Enhanced Independent Component Analysis of Temporal Human Expressions Using Hidden Markov model

J. J. Lee, MD. Zia Uddin, T.-S. Kim*

Abstract Facial expression recognition is an intensive research area for designing Human Computer Interfaces. In this work, we present a new facial expression recognition system utilizing Enhanced Independent Component Analysis (EICA) for feature extraction and discrete Hidden Markov Model (HMM) for recognition. Our proposed approach for the first time deals with sequential images of emotion-specific facial data analyzed with EICA and recognized with HMM. Performance of our proposed system has been compared to the conventional approaches where Principal and Independent Component Analysis are utilized for feature extraction. Our preliminary results show that our proposed algorithm produces improved recognition rates in comparison to previous works.

Keywords: Facial Expression Recognition, Hidden Markov Model(HMM) , Independent Component Analysis(ICA), Principal Component Analysis(PCA), Linear Discriminant Analysis(LDA).

This work was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology and Communication Research Center) support program supervised by the IITA (IITA-2006-(C1090-0602-0002)).

* Lee, Jee Jun : MS, Department of Biomedical Engineering, Kyung Hee University e-mail: ljj@khu.ac.kr
** MD. Zia Uddin : MS, Department of Biomedical Engineering, Kyung Hee University e-mail: ziauddin@khu.ac.kr
*** Kim, Tae Sung : Professor, Department of Biomedical Engineering, Kyung Hee University e-mail: tskim@khu.ac.kr

1. Introduction

In recent years, facial expression recognition (FER) has been regarded as one of fundamental technologies for human computer interaction (HCI) which enables computers to interrelate with humans in a way to human to human interaction. Since FER technology provides computers a way of sensing user's emotional information, facial expression analysis can contribute to make a HCI system respond to expressive states of humans and human behaviors.

Facial expressions to be recognized can be represented in images in two ways: a static facial expression image and a sequence of image frames of facial expressions. Each type of facial expression image(s) is then analyzed in two ways. The first is via facial action units (FAUs) and their association with facial expressions where facial expressions are represented by segmented key units called AUs such as eyes, lips, etc. The second is emotion specified expressions represented in an association of basis feature images which are those representing anger, joy, sadness, surprise, fear, and disgust.

The Facial Action Coding System (FACS) which was proposed by Ekman and Friesen defines 46 AUs to correspond to individually distinct muscle changes of a face and the combination of those to observe the expression [1]. For FAU detection, feature based approaches were introduced in [4]. Principal Component and Independent Component Analysis (PCA/ICA), Local Feature Analysis (LFA), Gabor Wablet representation were explored to separate FAUs and they demonstrated ICA and Gabor representation performed best to classify 12 facial actions. A neural network was also employed with multistate facial component models to capture facial actions [5]. After successful applications of Hidden Markov Model (HMM) to a speech recognition, a system based on HMM for FER was introduced and this approach was implemented with some Naive-Bayes classifiers and with facial animation parameters to recognize facial expressions [6][7][8]. Moreover, support vector machine (SVM) was used as a facial expression recognizer with FAUs extracted by geometric deformation features [9]. However, the drawback of FER based on FAU is a huge potential for expressions to be determined differently from the diverse combination of AUs, which causes an enormously complicated FER problem.

As for the emotion specific facial expression image processing methods, Andrew et al. used Principal component Analysis (PCA) using pixel intensities of images and Linear Discriminant Analysis (LDA) for revealing principal findings [2]. In [10], the two-dimensional cosine transform was utilized for extracting feature of the entire face images, and the neural network system was employed to recognize the facial expressions. Bartlett et al. introduced two ICA architectures where the first architecture found spatially local basis images for the faces and the second is for producing a factorial face code [11]. In their proposed settings, they found that the first ICA method is best for facial expression recognition meanwhile the second ICA approach is for face recognition. Kwak and Pedrycz introduced an Enhanced ICA which is the ICA method augmented with Fisher Linear Discriminant analysis (FICA) and this method outperforms the ICA method for the face task [12]. Nevertheless, the emotion-specified expression approach reveals an inadequacy that only the single image was in use to perform FER which ignores the temporal information. Since the temporal variations of the face are assembled into a specific expression, the temporal information should be deliberated for FER as well as for a real-time system.

In this paper, we focus on facial expression recognition from sequential images of facial expressions as we consider the importance of the temporal information. Also, we have used the entire face image for image processing in order to minimize the complexity due to the possibilities of different understanding of FAUs. The proposed architecture adopted the feature extractor appeared in [12] due to its improved capability of separating classes and we implemented HMM for the recognizer due to the fact that HMM can handle the relation between expression states and observed states of local facial movements of sequential images. In this study, we have recognized four emotional expressions: namely anger, joy, sadness, and surprise. Our preliminary results show that our method outperforms the conventional approaches of utilizing PCA and ICA in combination with HMM and previously reported results.

2. Methodology

2.1 Facial Expression Database

The facial expression database used in our experiment is the Cohn-Kanade AU-coded facial expression database consisting of facial expression sequences with a neutral expression as an origin to a target facial expression [13]. Facial expressions include 97 subjects with subsets of some expressions. In this work, we have dealt with four major expressions: namely anger, joy, sadness, and surprise. For image processing, 163 subsets of 97 subjects which contain 8 sequences per expression are selected. Fig. 1 shows a set of surprise facial expression sequences as an example.

2.2 Our Proposed Recognition System

In preprocessing, face alignment and histogram equalization have been performed. The face images are centered with the aligned eye and mouth locations and then cropped and scaled to 80 by 80 pixels [15]. In order to minimize lightening effect due to different light conditions, histogram equalization is performed.

In our implemented recognition system, we have used FICA feature extractor and HMM as a recognizer. The overall system is shown in Fig. 2.

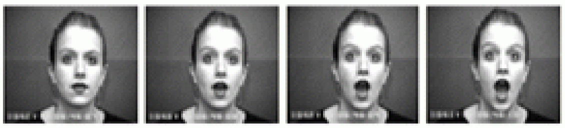

Fig. 1. A sequence of facial expression Images

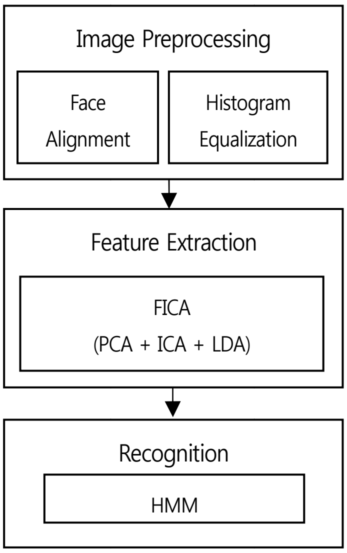

Fig. 2. The overall FER system

2.2.1 Feature Extraction using FICA

FICA is the combination of the ICA method based on learning high-order moments of the principal components and LDA which is a class specific learning algorithm. The purpose of this method is to find an optimal local presentation of face images. Extracting features consists of two major steps which are ICA based on eigenvectors and LDA. For the first step, each training image is assigned into a row of the data matrix X for ICA which finds a weight matrix W of independent source U. That is to say, the source images estimated by the rows of U are used as basis images to represent the dataset,

$$U = WX. \quad (1)$$

In this work, we apply ICA to the first *m* eigenvectors resulted from PCA on the dataset rather than to the original images due to the following reasons. ICA algorithm produces the same number of independent components (ICs) corresponding to the dimensions of the dataset, which means finding ICs from the 800 sequence images used for training takes disadvantage of long computing time and it encounters the memory limitation of capturing the number of ICs. Meanwhile, PCA transforms the high dimensional space to a reduced space capturing the maximum variability of the dataset so that principal components (PCs) still contains the high-order relationships which are not separated. PCs are shown in Fig. 3.

Fig. 3. Facial Expression representation onto the reduced feature space using PCA

Therefore, the ICA formula can be expressed as

$$U = W_{ICA} P_m^T \quad (2)$$

where U is the obtained basis images comprised with the coefficient W_{ICA} and the first *m* eigenvectors P_m^T applied onto X with zero mean, whose feature vector is $V = XP_m$. The reconstructed image set is then described as

$$\bar{\mathrm{X}} = \mathrm{VP_m^T} = \mathrm{VW_{ICA}^{-1}U}. \quad (3)$$

Therefore, IC representation U is computed by the rows of the form

$$R = VW_{ICA}^{-1}. \quad (4)$$

and IC basis images are shown in fig. 4.

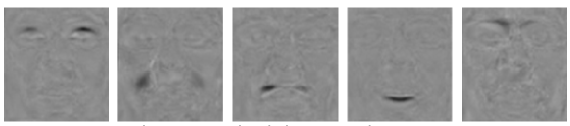

Fig. 4. Some basis images using ICA

For the second step, LDA, which is a class specific method using the class information maximizing the ratio of between-class scatter matrix and the within-class scatter matrix, is performed on IC feature vectors R. The within- and between scatter matrix are defined in the form as follows:

$$S_W = \sum_{i=1}^{c} \sum_{r_k \in C_i} (r_k - \bar{r}_i)(r_k - \bar{r}_i)^T \quad (5)$$

$$S_W = \sum_{i=1}^{c} N_i (\bar{r}_i - r_m)(\bar{r}_i - r_m)^T \quad (6)$$

where r_k is the feature vector from R, $\bar{r}_i$ the mean of class C_i, and r_m the mean of all feature vectors R. The optimal projection W_d, which maximizes the ratio $J(W_d) = |W_d^T S_B W_d| / |W_d^T S_W W_d|$, is derived by solving the generalized eigenvalue problem: $S_B W_d = \Lambda S_W W_d$, where Λ is the diagonal eigenvalue matrix. This discriminant vector W_d forms the basis of the $(c-1)$dimensional subspace for a c-class problem. Finally, the final feature vector G and the feature vector for testing image can be obtained by the criterion

$$G = RW_d^T \quad (7)$$

$$G_{test} = R_{test} W_d^T = X_{test} P_m W_{ICA}^{-1} W_d^T \quad (8)$$

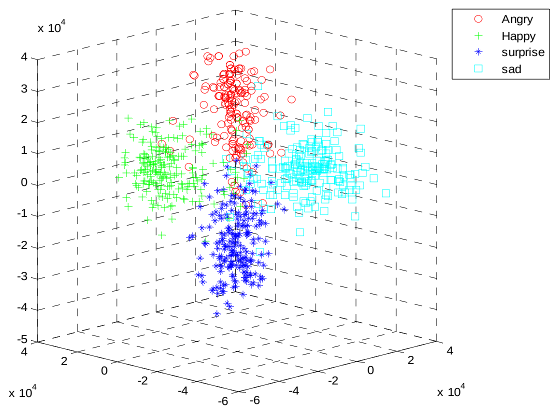

Fig. 5. Final vector plot of four expressions using FICA

As shown in Fig. 5, features related with a specific expression are concentrated in a particular region of the feature space. In our experiment, the features of the starting sequence images, neutral faces, are centered in the feature space and the feature of the target expressions are located in the center of the each expression region which means each expression feature region contains the temporal variation of face features.

2.2.2 Recognition using HMM

HMM used in our experiment is a left-right model useful to model a sequential event in a system which is generated from a fixed state and ends in a target state [6]. Generally, the purpose of HMM is to determine the model type λ with the highest probability of the likelihood $Pr(O|\lambda)$ when observing the sequential data O. The HMM model is denoted as $\lambda = \{A, B, \pi\}$ with the following elements:

$S = \{S_1, S_2, \ldots, S_N\}$: States in the model

$Q = \{q_1, q_2, \ldots, q_N\}$: States at time t

$A = \{a_{ij}\}, a_{ij} = Pr(q_{t+1} = S_j | q_t = S_j),\ 1 \le i,\ j \le N$
: State Transition Probability between hidden states

$B = \{b_i(O_t)\},\ b_j = Pr(O_t | q_t = S_j), 1 \le j \le N$
: Observation Symbol Probability in state j

$\pi = \{\pi_j\},\ \pi_j = \Pr(q_1 = S_j)$: Initial State Probability

In the learning step, we set the variable, $\delta_t(i,j)$, the probability of being in state q_i at time t and state q_i at time $t+1$, to re-estimate the model parameters as follows:

$$\delta_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{Pr(O|\lambda)} \quad (9)$$

$$\gamma_t(i) = \sum_{j=1}^{n} \delta_t(i,j) \quad (10)$$

where $\alpha_t(i)$ is the forward variables and $\beta_t(i)$ is the backward variables and the variable $\gamma_t(i)$ is the probability being in state q_i at time t [15].

Using the variables above, we can estimate the updated parameters A, B and π of the model of λ in the form

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \delta_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (11)$$

$$\bar{b}_j(k) = \frac{\sum_{t=1, O_t=k}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (12)$$

After training the model, we evaluate the observation sequences $O = \{O_1, O_2 \ldots, O_T\}$ from video dataset and determine the proper model whose $Pr\,(O|\lambda)$ is maximum. The likelihood of the observation O given the trained model λ can be determined via the forward variable in the form

$$Pr\,(O|\lambda) = \sum_{i=1}^{N} \alpha_t(i) \quad (13)$$

In our case, we used HMM having 4 states and 32 codebooks. The Linde, Buzo and Gray (LBG) algorithm was employed for vector quantization [14]. The structure and the transition probabilities of the HMM are illustrated in Figs. 6 and 7.

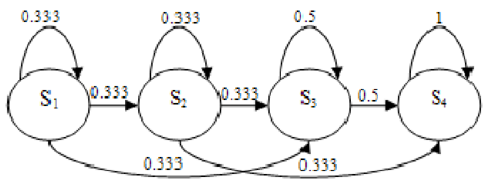

Fig. 6. HMM structure and transition probabilities for anger before training

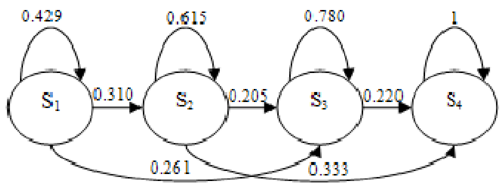

Fig. 7. HMM structure and transition probabilities for anger after training

3. Experimental Results

There are four expressions to be classified and recognized via the proposed approach. The expression sequences were collected from the Cohn-Kanade database introduced in Section 2.1. A total of 163 sequences are used and 8 images which only display the frontal view of the facial expressions are contained in each sequence. A total of 22 sequences of anger, 26 of joy, 26 of sadness, and 26 of surprise sequences are used in training and for the testing purpose, 11 of anger, 19 of joy, 13 of sadness, 20 of surprise subsets are used.

The first experiment is to determine the optimal number of features and the size of the codebook. To do this, we differentiate the number of feature selected in the PCA step to test the performance. As described in Fig. 8, 120 eigenvector shows the best performance. After selecting the number of eigenvectors we test the performance with the different size (2^n, n=4, 5, 6) of the codebook and 32 for the size of the codebook is found to be proper to be used for vector quantization as illustrated in Table. 1.

For comparison of three types of conventional methods and the proposed method, PCA, ICA, PCA_LDA, and, FICA, we fixed 120 as the number of selected features and 32 as the codebook size for parameter settings. The total mean of recognition rate from ICA representation of facial expression images is 69.84% which is higher than that of the PCA recognition rate. Moreover, the best recognition rate from our experiment is resulted from our proposed method presents 94.49% compared to the performance of 82.15% from the best conventional approach PCA_LDA. Table 2, 3, 4, and 5 compare the recognition results of each method.

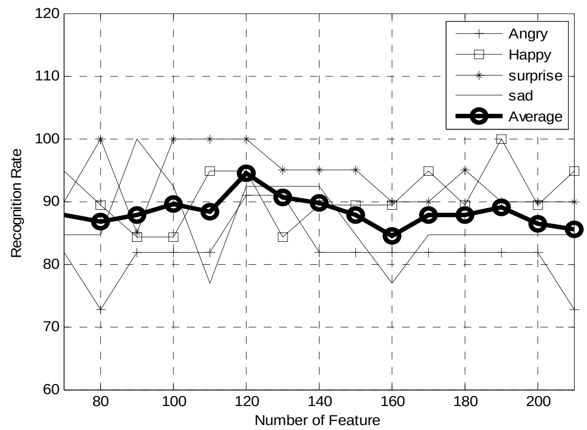

Fig. 8. Recognition rate of facial expressions using FICA with number of features

Table 1. Person independent confusion matrix using PCA

Emotion	Anger	Joy	Sadness	Surprise
Anger	**45.45(%)**	27.27	9.09	18.18
Joy	27.78	**66.67**	5.56	0
Sadness	8.33	8.33	**58.33**	25
Surprise	5.56	0	5.56	**88.89**

Table 2. Person independent confusion matrix using ICA

Emotion	Anger	Joy	Sadness	Surprise
Anger	**45.45(%)**	18.18	9.09	27.27
Joy	5.56	**83.33**	5.56	5.56
Sadness	8.33	8.33	**66.67**	16.67
Surprise	0	0	11.11	**88.89**

Table 3. Person independent confusion matrix using PCA_LDA

Emotion	Anger	Joy	Sadness	Surprise
Anger	**72.73(%)**	9.09	0	18.18
Joy	10.53	**84.21**	0	5.26
Sadness	0	15.38	**76.92**	7.69
Surprise	0	5	0	**95**

Table 4. Person independent confusion matrix using FICA

Emotion	Anger	Joy	Sadness	Surprise
Anger	**90.91(%)**	0	0	9.09
Joy	5.26	**94.74**	0	0
Sadness	7.69	0	**92.31**	0
Surprise	0	0	0	**100**

4. Conclusion

In this work, we have presented a new system for FER utilizing FICA for facial expression feature extraction, and HMM for recognition. We have illustrated the performance of our proposed method applied on sequential datasets for the four facial expression recognition problems. The result shows that FICA, which is the linear discriminant approach on ICA feature vectors from optimal representation of PCs, can improve the feature extraction task. Furthermore, HMM dealing with FICA processed sequential data can provide the higher recognition rate, 94.49%, in comparison to the results via the conventional feature extracting methods with HMM. We conclude that the whole face features can be used for the HMM method rather than usage of action units for HMM to recognize the partial face activities.

Reference

[1] Ekman, P., Priesen, W. V.," Facial Action Coding System: A technique for the Measurement of Facial Movement", Consulting Psychologists Press, Palo Alto, CA (1978)

[2] Andrew J. Calder, Mike Burton, A., Miller P., Andrew W. Y., Akamatsu S., "A principal component Analysis of facial Expressions", Vision Research, pp.1179-1208, 2001

[3] Ekenel, H. K., Sankur, B.: Feature selection in the independent component subspace for face recognition. Pattern Recognition Letters, pp. 1377-1388 (2004)

[4] Danato, G., Bartlett, M. S, Hagar, J. C., Ekman, P. and Sejnowski, T. J.: Classifying Facial Actions. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989 (1999)

[5] Tian, Y. L., Kanade, T., Cohn, J.: Evaluation of Gbor wavlet-based facial action unit recognition in image sequences of increasing complexity. 5th IEEE conf. Automatic Face and Gesture Recognition, pp.229-234 (2002)

[6] Rabiner, L. R.: A Tutorial on Hidden Markov Modes and selected application in speech recognition. Proc. IEEE, pp. 257-286 (1989)

[7] Cohen, I., Sebe, N., Garg, A., Chen, L. S. and Huang, T. S.: Facial expression recognition from video sequence: temporal and state modeling. Computer Vision and Image Understanding, pp. 160-187 (2003)

[8] Aleksic, P. S. and Katsaggelos, A. K.: Automatic Facial Expression Recognition Using Facial Animation Parameters and Multistream HMMs. IEEE Trans. Information Forensics and Security, vol. 1, No. 1, pp. 3-11 (2006)

[9] Kotsia, I. and Pitas, I.: Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machine. IEEE Trans. Image Processing, vol. 16, no. 1, pp. 172-187 (2007)

[10] Ma, L. and Khorasani, K.: Facial Expression Recognition Using Constructive Feedforward Neural Networks. IEEE Trans. Systems, Man, and Cybernetics, vol. 34, no. 3, pp. 1588-1595 (2004)

[11] Bartlett, M. S., Movella, J. R. and Sejnowski, T. J.: Face Recognition by Independent Component Analysis. IEEE Trans. Neural networks, vol. 13, no. 6, pp. 1450-1464 (2002)

[12] Kwak, K.-C., Pedrycz, W.: Face Recognition using Enhanced Independent Component Analysis Approach. IEEE Trans. Neural Network, vol. 18, no. 2, pp. 530-541 (2007)

[13] Cohn, J. F., Zlochower, A., Lien, J. and Kanade, T.: Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. Psychophysiology, pp. 35-43 (1999)

[14] Linde, Y., Buzo, A., Gray, R.: An algorithm for vector quantizer design. IEEE Trans. Communications, vol. 28(1), pp. 84-94 (1980)

[15] Iwai, Y., Hata, T. and Yachida, M.: Gesture Recognition based on subspace method and Hidden Markov Model. in proceedings of the IEEE/RSJ international conference on Intelligent Robots and Systems, pp. 960-966 (1997)