Volume 3, Issue 5, May 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Human Facial Expression Recognition based on Neural Network

Divanshu
Student of M.Tech (CSE)
Gurukul Vidyapeeth Institute of Engg & Tech
Patiala, Punjab, India.

Abstract: ***Facial expressions are a valuable source of information that accompanies facial biometrics. Early detection of physiological and psycho-emotional data from facial expressions is linked to the situational awareness module of any advanced biometric system for personal state re/identification. A new hidden Markov model (HMM) based feature generation scheme is proposed for face recognition (FR) in this paper. In this scheme, HMM method is used to model classes of face images. A set of Fisher scores is calculated through partial derivative analysis of the parameters estimated in each HMM. These Fisher scores are further combined with some traditional features such as log-likelihood, 2-D DCT and appearance based features to form feature vectors that exploit the strengths of both local and holistic features of human face. Neural Network is then applied to analyze these feature vectors for face recognition (FR). Performance improvements are observed over stand-alone HMM method and Fisher face method which uses appearance based feature vectors. A further study reveals that, by reducing the number of models involved in the training and testing stages of Neural network, the proposed feature generation scheme can maintain very high discriminative power at much lower computational complexity comparing to the traditional HMM based face recognition (FR) system. Experimental results on a public available face database are provided to demonstrate the viability of this scheme. The performance of this system has been validated on three public databases: the JAFFE, the Cohn-Kanade, and the MMI image.***

Keywords: *Face recognition, Hidden Markov model, Fisher scores, Facial Expression, neural network.*

1. INTRODUCTION

We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace the content with your own material. One of the most popular appearance based methods [1–3] for face recognition (FR) developed in recent years is the Fisher face method. The Fisher face method performs LDA of feature vectors obtained as one-dimensional representation of a face image and retrieves the identity of person based on the nearest-neighbor classification criterion in the LDA space. This method is insensitive to large variation in lighting direction and facial expression [2]. Meanwhile, statistical model based methods such as hidden Markov model (HMM) have also been proposed for FR problems [4–8]. This method uses HMM to describe the statistical distribution of observation vector sequences which are generated from small sub-image blocks of face image. Classification is usually based on Bayesian decision rule, e.g., maximum a posteriori (MAP) criterion. Comparing with appearance based methods; HMM methods focus mainly on local characteristics of human faces. These methods have the flexibility to incorporate information from different instances of faces at different scales and orientations [5]. However, in these existing statistical model based methods, only the calculated likelihood of a particular observation on each established model is used as the measure of closeness of the observation towards the corresponding class. In this work, we present a new feature vector generation scheme from HMMs. The scheme generates feature vectors which represent the influence of the model parameters of several competing HMMs on the generation of a particular observation vector sequence. Similar methods were proposed and used in biosequence analysis, speech recognition, and speaker identification [9, 14, and 15]. Unlike previous schemes which are inherently two-class problem oriented, the proposed scheme in this work is multi-class problem oriented and the resulting feature vectors appear to be more effective. We also explore the strengths of both Fisher face method, 2-D DCT and HMM method by combining appearance based features (as seen in Fisher face approaches) and statistical model based features together to form new feature vectors, which may have greater discriminative power over those used separately. Furthermore, in a typical multi-class HMM method, one HMM is established for each class of object (e.g. faces of one person), and a test observation is compared to all the available classes in order to determine its identity. In this work we attempt to reduce the number of HMMs involved in this process and manage to achieve a comparable recognition performance as when all HMMs are used. Apparently the model reduction translates to a significant computational advantage, which effectively improves the scalability of such statistical model based methods.

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, fisher scores 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.

2. 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.

3. PROPOSED EXPRESSION RECOGNITION TECHNIQUE

The new technique uses a 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.

3.1 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.

3.2 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×N. The 2-D DCT coefficients of a square block with size L1×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.

3.3 Fisher scores for HMM

The computation of the Fisher scores depends on the structure of the statistical model. The statistical model we choose for FR is a one-dimensional ergodic HMM which assumes the observation distribution density as Gaussian with diagonal covariance matrix. For a Gaussian HMM, the parameters needed to represent the model include three components, i.e., the state transition distribution *A*, the observation probability distribution *B* and the initial states distribution [16]. In order to completely represent the gradients, all three components should be considered.

3.4 Neural Network Training

The dimension of the input vector of the neural network is M ×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.

4. EXPERIMENT 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×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). 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 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.

4.1 Neural Network Training and Testing Results

The proposed network was trained with feature vector data cases. 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 28.60 minutes. The testing process is done for 400 cases. These 400 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 3.

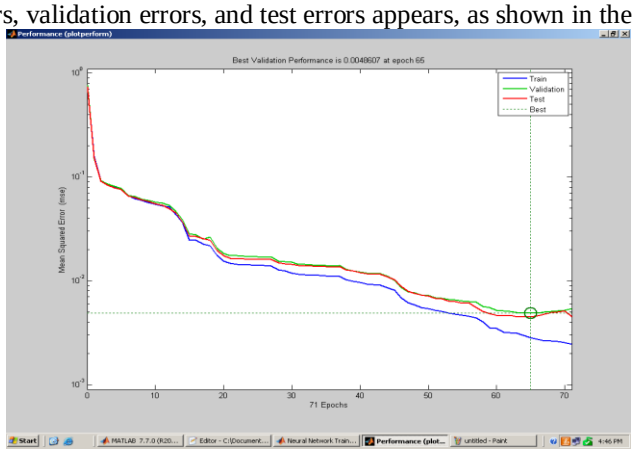

Figure 3: Performance plot

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 4.

Figure 4: ROC Plot

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 5.

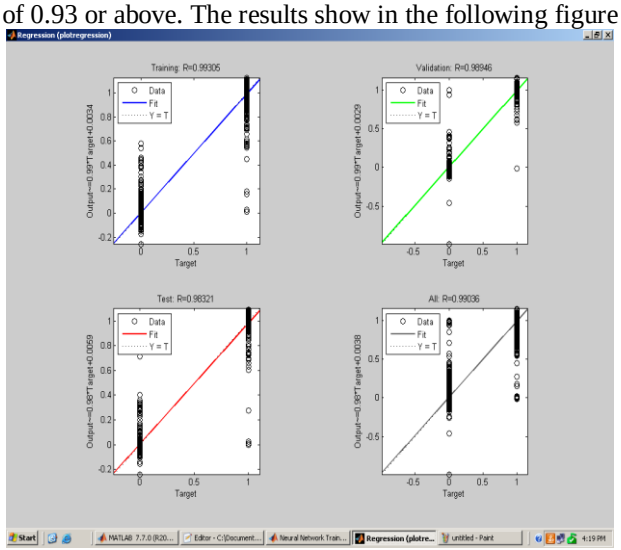

Figure 5: Regression Plots

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 6.

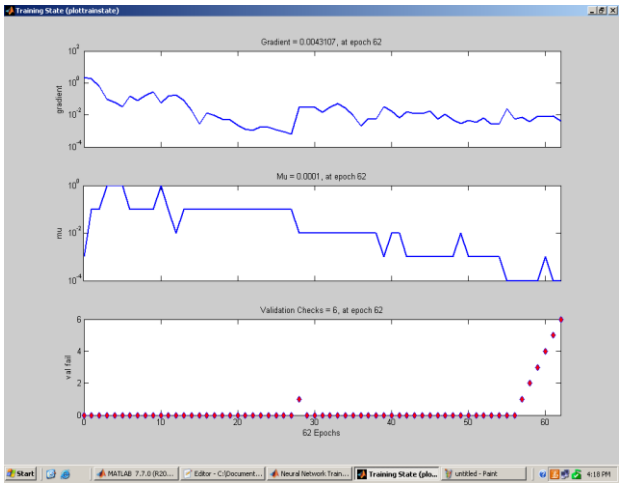

Figure 6: Training State Plot

Confusion Matrix: This figure shows the confusion matrices for training, testing, and validation, and the three kinds of data combined. The network outputs are very accurate, as you can see by the high numbers of correct responses in the green squares and the low numbers of incorrect responses in the red squares. The lower right blue squares illustrate the overall accuracies. The diagonal cells show the number of cases that were correctly classified, and the off-diagonal cells show the misclassified cases. The blue cell in the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red). The results show very good recognition.

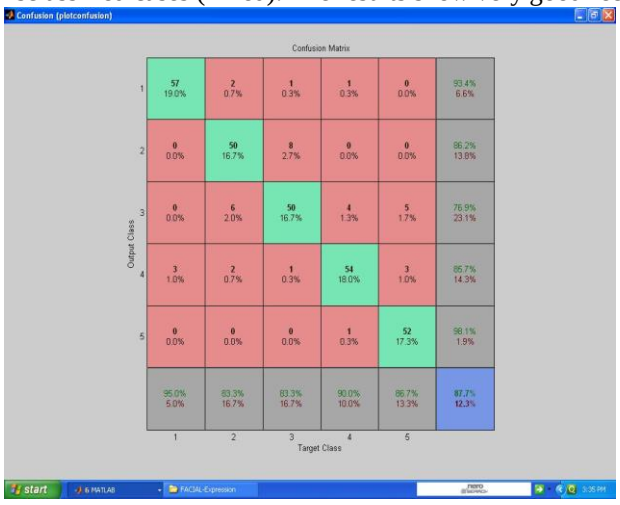

Figure 7: Confusion Matrix

5. Conclusions

In this paper, a new facial expression recognition technique is proposed which uses 2-D DCT, HMM, Fisher Scores and neural network to separate the facial expressions systematically. The 2-D DCT and fisher scores are 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 50 men and 50 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).
[7] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1991) 71–86.
[8] P. Bellhumeur, J. Hespanha, D.J. Kriegmand, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 711–720.
[9] K. Etemad, R. Chellappa, Face recognition using discriminant eigenvectors, in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1996, pp. 2148–2151.
[10] F. Samaria, Face recognition using hidden Markov model, Ph.D. Thesis, University of Cambridge, 1995.
[11] A. Nefian, A hidden Markov model-based approach for face detection and recognition, Ph.D. Thesis, Georgia Institute of Technology, 1999.
[12] A. Nefian, Embedded Bayesian networks for face recognition, ICME 2002—IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, August 2002, pp. 133–136.
[13] S. Eickeler, S. Müller, G. Rigoll, High performance face recognition using pseudo-2-D hidden Markov models, European Control Conference (ECC), August 1999.
[14] H. Othman, T. Aboulnasr, Low-complexity 2-D hidden Markov model face recognition, ISCA 2000—IEEE International Symposium on Circuits and Systems, Geneva, Switzerland, May, 2000, pp. V33–V36.
[15] T. Jaakkola, D. Haussler, Exploiting generative models in discriminative Classifiers, in: S.A. Solla, T.K. Leen, K.R. Müller (Eds.), Advances in Neural Information Processing Systems, vol. 12, MIT Press, Cambridge, MA, 2000.
[16] T. Jaakkola, D. Haussler, Maximum entropy discrimination, Technical Report AITR-1668 MIT, 1999.
[17] M. Seeger, Covariance kernels from Bayesian generative models, in: T.G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems, vol. 14, MIT Press, Cambridge, MA, 2002.
[18] K. Tsuda, S. Akaho, M. Kawanabe, K.R. Müller, Asymptotic properties of the Fisher kernel, Neural Comput. 16 (1) (2004) 115–137.
[19] S. Amari, Differential-Geometrical Methods in Statistics, Lecture Notes in Statistics, vol. 28, Springer, NewYork, 1985.
[20] N. Smith, M. Gales, Using SVMs to classify variable length speech patterns, Technical Report CUED/F-INFENG/TR.412, Cambridge University Engineering Department, June 2001.
[21] S. Fine, J. Navrátil, R. Gopinath, A Hybrid GMM/SVM approach to speaker identification, ICASSP 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Utah, USA, May, 2001, pp. 417–420.
[22] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989).
[23] K. Fukunaga, Introduction to Statistical Pattern Recognition,second ed., Academic Press, Boston, 1990.
[24] W. Zhao, R. Chellappa, A. Krishnaswamy, Discriminant analysis of principal components for face recognition, AFGR 1998—IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, April, 1998, pp. 336–341.
[25] A. Martínez, A. Kak, PCA versus LDA, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 228–233.

