure 6: Graph showing variation of average minimum MSE on Training and CV dataset with Transfer function.

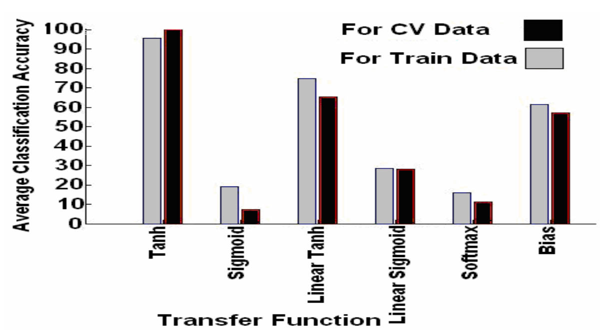

Figure 7: Graph showing variation of % average classification accuracy with Transfer function.

ISSN 0974-1003

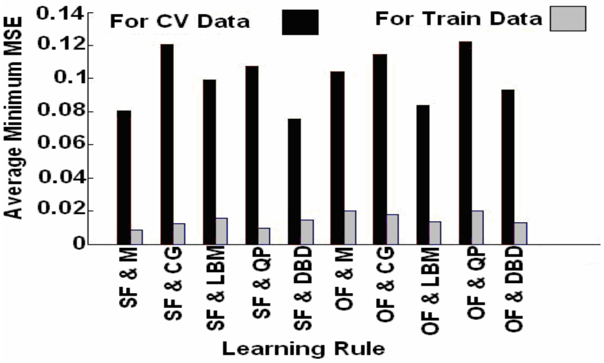

Figure 8: Graph showing variation of average minimum MSE on Training and CV dataset with Learning rule.

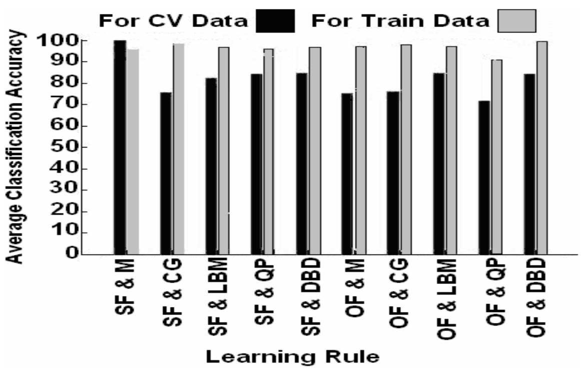

Figure 9: Graph showing variation of % average classification accuracy with learning rule.

Table 1. Confusion Matrix for training data set using PCA

Output/ Desired	**Angry**	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	**24**	0	0	0	0	0	0
Disgust	0	**22**	1	0	0	0	0
Fear	1	1	**25**	0	0	0	0
Happy	0	0	0	**29**	0	1	1
Neutral	0	0	0	0	**27**	0	1
Sad	0	0	0	0	0	**18**	1
Surprise	0	0	1	0	0	0	**25**

Table 2. Performance parameters for training data sheet using PCA

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.01199	0.02391	0.03102	0.01305	0.01842	0.02238	0.02372
% Correct	96	95.65	92.59	100	100	94.74	89.29

The overall accurate recognition of emotion is = 95.47%

Table 3. Performance parameters for training data sheet using PCA

Output/ Desired	**Angry**	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	**5**	0	0	0	0	0	0
Disgust	0	**7**	0	0	0	0	0
Fear	0	0	**3**	0	0	0	0
Happy	0	0	0	**1**	0	0	0
Neutral	0	0	0	0	**3**	0	0
Sad	0	0	0	0	0	**11**	0
Surprise	0	0	0	0	0	0	**2**

Table 4Confusion Matrix for test data set using PCA

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.00454	0.00863	0.01924	0.00553	0.01376	0.01731	0.00372
% Correct	100	100	100	100	100	100	100

The overall accurate recognition of emotion is = 100%

ii) Human Emotion Recognition Using FFT

The randomized data is fed to the PCA neural network and is trained for different hidden layers. It is observed that PCA with three hidden layer gives better performance. The number of Processing Elements (PEs) in the hidden layer is varied. The network is trained for 20 principal components (PC) and average minimum MSE on CV data is obtained when 71 PEs are used in all three hidden layers HL1, HL2 & HL3. Rigorous experimentation is done by varying number of exemplars for training and cross validation data using 71 PEs in all three hidden layers. The network is again trained three times. Average Minimum MSE on Trained and CV data set along with average classification accuracy is calculated and is shown in figure 10 & 11. It is observed that best results obtained when 10% exemplars are used for cross validation (CV) data and 90% for training.

With 10% CV and 90% train data various transfer functions are used for training the network and average minimum MSE on training and CV data is measured (Figure 12). Average classification accuracy for different transfer function is plotted in figure 13. It is observed that tanh is most suitable transfer function. Using tanh transfer function the PCA neural network is trained using learning rules namely Sangers full, Ojas full, Momentum, Conjugate-Gradient (CG), Quick Propagation (QP), Delta Bar Delta (DBD). Minimum MSE on training and CV data set is measured and is indicated in figure 14. Finally network is tested on training and testing dataset (figure 15). It is observed that Sangers full and Delta Bar Delta are most suitable learning rules for our neural network.

International Journal Of Computer Science And Applications Vol. 1, No. 2, August 2008
ISSN 0974-1003

From above experimentation selected parameters for PCA neural network are given below.

PCA NN (71-71-71-71-7), Number of epochs = 5100
Unsupervised learning control epochs = 100
Supervised learning control epochs = 5000
Number of principal components = 20
Exemplars for cross validation = 10%
Exemplars for training = 90%
Termination is at 500 epochs without improvement.

Layers of PCA	Transfer Function	Learning Rule	Step size	Momentum
Input Layer	Tanh	Sangers full	-	-
HL#1	Tanh	Delta Bar Delta	0.5	0.7
HL#2	Tanh	Delta Bar Delta	0.09	0.7
HL#3	Tanh	Delta Bar Delta	0.05	0.7
Output layer	Tanh	Delta Bar Delta	0.005	0.7

Time elapsed per epochs per exemplar = 0.2191 m.Sec. Number of free parameters (P) of PCA = 15840
Number of exemplars in training dataset (N) = 189
(N/P) ratio = 0.0119
Finally designed PCA is tested on training and CV dataset and results are shown in table 5 to 8.

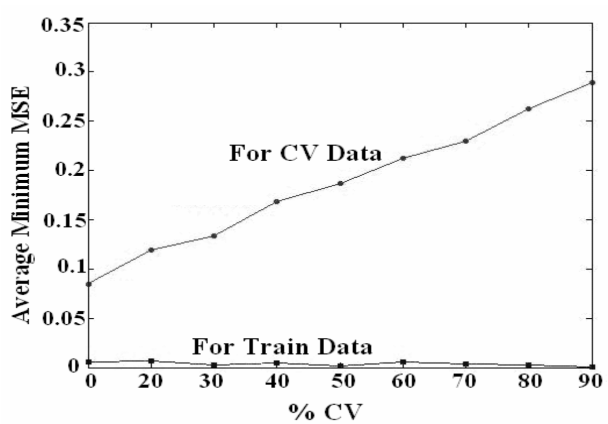

Figure 10: Graph showing variation of average minimum MSE on Training and CV dataset with % of CV data.

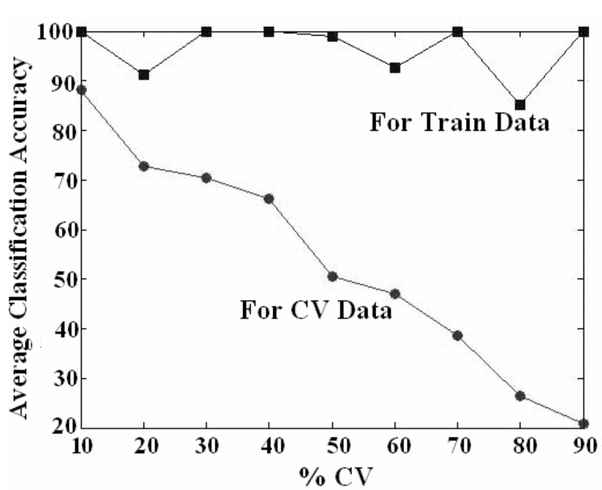

Figure 11: Graph showing variation of % average classification accuracy with % of CV data.

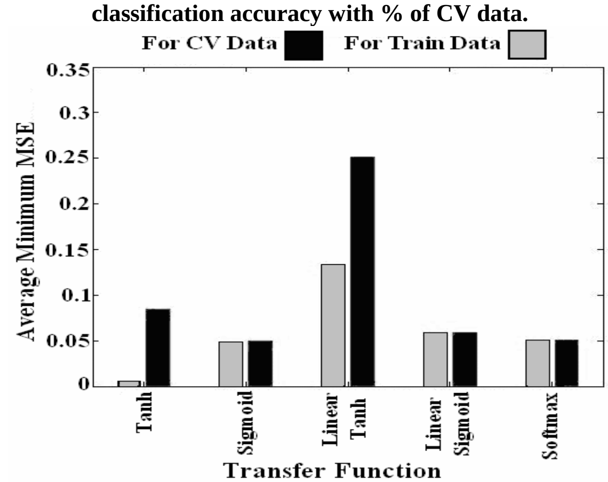

Figure 12: Graph showing variation of average minimum MSE on Training and CV dataset with Transfer function.

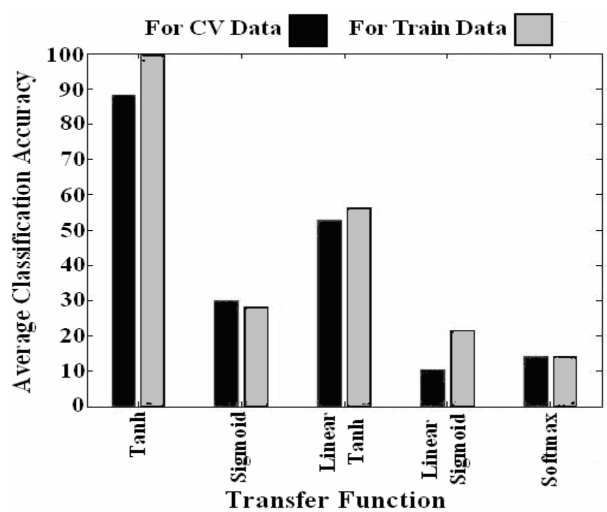

Figure 13: Graph showing variation of % average classification accuracy with Transfer function.

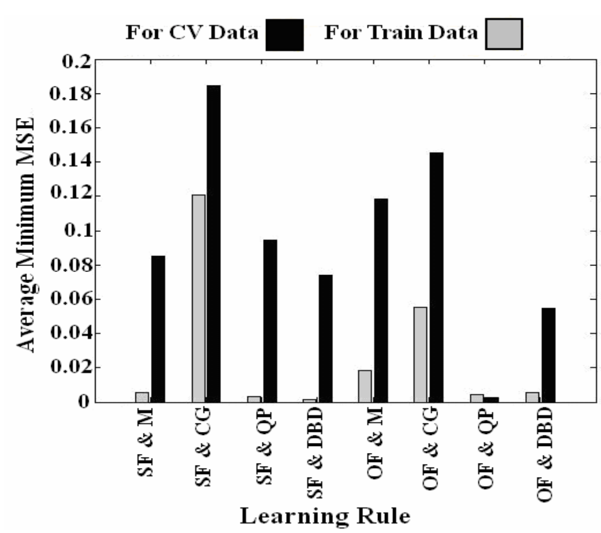

Figure 14: Graph showing variation of average minimum MSE on Training and CV dataset with Learning rule.

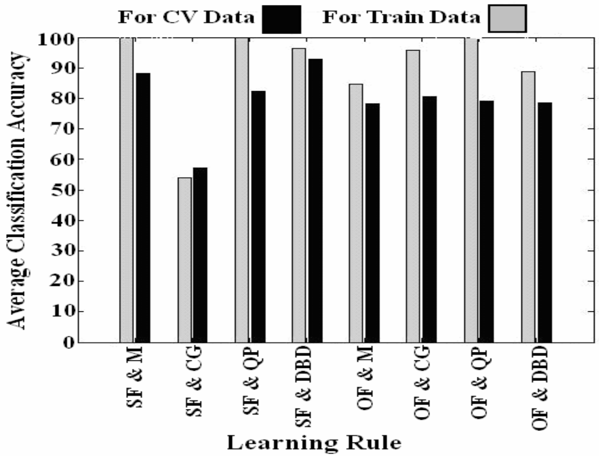

Figure 15: Graph showing variation of % average classification accuracy with learning rule.

Table 5. Confusion Matrix for training data set using PCA

Output/ Desired	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	**29**	0	0	0	0	0	0
Disgust	0	**27**	0	0	0	0	0
Fear	0	0	**28**	0	0	0	0
Happy	0	0	0	**25**	0	0	0
Neutral	0	0	0	0	**25**	0	0
Sad	0	0	0	0	0	**27**	0
Surprise	0	0	0	0	0	0	**28**

Table 6. Performance parameters for training data sheet using PCA

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.00111	0.00082	0.00062	0.00089	0.00080	0.00141	0.00079
% Correct	100	100	100	100	100	100	100

The overall accurate recognition of emotion is = 100%

Table 7 Performance parameters for training data sheet using PCA

Output/ Desired	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	**1**	0	0	0	0	0	0
Disgust	0	**3**	1	0	0	0	0
Fear	0	0	**1**	0	0	0	0
Happy	0	0	0	**5**	0	0	0
Neutral	0	0	0	0	**5**	1	0
Sad	0	0	0	0	0	**2**	0
Surprise	0	0	0	0	0	0	**2**

Table 8. Confusion Matrix for test data set using PCA

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.01481	0.05434	0.05718	0.05776	0.04910	0.06015	0.07484
% Correct	100	100	50	100	100	66.67	100

The overall accurate recognition of emotion is = 88.10%

5. RESULT

In this paper, the authors evaluated the performance of the two Feature Extraction Techniques namely DCT & FFT and compare there performance for the design of PCA to recognize human emotions.

Table 1 to 4 show emotion recognition results on training and testing data set for optimally designed PCA when DCT is used. The accuracy of recognition is 95.47% on train dataset and 100% on test dataset for all the emotions namely angry, disgust, fear, happy, neutral, sad and surprise.

Table 5 to 8 show emotion recognition results on training and testing data set for optimally designed MLP when FFT is used. The accuracy of recognition is 100% on train dataset and 88.10% on test dataset for all the emotions namely angry, disgust, fear, happy, neutral, sad and surprise.

6. CONCLUSION

The performance comparison for the various extraction methods is given below.

Neural Network Model	Average Minimum MSE		% Average Classification Accuracy		Time elapsed per epoch per exemplar	N/P
	Train	CV	Train	CV		
DCT	0.0087	0.0803	95.47	100	0.0129mSec.	0.118
FFT	0.0057	0.0848	100	88.10	0.2191mSec.	0.0119

When we compare the average minimum MSE on train data for PCA network, when DCT & FFT are used for human emotion recognition, it is observed that the value is lower for FFT and obtained more classification accuracy than DCT. For CV data (unknown exemplars) average minimum MSE is lower for DCT and authors obtained higher human emotion recognition classification accuracy.

When we compare the time elapsed per epoch per exemplar for the PCA neural network, it is lower when DCT is used for emotion recognition. It indicates that PCA neural network is quite faster in training when DCT is used. Ratio of number of exemplars in training the dataset (N) to the number of free parameters (P) of PCA, network is quite low for FFT. This shows that the design of PCA neural network when FFT is used for human emotion recognition, is quite simple and easy for synthesis. From the comparative study authors recommend DCT for human emotion recognition for PCA neural network.

7. REFERENCES

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

[2] R.W. Picard, Affective Computing Cambridge. MA: MIT Press, 1997.

[3] R. W. Picard, Toward agents that recognize emotion in Proc. IMAGINA, 1998, pp. 153–155.

[4] D. Beymer, A. Shashua, and T. Poggio, Example Based Image Analysis and Synthesis, M.I.T. A.I. Memo No. 1431, 1993.

[5] I.A. Essa and A. Pentland, A Vision System for Observing and Extracting Facial Action Parameters, Proc. IEEE CVPR, pp.76-83, 1994

[6] H. Li, P. Roivainen, and R. Forcheimer, 3D Motion Estimation in Model-Based Facial Image Coding, IEEE Trans. Pattern Analysis and Machine intelligence, vol. 15, pp. 545-555,1993.

[7] K. Mase, Recognition of Facial Expression from Optical Flow, IEICE Trans., vol. E 74, pp. 3,474-3483, 1991.

[8] K. Matsuno, C. Lee, and S.Tsuji, Recognition of Human Facial Expressions without Feature Extraction, Proc. ECCV, pp. 513-520, 1994.

[9] M. Rosenblum, Y. Yacoob, and L.S. Davis, Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture, IEEE Workshop Motion of Non-Rigid and Articulated Objects, Austin, Texas, pp. 43-49, Nov. 1994.

[10] D. Terzopoulos and K. Waters, Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, pp. 569-579, 1993.

[11] Y.Yacob and L Devis. Recognizing Human facial expression from long image sequences using optical flow. IEEE transaction on Pattern Analysis and Machine Intelligence [PAMI], 18{6}:636-642, 1996.

[12] J. N. Bassili, Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face, J. Personality and Social Psych., vol. 37, pp. 2049-2059, 1979.

[13] Irfan A. Essa and Alex P. Pentland, Coding, Analysis, Interpretation and Recognition of Facial Expressions. IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 19, No. 7, July 1997.

[14] Analysis, Interpretation and Synthesis of Facial Expressions by Irfan Aziz Essa, Ph.D. thesis, MIT February 1995.

[15] Gianluca Donato, Marian Stewart Bartlett, Joseph C. Hager, Paul Ekman, and Terrence J. Sejnowski, Classifying Facial Actions, IEEE Transaction Pattern Analysis and Machine Intelligence, vol. 21,No. 10 October 1999.

[16] I. Essa and A. Pentland, Coding, Analysis, Interpretation and Recognition of Facial Expressions, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757-763, July 1997.

[17] Mark Rosenblum, Yaser Yacoob, and Larry S. Davis, Human Expression Recognition from Motion Using; a Radial Basis Function Network Architecture. IEEE Transactions On Neural Networks. Vol. 7, No. 5, September 1996

[18] P. Ekman and W.Friesen, The Facial Action Coding System. San Francisco CA: Consulting Psychologists Press, 1978.

[19] Y. Yacoob and L. S. Davis, Computing spatio-temporal representations of human faces, in IEEE trans. on Computer Vision and Pattern Recognition, 1994, pp. 70-75.

[20] K. Mase and A. Pentland, Recognition of facial expression from optical flow, IEICE Trans. E, vol. 74, pp. 408–410, 1991.

[21] C. Padgett and G. W Cottrell, A simple neural network models categorical perception of facial expressions, in Proc. 20th Annu. Cognitive Science Conf., 1998, pp. 806–807.

[22] B. Fasel and J. Luttin, Recognition of Asymmetric Facial Action Unit Activities and Intensities, IDIAP Research Rep., 1999.

[23] A. Lanitis, C. J. Taylor, and T. F. Cootes, Automatic interpretation and coding of face images using flexible models, IEEE Trans. Pattern Analysis and Machine Intelligence., vol. 19, no. 7, pp. 743–756, Jul. 1997.

[24] M.Matsugu, K.Mori, Y.Mitari, and Y.Kaneda, Subject independent facial expression recognition with robust face detection using a convolutional neural network, Neural Network., vol. 16, pp. 555–559, 2003.

[25] M.Gargesha and P.Kuchi, Facial expression recognition using artificial neural networks, Artificial Neural Computing System, pp. 1–6, 2002.

[26] A.Johnston, P. W. McOwan, and C. P. Benton, Robust velocity computation from a biologically motivated model of motion perception, in Proc. Roy. Soc. Lon., vol. 266, 1999, pp. 509–518.

[27] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, Performance of optical flow techniques, Int. Journal of Computer Vision, vol. 12, pp. 43–77, 1994.

[28] Mohammed Yeasin, Baptiste Bullot, and Rajeev Sharma, Recognition of Facial Expressions and Measurement of Levels of Interest From Video IEEE Transactions On Multimedia, Vol. 8, No. 3, June 2006.

[29] M. Black and Y. Yacoob, Recognizing facial expressions in image sequences using local parameterized models of image motion, presented at the Int. Conf. Computer Vision, 1995