35

Human Facial Expression Recognition Using Fisher Independent Component Analysis and Hidden Markov Model

Tae-Seong Kim and Jee Jun Lee
Kyung Hee University, Dept. of Biomedical Engineering
Republic of Korea

1. Introduction

Facial expression recognition (FER), a special part of gesture recognition, is one of the fundamental technologies for Human Computer Interface (HCI), which allows computers or other devices to interact with humans in a similar way to human to human interactions (Mitra & Acharya, 2007). A FER system can contribute to a HCI system by responding to the expressive states of a user upon recognizing his or her emotional state from the facial images. Various approaches have been attempted to solve the FER problem so far. In one major category of approaches known as the Facial Action Units (FAU)-based FER, FAUs are identified and classified to understand facial muscle movements according to facial expressions such as joy, anger, disgust, fear, sadness, and surprise. In another category of approaches known as the emotion-specified FER, facial expressions are represented by holistic or local features and the combination of these features are used to recognize each specific expression. In general, FER involves various feature extraction and recognition or classification techniques.

Among the various feature extraction techniques, the most commonly utilized technique so far in the FER research community is Principle Component Analysis (PCA). PCA is a second-order statistical method that produces orthogonal basis of given data. In general, PC images provide global features of facial expressions. In the FAU-based FER, Padgett & Cottrell (1997) applied PCA on facial expression images to identify FAUs and to recognize facial expressions in which their best analysis was performed on the separated face regions such as eyes and mouth. Donato et al. (1999) also employed PCA for FER with their Facial Action Coding System (FACS). In the later approaches of FER, PCA is commonly augmented with Fisher Linear Discriminant (FLD) which is based on the class information which projects the data onto a subspace with the criterion that the between-class scatter of the projected data is maximized and the within-class scatter is minimized. This PCA-FLD augmentation much improved the accuracy of the classification and the performance of FER. For example, Calder et al. (2001) employed PCA and FLD using the pixel intensities of holistic face images and Dubuisson et al. (2002) performed the PCA-FLD based feature extraction on facial images to classify the expressions. However, the nature of PCA, which only relies on the second-order statics and decorrelation of data, limits the performance of

FER. To overcome this deficiency, some higher order statistical methods have been suggested lately.

As one of the higher order feature extraction techniques, lately independent component analysis (ICA) has been utilized extensively for FER tasks due to its ability to extract local features (Donato et al., 1999; Bartllet et al., 2002; Chen & Kotani, 2008; Buciu & Kotropoulos, 2003). ICA is a generalization of PCA which learns the higher order statistics of data, producing the facial features that are statistically independent to each other. It actually performs blind source separation with the assumption that the given data is a linear mixture of sources which produces statistically independent basis and coefficients (Karklin & Lewicki, 2003). For the FAU-based FER, Bartlett et al. (2002) extracted the features using ICA to classify twelve facial actions for facial expressions coding referred to as FACS. Also, Chuang & Shih (2006) utilized ICA to extract the independent features of facial expression images and recognized the upper and lower parts of FAUs and the whole face parts of FAUs. However, these works mostly focused on the successful extraction of FAUs not on the recognition of facial expressions. Also, their works encountered the limitation of AUs due to the fact that the separate facial actions do not directly draw the comparisons with human data (Calder et al., 2000). For the emotion-specified FER, Buciu et al. (2003) reported their highest recognition accuracy of 86.39% when applying ICA on the Japanese female facial expression data (Lyons et al., 1998). However, they only used the peak expressional state of each expression images to recognize the expressions while the temporal information of facial expression changes is ignored. Bartlett et al. (2002) later utilized the enhanced ICA (EICA) with the two different architectures where the first architecture finds spatially local basis images for the faces and the second produces the factorial face codes. In their proposed settings, they found that the local features of face images are sensitive to FER meanwhile the factorial code is preferred for face recognition. Later, Kwak and Pedrycz introduced the fisher version of independent component analysis (FICA) in the two different architectures similar to the way of Bartlett's, and they reported that FICA outperforms the generic ICA and EICA methods in face recognition (Kwak & Pedrycz, 2007). So far, EICA and FICA are successfully utilized for face recognition, but rarely utilized for FER.

As for the recognition techniques in FER, various distance measures such as Euclidian, Mahalonobis, and Cosine distances have been used to assign the facial data to a relevant expression class (Lyons et al., 1998). Neural network systems have been employed to extract FAUs with the Gabor wavelet based features (Tian et al., 2002). Then, a bank of Support Vector Machine (SVM) classifiers was used on the regions of facial expression images in FER (Chuang & Shih, 2006; Kotsia & Pitas, 2007). For instance, Chuang & Shih (2006) used independent components as facial expression features, and SVM was performed to extract FAUs. Recently, Hidden Markov Model (HMM), a method for dynamic and sequential event recognition, has been popularly applied for FER (Otsuka & Ohya, 1997; Aleksic & Katsaqqelos, 2006). Zhu et al. (2002) used HMM to recognize the emotions with moment invariants as their feature. In their work, they report the overall average accuracy of 96.77% where only four expressions were recognized. Cohen et al. (2003) also adopted HMMs for temporal and static modeling of FER. They used the dynamic HMM models with the facial expression features extracted using the Naïve-Bayes classifiers and they reached their best accuracy rate of 73.22% on the Cohn-Kanade database. More recently, Aleksic & Katsaqqelos (2006) introduced a FER system using multi-stream HMMs and demonstrated the best performance of 93.66%. In their system, however, only separated facial animation

parameters such as eyebrow and outer-lip were used as decision cues for FER and they trained the HMM models using FAUs as input parameters. In other words, only the spatial information was utilized for training while the temporally changing patterns are not utilized. In summary, most of these works have used static facial expression images for FER, not utilizing the information of temporal and sequential changes of facial expression features.
In this chapter, we present a novel spatiotemporal modeling approach of FER, fully dealing with sequential facial expression images analyzed with FICA and recognized with HMM (Lee et al., 2008a; Lee et al., 2008b). The fundamental differences between our method and the previous approaches are i) we focus on the sequential holistic facial images to derive local spatiotemporal features so that we can model temporally evolving spatial patterns of the facial expression feature changes and ii) we exploit the natural feature behaviors rather than finding FAUs, and lastly, iii) we come up with the spatiotemporal model of each facial expression via HMM. In our method, spatially local independent features are extracted using FICA from sequential facial expression images. These FICA features are coded of temporally changes of the spatial feature via codebook generation which provides the symbolized spatiotemporal signatures. Then, discrete HMMs are constructed and trained for dynamic FER, encoding the spatiotemporal signatures of various facial expressions. To analyze the performance of our presented approach, some conventional feature extractors including PCA, ICA, EICA, and FLD over PCA (i.e., PCA-LDA) in combination with HMM are tested. Our results show that our presented method significantly outperforms the conventional approaches.

2. Methods

2.1 System Overview

Our FER system consists of i) spatiotemporal FE feature extraction including preprocessing of sequential FE images, ii) codebook generation for temporal signature coding, iii) modeling and training of HMMs, and iv) recognition via HMMs. Fig. 1 shows the overall architecture of our FER system where, L denotes the likelihood, O the observation, and H HMM (Lee et al., 2008a).

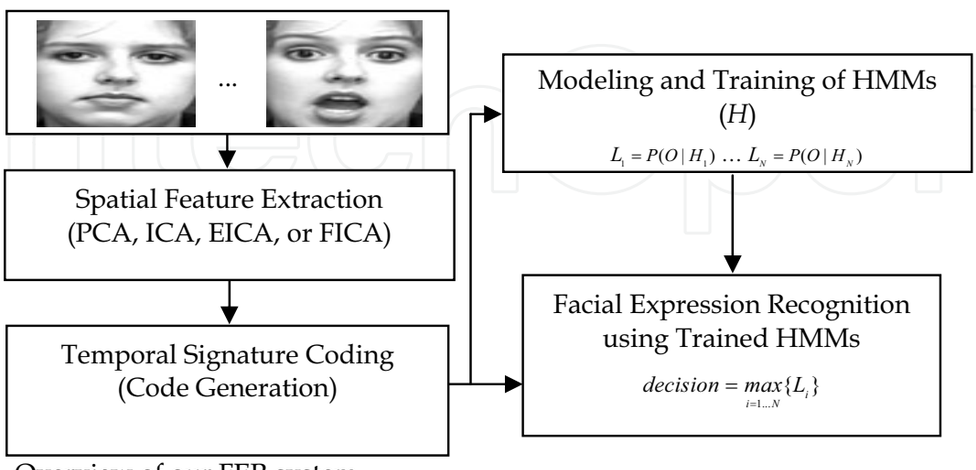


Fig. 1. Overview of our FER system.

2.2 Facial Expression Feature Extraction

The major objective of feature extraction is to find efficient FE spatial features in a lower dimensional subspace and reveal the patterns of temporal behaviors of the FE spatial features corresponding to the changes of facial expression from a neutral to a peak state of each facial expression. In preprocessing, face re-alignment, histogram equalization, and delta image generation have been performed to obtain the input images to our FER system. Then, temporally evolving spatial features are obtained in the feature extraction step.

2.2.1 Preprocessing

In preprocessing of sequential images of facial expressions, image alignment is performed first to realign the common regions of the face. In this study, we have utilized a face alignment approach described in (Zhang & Cottrell, 2004) by manually matching the eyes and mouth of the face in the designated coordinates. The facial images are scaled and translated so that the sum of square distance between the target coordinates and those of the transferred features was minimized: the triangular shape from two eyes and one mouth is scaled and translated to fit the reference location. The typical realigned image consisted of 60 by 80 pixels. Histogram equalization is then performed on the realigned images for lighting correction. Afterwards, the first frame of each input sequence is subtracted from the following frames to obtain the delta image (Donato et al., 1999), reflecting the facial expression change differences over time. Fig. 2 shows an exemplar set of a facial expression sequence of anger and the corresponding delta images.

(a)

(b)

Fig. 2. Sequential facial expression images of anger: (a) a set of the aligned images and (b) the corresponding delta images.

2.2.2 Feature Extraction Using FICA

The key idea of FICA is the combination of ICA and FLD with the purpose to extract the local features such that the facial expression images are represented with them in a low dimensional space: the extracted features must be well separated in space in conjunction with temporally evolving spatial patterns. Extracting theses features via FICA consists of three fundamental stages: i) PCA is first performed for dimension reduction, ii) ICA is applied on the data in the reduced PCA subspace to find statistically independent basis features (i.e., images) for the corresponding facial expression images, and iii) FLD is finally employed to compress the features such that the similar features put close as possible and the different features put as far as possible.

For the first stage of FICA, we initially apply PCA to the data to reduce the dimension of data: PCA is a popular subspace projection method that transforms the high dimensional space to a reduced space by capturing the maximum variability which still contains the associated high order relationships. Basically, PCA basis vectors are obtained by the eigenvectors of the covariance data matrix such that

$$P^T(YY^T)P=\Lambda \tag{1}$$

where Y is the data matrix, Λ the diagonal matrix of eigenvalues, and P the orthogonal eigenvector matrix. The eigenvector associated with the largest eigenvalue indicates the axis of maximum variance and the following eigenvector with the second largest eigenvalue indicates the axis of the second largest variance. Thus, a certain number of eigenvectors associated with the higher eigenvalues defines the reduced subspace.

If we denote the image dataset as $X=(x_1,x_2,...,x_n)^T$ and its corresponding feature vectors in the reduced dimension by $V=(v_1,v_2,...,v_n)^T$. Then, the PCA algorithm finds the following feature vectors such that

$$V=XP_m \tag{2}$$

where P_m is the few selected m eigenvectors. Fig. 3 shows the PCs of the facial expression images.

As the second stage of FICA, ICA is applied to the PCA-reduced facial features to obtain statistically independent features. Basically ICA finds a linear transformation matrix to extract linear combinations of statistically independent sources from a set of random data. If we denote the observed data matrix again by $X=(x_1,x_2,...,x_n)^T$ and the sources of ICs by $S=(s_1,s_2,...,s_m)^T$. The linear ICA assumes that the data X is a linear combination of ICs such that

$$X = MS \tag{3}$$

where M is the matrix of size of $n\times m$ containing the mixing coefficients. The ICA algorithm performed on the data matrix finds an unmixing matrix W and independent source S. That is to say, the source images estimated in the rows of S are used as basis images to represent the dataset and are described as

$$U = WX \tag{4}$$

where U is an estimated independent sources and W represents the linear unmixing matrix.

Generally, ICA performance can be improved with a proper dimension reduction procedure such as PCA as a preprocessing step. ICA performed on the PCA space is also known as

EICA (Liu, 2004). Here, we apply the ICA algorithm on P_m^T which is in the reduced subspace containing the first m eigenvectors. To find the statistically independent basis images, each PCA basis image is the row of the input variables and the pixel values are observations for the variables. Thus,

$$U = W_{ICA} P_m^T \tag{5}$$

where U is the obtained basis images comprised with the coefficient W_{ICA} and the eigenvectors P_m^T. Some of the basis images are shown in Fig. 4. The reconstructed image set $\widetilde{X}$ is then described as

$$\widetilde{X} = VP_m^T = VW_{ICA}^{-1}U. \tag{6}$$

Therefore, the IC representation U can be computed by the rows of the feature vector R followed as

$$R = VW_{ICA}^{-1}. \tag{7}$$

For the final step of FICA, FLD is performed on the IC feature vectors of R. FLD is based on the class specific information which maximizes the ratio of the between-class scatter matrix and the within-class scatter matrix. The formulas for the within, S_W and between, S_B scatter matrix are defined as follows:

$$S_W = \sum_{i=1}^{c} \sum_{r_k \in Ci} (r_k - \tilde{r}_i)(r_k - \tilde{r}_i)^T, \tag{8}$$

$$S_B = \sum_{i=1}^{c} N_i (\tilde{r}_i - r_m)(\tilde{r}_i - r_m)^T \tag{9}$$

where c is the total number of classes, N_i the number of facial expression images, r_k the feature vector from all feature vector R, $\tilde{r}_i$ the mean of class C_i, and r_m the mean of all feature vectors R.

The optimal projection W_d is chosen from the maximization of ratio of the determinant of the between class scatter matrix of the projection data to the determinant of the within class scatter matrix of the projected samples as

$$J(W_d) = |W_d^T S_B W_d| / |W_d^T S_w W_d| \tag{10}$$

where W_d is the set of discriminant vectors of S_B and S_W corresponding to the $c-1$ largest generalized eigenvalues. The discriminant ratio is derived by solving the generalized eigenvalue problem such that

$$S_B W_d = \Lambda S_W W_d \tag{11}$$

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.

Fig. 3. Facial expression representation onto the reduced feature space using PCA. These are also known as eigenfaces.

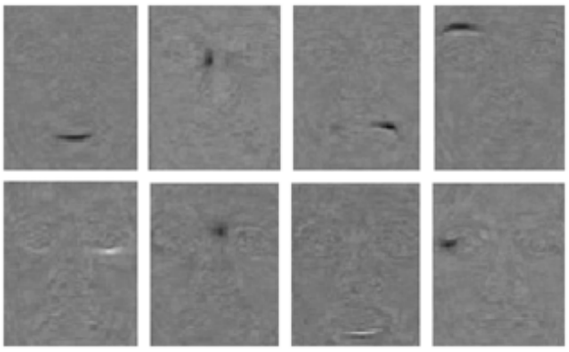


Fig. 4. Sample IC basis images.

Finally, the final feature vector G and the feature vector G_{test} for testing images can be obtained by the criterion

$$G = R W_d^T, \tag{12}$$

$$G_{test} = R_{test} W_d^T = X_{test} P_m W_{ICA}^{-1} W_d^T. \tag{13}$$

As the result of FICA, the vectors of each separated classes can be obtained. As can be seen in Fig. 5, the feature vectors associated with a specific expression are concentrated in a separated region in the feature space showing its gradual changes of each expression. The features of the neutral faces are located in the centre of the whole feature space as the origin of the facial expression, and the feature vectors of the target expressions are located in each

expression region: within each expression feature region contains the temporal variations of the facial features. As shown in Fig. 6, a test sequence of sad expression is projected onto the sad feature region. The projections are evolving according to the time from $P(t_1)$ to $P(t_8)$, describing facial feature changes from the neural to the peak of sad expression.

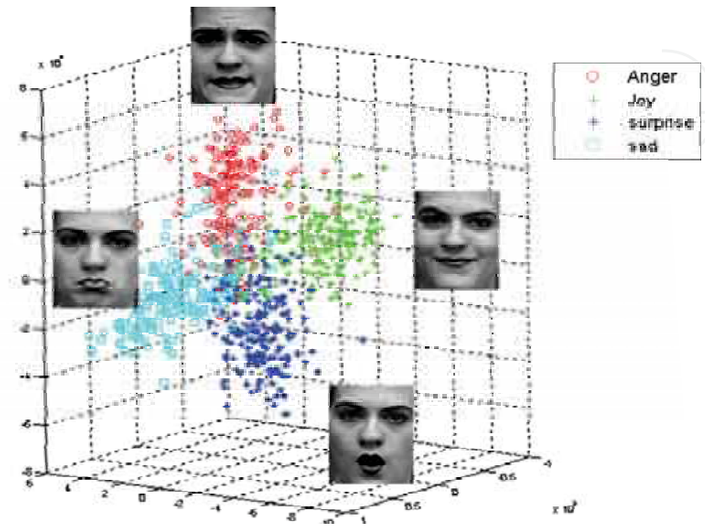
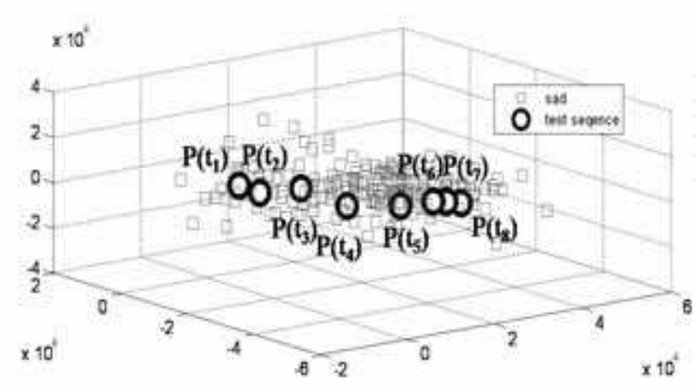


Fig. 5. Exemplar feature plot for four facial expressions.

(a)



(b)

Fig. 6. (a) Test sequences of sad expression and (b) their corresponding projections onto the feature space.

2.3 Spatiotemporal Modelling and Recognition via HMM

Hidden Markov Model (HMM) is a statistical method of modeling and recognizing sequential information. It has been utilized in many applications such as pattern recognition, speech recognition, and bio-signal analysis (Rabiner, 1989). Due to its advantage of modeling and recognizing consecutive events, we also adopted HMM as a modeler and recognizer for facial expression recognition where expressions are concatenated from a neutral state to a peak of each particular expression. To train each HMM, we first perform vector quantization of training dataset of facial expression sequences to model sequential spatiotemporal signatures. Those obtained sequential spatiotemporal signatures are then used to train each HMM, learning each facial expression. More details are given in the following sections.

2.3.1 Code Generation

As HMM is normally trained with the symbols of sequential data, the feature vectors obtained from FICA must be symbolized. The symbolized feature vectors then become a codebook which is a set of symbolized spatiotemporal signature of sequential dataset, and the codebook is then regarded as a reference for recognizing the expression. To obtain the codebook, vector quantization is performed on the feature vectors from the training datasets. In our work, we utilize the Linde, Buzo and Gray (LBG)'s clustering algorithm for vector quantization (Linde et al, 1980). The LBG approach selects the first initial centroids and splits the centroids of the whole dataset. Then, it continues to split the dataset according to the codeword size.

After vector quantization is done, the index numbers are regarded as the symbols of the feature vectors to be modeled with HMMs. Fig. 7 shows the symbols of the codebook with the size of 32 as an example. The index of codeword located in the center of the whole feature space indicates the neutral faces and the other index numbers in each class feature space represents a particular expression reflecting gradual changes of an expression in time.

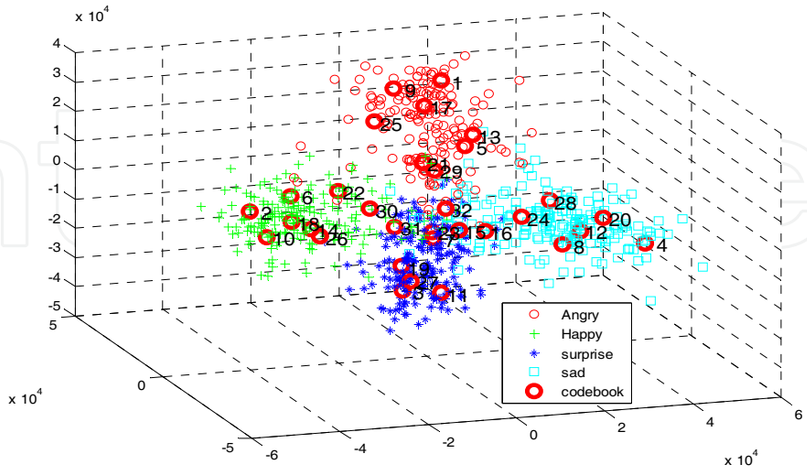


Fig. 7. Exemplary symbols of the codebook in the feature space. Only four out of six expressions are shown for clarity of presentation.

2.3.2 HMM and Training

HMM used in this work is a left-to-right model useful to model a sequential event in a system (Rabiner, 1989). Generally, the purpose of HMM is to determine the model parameter λ with the highest probability of the likelihood $\Pr(O|\lambda)$ when observing the sequential data $O=\{O_1, O_2, \cdots, O_T\}$. A HMM model is denoted as $\lambda = \{A, B, \pi\}$ and each element can be defined as follows (Zhu et al., 2002). Let us denote the states in the model by $S = \{s_1, s_2, \cdots, s_N\}$ and each state at a given time t by $Q = \{q_1, q_2, \cdots, q_t\}$. Then, the state transition probability A, the observation symbol probability B, and the initial state probability π are defined as

$$A = \{a_{ij}\}, \quad a_{ij} = \Pr(q_{t+1} = S_j \mid q_t = S_i), \quad 1 \le i, j \le N, \tag{14}$$

$$B = \{b_j(O_t)\}, \quad b_j = \Pr(O_t \mid q_t = S_j), \quad 1 \le j \le N, \tag{15}$$

$$\pi = \{\pi_j\}, \quad \pi_j = \Pr(q_1 = S_j). \tag{16}$$

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

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\Pr(O \mid \lambda)}, \tag{17}$$

$$\gamma_t(i) = \sum_{j=1}^{N} \xi(i,j) \tag{18}$$

where $\alpha_t(i)$ is the forward variable and $\beta_t(i)$ is the backward variable such that

$$\alpha_1(i) = \pi_i b_i(O_1), \quad (1 \le i \le q) \tag{19}$$

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(O_{t+1}), \quad (t = 1, 2, \ldots, T-1) \tag{20}$$

$$\beta_T(i) = 1, \quad (1 \le i \le q) \tag{21}$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(i). \quad (t = T-1, T-2, \ldots, 1) \tag{22}$$

Using the variables above, we can estimate the updated parameters A and B of the model of λ via estimating probabilities as follows

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \tag{23}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1, O_t=k}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{24}$$

where $\bar{a}_{ij}$ is the estimated transition probability from the state i to the state j and $\bar{b}_j(k)$ the estimated observation probability of symbol k from the state j.

When training each HMM, a training sequence is projected on the FICA feature space and symbolized using the LBG algorithm. The obtained symbols of training sequence are compared with the codebook to form a proper symbol set to train the HMM. Table 1 describes the examples of symbol set for some expression sequences. Symbols in the first two frames are revealing the neutral states whose symbols are on the center of the whole feature subspace and the symbols are assigned into each frame as each expression gradually changes to its target state.

After training the model, the observation sequences $O=\{O_1, O_2, \cdots, O_T\}$ from a video dataset are evaluated and determined by the proper model with the likelihood $\Pr(O|\lambda)$. The likelihood of the observation O given the trained model λ can be determined via the forward variable in the form

$$\Pr(O|\ \lambda) = \sum_{i=1}^{N} \alpha_T(i)\ . \tag{25}$$

The criterion for recognition is the highest likelihood value of each model. Figs. 8 and 9 show the structure and transition probabilities for the anger case before and after training with the codebook size of 32 as an example.

Expression	Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8
Joy	24	32	30	30	14	14	10	10
Sad	32	32	24	16	13	12	4	12
Angry	21	21	13	9	7	8	22	25
Surprise	23	34	34	26	19	19	27	27

Table 1. Example of codebook symbols of the training expression data.

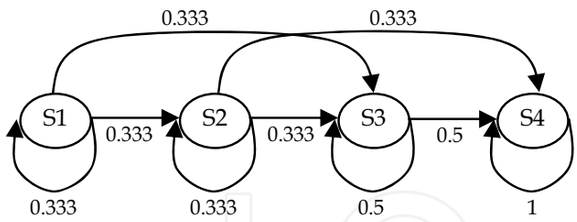


Fig. 8. HMM structure and transition probabilities for anger before training.

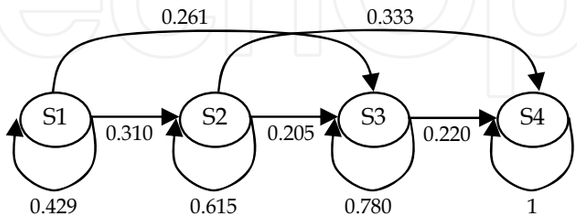


Fig. 9. HMM structure and transition probabilities for anger after training.

3. Experimental Setups

To assess the performance of our FER system, a set of comparison experiments were performed with each feature extraction method including PCA, generic ICA, PCA-LDA, EICA, and FICA in combination with the same HMMs. We recognized six different, yet commonly tested expressions: namely, anger, joy, sadness, surprise, fear, and disgust. The following subsections provide more details.

3.1 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 (Cohn et al., 1999). The image data in the Cohn-Kanade AU-coded facial expression database displays only the frontal view of the face and each subset is comprised of several sequential frames of the specific expression. There are six universal expressions to be classified and recognized. Facial expressions include 97 subjects with the subsets of some expressions. For data preparation, 267 subsets of 97 subjects which contain 8 sequences per expression are selected. A total of 25 sequences of anger, 35 of joy, 30 of sadness, 35 of surprise, 30 of fear, and 25 of disgust sequences are used in training and for the testing purpose, 11 of anger, 19 of joy, 13 of sadness, 20 of surprise, 12 of fear, 12 of disgust subsets are used.

3.2 Recognition Setups for RGB Images

From the database mentioned above, we selected 8 consecutive frames from each video sequences. The selected frames are then realigned with the size of 60 by 80 pixels. Afterwards, histogram equalization and delta image generation were performed for the feature extraction. A total of 180 sequences from all expressions were used to build the feature space.

As we tried to assess our FER system, we applied a total of 180 and 87 image sequences for training and testing respectively. Next, we performed the experiments to empirically determine the optimal number of features and the size of the codebook. To do this, we tested a range of feature numbers selected in the PCA step. Once the optimal number of features was determined, the experiment for the size of the codebook was conducted. We test the performance with the different sizes (2^n, n=4, 5, 6) of the codebook for vector quantization along with HMM in order to determine the optimal settings.
Finally, we compared the different feature extraction methods under the same HMM structure. Previously, PCA and ICA have been extensively explored due to its strong ability of building a feature space, and PCA-LDA has been one of the good feature extractor because of the LDA classifier that finds out the best linear discrimination from the PCA subspace. In this regard, our FICA results have been compared with the conventional feature extraction methods namely PCA, generic ICA, EICA, and PCA-LDA based on the results for the optimal number of features with the same codebook size, and HMM procedure.

3.3 Recognition Setups for Depth Images

Some drawbacks associated with RGB images are known that they are highly affected by lighting conditions and colors causing the distortion of the facial shapes. As one way of overcoming these limitations is the use of depth images. These depth images generally reflect 3-D information of facial expression changes. In our study, we performed preliminary studies of testing depth images and examined their performance for FER. Fig. 10 shows a set of facial expression of surprise from a depth camera called Zcam (www.3dvsystems.com). We tested only four basic expressions in this study: namely, anger, joy, sadness, and surprise using the method presented in the previous section (Lee et al., 2008b).

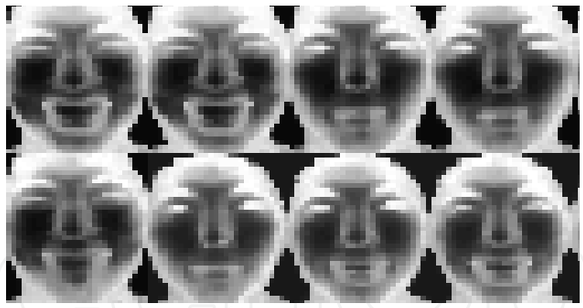


Fig. 10. Depth facial expression images of joy.

4. Experimental Results

Before testing the presented FER system, the system requires setting of two parameters: namely the number of features and the size of codebook. In our experiments, we have tested the eigenvectors in the range from 50 to 190 with the training data and have decided empirically 120 as the optimal number of eigenvectors since it provided the best overall recognition rate. As for the size of the codebook, we have tested the codebook size of 16, 32, and 64, and then decided 32 as the optimal codebook size since it provided the best overall recognition rate for the test data (Lee et al., 2008a).

4.1 Recognition via RGB Images

For recognition comparison between FICA and four other types of conventional feature extraction methods including PCA, ICA, EICA, and PCA-LDA, all extraction methods mentioned above were implemented with the same HMMs for recognition of facial expressions. The results from each experiment in this work represent the best recognition rate with the empirical settings of the selected number of features and the codebook size. For the PCA case, we computed eigenvectors of all the dataset and selected 120 eigenvectors to train the HMMs. As shown in Table 2, the recognition rate using the PCA method was 54.76%, the lowest recognition rate. Then, we employed ICA to extract the ICs from the dataset. Since the ICA produces the same number of ICs as the number of original dimensions of dataset, we empirically selected 120 ICs with the selection criterion of kurtosis values for each IC for training the model. The result of ICA method in Table 3 shows the improved recognition rate than the result of PCA. We also compared the EICA method. We first chose the proper dimension in the PCA step, and processed ICA from the selected eigenvalues to extract the ECIA basis. The results are presented in Table 4, and the total mean of recognition rate from EICA representation of facial expression images was 65.47% which is higher than the generic ICA and PCA recognition rates. Moreover, the best conventional approach PCA-LDA was performed for the last comparison study and it achieved the recognition rate of 82.72% as shown in Table 5. Using the settings above, we conducted the experiment of FICA method implemented with HMMs, and it achieved the total mean of recognition rate, 92.85% and expression labeled as surprise, happy, and sad were recognized with the high accuracy from 93.75% to 100% as shown in Table 6.

Label	Anger	Joy	Sadness	Surprise	Fear	Disgust
Anger	30	0	20	0	10	40
Joy	4	48	8	8	28	4
Sad	0	6.06	81.82	12.12	0	0
Surprise	0	0	0	68.75	12.50	18.75
Fear	0	8.33	50	8.33	33.33	0
Disgust	0	8.33	25	0	0	66.67
Average	54.76					

Table 2. Person independent confusion matrix using PCA (unit : %).

Label	Anger	Joy	Sadness	Surprise	Fear	Disgust
Anger	30	0	10	30	10	20
Joy	4	60	0	0	36	0
Sad	0	6.06	87.88	6.06	0	0
Surprise	0	0	12.50	81.25	0	6.25
Fear	0	25	25	8.33	33.33	8.33
Disgust	0	8.33	25	0	0	66.67
Average	59.86					

Table 3. Person independent confusion matrix using ICA.

Label	Anger	Joy	Sadness	Surprise	Fear	Disgust
Anger	60	0	0	0	20	20
Joy	4	72	8	4	12	0
Sad	0	6.06	87.88	6.06	0	0
Surprise	0	0	12.50	81.25	0	6.25
Fear	0	16.67	16.67	8.33	50	8.33
Disgust	25	8.33	25	0	0	41.67
Average	65.47					

Table 4. Person independent confusion matrix using EICA.

Label	Anger	Joy	Sadness	Surprise	Fear	Disgust
Anger	60	0	10	0	0	30
Joy	0	88	0	0	8	4
Sad	0	6.06	87.88	6.06	0	0
Surprise	0	0	0	93.75	6.25	0
Fear	0	8.33	8.33	8.33	75	0
Disgust	0	0	0	0	8.33	91.67
Average	82.72					

Table 5. Person independent confusion matrix using PCA-LDA.

Label	Anger	Joy	Sadness	Surprise	Fear	Disgust
Anger	80	0	0	0	0	20
Joy	0	96	0	0	4	0
Sad	0	0	93.75	0	6.25	0
Surprise	0	0	0	100	0	0
Fear	0	8.33	0	0	91.67	0
Disgust	0	0	0	0	8.33	91.67
Average	92.85					

Table 6. Person independent confusion matrix using FICA.

As mentioned above, the conventional feature extraction based FER system produced lower recognition rate than the recognition rate of our method, 92.85%. Fig. 11 shows the summary of recognition rate of the conventional compared against our FICA-based method.

4.2 Recognition via Depth Images

A total of 99 sequences were used with 8 images in each sequence, displaying the frontal view of the faces. A total of 15 sequences for each expression were used in training, and for the testing purpose, 10 of anger, 10 of joy, 8 of surprise, and 11 of sadness subsets were used. We empirically selected 60 eigenvectors for dimension reduction, and test the performance with the codebook size of 32. On the data set of RGB and depth facial expressions of the

same face, we applied our presented system to compare the FER performance. Table 7 and 8 show the recognition results for each case. More details are given in Lee at al. (2008b).

-

Fig. 11. Recognition rate of facial expressions using the conventional feature extraction methods and the presented FICA feature extraction method.

Label	Anger	Joy	Sadness	Surprise
Anger	100	0	20	0
Joy	10	90	0	0
Sadness	9.09	9.09	81.82	0
Surprise	12.5	12.5	0	75
Average	86.5			

Table 7. Person independent confusion matrix using the sequential RGB images (unit :%).

Label	Anger	Joy	Sadness	Surprise
Anger	100	0	0	0
Joy	0	100	0	0
Sadness	0	0	100	0
Surprise	0	0	0	100
Average	100			

Table 8. Person independent confusion matrix using the sequential depth images.

5. Conclusion

In this work, we have presented a novel FER system utilizing FICA for facial expression feature extraction and HMM for recognition. Especially in the framework of FICA and HMM, the sequential spatiotemporal feature information from holistic facial expressions is modeled and used for FER. The performance of our presented method has been investigated on sequential datasets of six facial expressions. The result shows that FICA can extract optimal features which are well utilized in HMM, outperforming all other conventional feature extraction methods. We have also applied the presented system to 3-D depth facial expression images and showed its improved performance. We believe that our presented

FER system should be useful toward real-time recognition of facial expressions which could be also useful in many other applications of HCI.

6. Acknowledgement

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA-2009-(C1090-0902-0002)).

7. Reference

Aleksic, P. S. & Katsaggelos, A. K. (2006). Automatic facial expression recognition using facial animation parameters and multistream HMMs, *IEEE trans. Information and Security*, Vol. 1, Nol. 1, pp. 3-11, ISSN. 1556-6013

Bartlett, M. S.; Donato, G. ; Movellan, J. R.; Hager, J. C.; Ekman, P. & Sejnowski, T. J. (1999). Face Image Analysis for Expression Measurement and Detection of Deceit, *Proceedings of the 6th Joint Symposium on Neural Computation*, pp. 8-15

Bartlett, M. S.; Movellan, J. R. & Sejnowski, T. J. (2002). Face Recognition by Independent Component Analysis, *IEEE trans. Neural Networks*, Vol. 13, No. 6, pp. 1450-1464, ISSN. 1045-9227

Buciu, I.; Kotropoulos, C. & Pitas, I. (2003). ICA and Gabor Representation for Facial Expression Recognition, *Proceedings of the IEEE*, pp. 855-858

Calder, A. J.; Young, A. J.; Keane, J. & Dean, M. (2000). Configural information in facial expression perception, *Journal of Experimental psychology, Human Perception and Performance. Human perception and performance*, Vol. 26, No. 2, pp. 527-551

Calder, A. J.; Burton, A. M.; Miller, P.; Young, A. W. & Akamatsu, S. (2001). A principal component analysis of facial expressions, *Vision Research*, Vol.41, pp. 1179-1208

Chen, F. & Kotani, K. (2008). Facial Expression Recognition by Supervised Independent Component Analysis Using MAP Estimation, *IEICE trans. INF. & SYST.*, Vol. E91-D, No. 2, pp. 341-350, ISSN. 0916-8532

Chuang, C.-F. & Shih, F. Y. (2006). Recognizing Facial Action Units Using Independent Component Analysis and Support Vector Machine, *Pattern Recognition*, Vol. 39, No. 9, pp. 1795-1798, ISSN. 0031-3203

Cohen, I.; Sebe, N.; Garg, A; Chen, L. S. & Huang, T. S. (2003). Facial expression recognition from video sequences: temporal and static modeling, *Computer Vision and Image Understanding*, Vol. 91, ISSN. 1077-3142

Cohn, J. F.; Zlochower, A.; Lien, J. & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding, pp. 35-43, *Psychophysiology*, Cambridge University Press

Danato, G.; Bartlett, M. S.; Hagar, J. C.; Ekman, P. & Sejnowski, T. J. (1999). Classifying Facial Actions, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21(10), pp. 974-989

Dubuisson, S.; Davoine, F. & Masson, M. (2002). A solution for facial expression representation and recognition, *Signal Processing: Image Communication*, Vol. 17, pp. 657-673

Lee, J. J.; Uddin, M. D. & Kim, T.-S. (2008a). Spatiotemporal human facial expression recognition using fisher independent component analysis and Hidden Markov Model, *Proceedings of the IEEE Int. Conf. Engineering in Medicine and Biology Society, pp. 2546-2549*

Lee, J. J.; Uddin, M. D.; Truc P. T. H. & Kim, T.-S. (2008b). Spatiotemporal Depth Information-based Human Facial Expression Recognition Using FICA and HMM, *Int. Conf. Ubiquitous Healthcare, IEEE, Busan, Korea*

Lyons, M.; Akamatsu, S.; Kamachi, M. & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets, *Proceedings of the Third IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 200-205

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-286

Linde, Y.; Buzo, A. & Gray, R. (1980). An Algorithm for Vector Quantizer Design, *IEEE Transaction on Communications*, Vol. 28, No. 1, pp. 84–94, ISSN. 0090-6778

Liu, C. (2004). Enhanced independent component analysis and its application to content based face image retrieval, *IEEE trans. Systems, Man, and Cybernetics,* Vol. 34, No. 2, pp. 1117-1127

Karklin, Y. & Lewicki, M. S. (2003). Learning higher-order structures in natural images, *Netw. Comput. Neural Syst.*, Vol. 14, pp. 483-499

Kwak, K. C. & Pedrycz, W. (2007). Face recognition using an enhanced independent component analysis approach, *IEEE Trans. Neural Network*, Vol. 18, pp. 530-541, ISSN. 1045-9227

Kotsia, I. & Pitas, I. (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machine, *IEEE trans. Image Processing*, Vol. 16, pp. 172-187, ISSN. 1057-7149

Mitra, S. & Acharya, T. (2007). Gesture Recognition: A survey, *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 37, No. 3, pp. 331-324, ISSN. 1094-6977

Otsuka, T. & Ohya, J. (1997). Recognizing multiple person's facial expressions using HMM based on automatic extraction of significant frames from image sequences. *Proceedings of the IEEE Int. Conf. Image Processing*, pp. 546-549

Padgett, C. & Cottrell, G. (1997). Representation face images for emotion classification, *Advances in Neural Information Processing Systems,* vol. 9, Cambridge, MA, MIT Press

Tian, Y,-L.; Kanade, T. & Cohn, J. F. (2002). Evaluation of Gabor wavelet based facial action unit recognition in image sequences of increasing complexity, *Proceedings of the 5th IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 229-234

Zhang, L. & Cottrell, G. W. (2004). When Holistic Processing is Not Enough: Local Features Save the Day, *Proceedings of the Twenty-sixth Annual Cognitive Science Society Conference*

Zhu, Y.; De Silva, L. C. & Ko, C. C. (2002). Using moment invariants and HMM in facial expression recognition, *Pattern Recognition Letters*, Vol. 23, pp. 83-91, ISSN. 0167-8655

Biomedical Engineering

Edited by Carlos Alexandre Barros de Mello

ISBN 978-953-307-013-1

Hard cover, 658 pages

Publisher InTech

Published online 01, October, 2009

Published in print edition October, 2009

Biomedical Engineering can be seen as a mix of Medicine, Engineering and Science. In fact, this is a natural connection, as the most complicated engineering masterpiece is the human body. And it is exactly to help our “body machine” that Biomedical Engineering has its niche. This book brings the state-of-the-art of some of the most important current research related to Biomedical Engineering. I am very honored to be editing such a valuable book, which has contributions of a selected group of researchers describing the best of their work. Through its 36 chapters, the reader will have access to works related to ECG, image processing, sensors, artificial intelligence, and several other exciting fields.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen