Human Emotion Recognition Using PCA, ICA and NMF

Paresh Chandra Barman, Chandra Shekhar Dhir, Soo-Young Lee

Abstract—Recognizing human facial expression and emotion by computer is an interesting and challenging problem. In this paper, Principal component analysis (PCA), independent component analysis (ICA) and Non-negative Matrix Factorization (NMF) are exploited for feature extraction of face images. The features are low-dimensional representation of the original multivariate high dimensional data with minimal loss in data representation [1]. In addition, the features are also required to give good class discrimination for recognition experiments. Feature selection based on information gain criterion has been studied for finding efficient features to improve the classification performance of face recognition tasks [2]. This work presents a detailed study on the application of information gain for efficient feature selection and is compared with Fisher criterion. Individual, emotion recognition experiments using face images of Korean nationals are performed to compare the two feature selection criteria (Information gain and Fisher criterion). The face images of Korean nationals are obtained from the Postech Faces '01 (PF01) database [3].

Keywords—Human emotion recognition, NMF, PCA, ICA.

1 Introduction

HUMAN-computer intelligent interaction (HCII) is an emerging field of science aimed at providing natural ways for humans to use computers as aids. It is argued that for the computer to be able to interact with humans, it needs to have the communication skills of humans. One of these skills is the ability to understand the emotional state of the person. The most expressive way humans display emotions is through facial expressions. In recent years there has been a growing interest in improving all aspects of the interaction between humans and computers. Faces are much more than keys to individual identity. Human beings possess and express emotions in day to day interactions with others.

The problem of handling facial images for recognition is also due to the large pixel size and demands on computational resources. Therefore, several researchers have exploited linear low-dimensional representation of images using orthogonal basis by Principal component analysis (PCA) and independent features using Independent component analysis (ICA). ICA is a multivariate approach of data representation with statistically independent features [5]. It has also found wide applications in the field on blind source separation. In [15] Cauchy Naive Bayes Classifier has been used for emotion recognition.

Face recognition using eigenfaces was one of the preliminary works in the field of face recognition using low dimensional features [6]. While ICA and NMF give more local representation of data, PCA features are global. Although, feature extraction methods helps in low-dimensional representation of multivariate data by using unsupervised methods, all the feature may not be important for classification. Hence, the features extracted give a good data representation and the next task is to select proper features for good classification performance. Variance of features and Fisher criterion (between class variance over within class variance) has been widely used for selection of features for classification [7]. Variance of feature gives an estimate of the power of the features that can be considered important. Recently, feature selection criterion based on information gain to face images is studied for efficient feature selection and improving classification performance [2].

A comparative study on the performances of the features will be done on feature selection criterion based on information gain to facial images is studied for efficient feature extraction and improving classification performance. Information gain criterion is extensively studied in the field of text categorization [8]. The motivation to apply information gain was to maximize the information between the class and the given features. Since, the ICA features are independent we can get a score value for each feature based on information gain and a proper number of features can be selected. However, in case of PCA and NMF features which are dependent this criterion in its crude form cannot be applied. The information gain criterion presented in the paper does not consider the dependency among features which may be present in case of PCA and NMF features.

In section II, feature extraction using PCA, ICA and NMF are discussed. Section III concentrates on feature selection using Fisher criterion and proposed information gain for application to face recognition. Experimental setup is presented in section IV followed by results in section V. Sec-

- *Paresh Chandra Barman is with the Department of Information and Communication Engineering, Kushtia, Bangladesh. E-mail: pcb@iceiu.info.*
- *Chandra Shekhar is with the Department of Bio and Brain Engineering, KAIST, Republic of Korea.*
- *Soo-Young Lee is with the Department of Electrical Engineerig, KAIST, Republic of Korea.*

Manuscript received on 31 July 2010 and accepted for publication on 27 August 2010.

tion V1 discusses the experimental results followed by conclusions.

A brief overview on the two dataset used is given below:

In this paper we utilize the publicly available Postech Faces '01 (PF01) [3] face database. PF01 has 56 males' faces with 4 emotions of each, and the resolution is 150×150. The four different emotion categories are: *Smile*, *Surprised*, *Sad* and *Closed eyes* as shown in figure1. There are 168 training samples per class, 56 test samples per class. We used 4 fold cross-validations for recognition process.

Figure 1: Example of 4 facial emotion expressions.

2 Feature Extraction Methods

Classification of emotion images with high resolution is a difficult problem and is computationally demanding if no pre-processing is done on the raw images. In an attempt to reduce the computational burden several efficient feature extraction methods have been studied. PCA is a well known method in image processing and has been widely used to extract meaningful features. Working on same philosophy, ICA and NMF are also used for representing low dimension feature space.

2.1 Principal Component Analysis (PCA)

PCA tries to obtain a representation of the inputs based on uncorrelated variables. The eigenfaces, E, are the orthogonal axis or the uncorrelated basis for representation of face data. Let X_{PxN} be the data matrix consisting on image data (each column consists of pixels of one image). Here P is the number of pixels and N is the number of images. Using singular value decomposition, the data can be represented as:

$$X = EDV^T = \sum_{r=1}^{R} \lambda_r e_r v_r^T \qquad (1)$$

where D is a diagonal matrix with elements λ_r, and V is a orthonormal matrix with elements $\{v_{ij}\}_{NxR}$. Hence, we can choose the number of basis (eigenfaces) depending on the square of the value of λ_r, which is the same as the eigenvalues of covariance matrix of X. Thus, the dimension of the images is reduced from P pixels to R coefficients (P << R) with minimal loss in data representation.

2.2 Independent Component Analysis (ICA)

ICA is an unsupervised learning algorithm that tries to remove higher order dependencies among the basis of natural scenes [6]. The observed image is assumed to be a linear combination of basis images scaled by independent coefficients. The task is to find basis, A= [a_1 a_2 …a_P], such that the coefficients, U_{PxN} are independent of each other and **X ≈ AU**, where M is the number of basis and N is the number of images. This is also referred as factorial code representation [1].

Considering the high dimensionality of X it is often beneficial to represent the data by the coefficients pertaining to the important principal components. PCA is performed on the data and the dimension is reduced to R×N. The input to the ICA network is the low-dimensional representation of X which comprises of R eigenfaces. Let the low-dimensional representation of X be XPCA. ICA is performed using FASTICA algorithm to find the independent features. FASTICA uses the criterion of maximization of negentropy and provides fast convergence [11, 12]. Since, ICA features extracted using informax learning algorithm is reported to be similar to the performance using FastICA the informax learning algorithm is not considered in the simulations [13].

2.3 Nonnegative Matrix Factorization (NMF)

Given a non-negative data matrix $X_{n_x m}$ is factorized into a non-negative basis factors, $W_{n_x r}$, and a coefficient factor Hr×m, such that: X ≈ WH or

$$X_{ij} \approx (WH)_{ij} = \sum_a W_{ia} H_{aj} \qquad (2)$$

where r is chosen as $r < nm/(n+m)$. To obtain the factors W and H we used the well known multiplicative update rule of [4].

3 Feature Selection Methods

The basis of orthogonal decomposition of data matrix using PCA is arranged in the order of the descending magnitude of eigenvalues. Fisher criterion and Information Gain is used to select good basis from already extracted basis mentioned in section 2 for classification performance.

3.1 Class Discrimination using Fisher Criterion

Class With results from nearest mean classifier (NMC), efficient feature selection or reducing the number of basis based on variation ratio (Fisher criterion score). The class discriminability ratio is given as [1]

$$r = \frac{\sum_j \left(\bar{x}_j - \bar{x}\right)^2}{\sum_j \sum_i \left(x_{ij} - \bar{x}_j\right)^2} \qquad (3)$$

where, x_j are the features, $\bar{x}_j$ is the mean of feature corresponding to class j, and $\bar{x}$ is the mean of the feature. The bases are rearranged in the descending order based on the value of r for each feature.

3.2 Class Discrimination using Information Gain

Feature selection for proper classification can also be done using information gain criterion. Information gain is defined as [2]

$$\begin{aligned} I &= \sum_j p(j,x)\log_2 \frac{p(j,x)}{p(j)p(x)} \\ &= \sum_j p(j)p(x|j)\log_2 \frac{p(x|j)}{p(x)} \end{aligned} \quad (4)$$

where p(j,x) is the joint distribution of the class j and feature x. p(x) is the probability distribution of the feature and p(j) is the probability distribution of the class. The features rearranged based on the value of I and the performance of feature selection using information gain is compared with fisher criterion.

4 Experimental Setup

Experiments are performed on PF01 (Postech Faces) as described in section 1. In this recognition experiment, the features obtained from the training data set will be utilized to recognize the test image labeled according to the emotions. The expressions images of 56 male individuals are used for the emotion recognition problems. Each cross validation training data set consists of 168 images consisting of 42 images per emotion. The cross validation test data set consists of 56 images with 16 images per emotion. The following classifier has been used to empirically find the classification performance.

4.1 Nearest Mean Classifier (NMC)

Four distances metric: L_1-metric, L_2-metric, cosine distance and Mahanalobis distance are used to test the performance of nearest neighbor classifier.

L2-metric is the Euclidean distance between the two vectors a and b and is given as:

$$L_2(a,b) = \sqrt{\sum_{i=1}^{k}(a_i - b_i)^2} \ .$$

Cosine distance (CD) is the similarity measure between the two vectors and is given as $CD(a,b) = -\frac{\langle a,b \rangle}{\|a\| \cdot \|b\|}$

4.2 Support Vector Machines

Classification analysis with Support Vector Machine (SVM) will also be performed with inputs having fixed number of basis. SVM light is used in our experimental study [14]. The inputs to the SVM are features extracted using PCA and ICA. Linear, polynomial and Gaussian kernels are used for the SVM classifier. For the linearly separable case, SVM provides the optimal hyper-plane that separates the training patterns. The optimal hyper-plane maximizes the sum of the distances to the closest positive and negative training patterns. This sum is called margin. For the non-linearly separable case, the training patterns are mapped onto a high-dimensional space using a kernel function. In this space the decision boundary is expected to be linear. The most commonly used kernel functions are polynomials, Radial basis function (RBF) and sigmoid functions.

5 Experimental Results

Figures and For all tables the following notations are applicable: M- Maximum recognition performance is obtained using M features, where the features are arranged in the descending order of fisher score (FS (f_1) > FS(f_2) >…> FS(f_M). (N, L)- Maximum recognition performance is obtained using N number of bins and L number of features. Features are arranged in descending order of mutual information with respect to class, i.e., MI(f_1,C)> MI(f_2,C)>……> MI(f_L,C). For example, Table 1 represents recognition performance for the PCA extracted features. Let us consider 20 is the predefined PCA dimension, it means we extracted 20 PCA feature vectors then out of these 20 feature vectors using Fisher Score or Mutual Information score we used M selected features. In case of the element (3, 3) of

Table 1. Emotion recognition with PCA features

Training Set		{1}		{2}		{3}		{4}		Average	
# of features	FSC Dist.	FS (M)	MI (N, L)	FS (M)	MI (N, L)	FS (M)	MI (N, L)	FS (M)	MI (N, L)	FS	MI
20	EUC	83.93 (14)	87.5 (24,14)	85.71 (16)	91.07 (20.13)	83.93 (17)	83.93 (4,18)	89.89 (18)	89.29 (12,10)	85.71	**87.95**
	COS	82.14 (10)	87.5 (22,15)	87.5 (9)	94.64 (24,14)	82.14 (14)	85.71 (2,16)	89.28 (7)	92.86 (6,8)	85.26	**90.18**
30	EUC	85.71 (15)	89.29 (4,15)	85.71 (15)	87.5 (14,19)	83.93 (17)	85.71 (8,25)	87.5 (16)	91.07 (8,11)	85.71	**88.40**
	COS	82.15 (10)	89.29 (12,19)	89.29 (22)	94.64 (22,9)	82.15 (14)	83.92 (2,17)	89.29 (7)	94.64 (8,11)	85.71	**90.62**
40	EUC	83.92 (13)	89.29 (12,14)	85.71 (16)	89.29 (20,14)	85.71 (20)	85.71 (2,27)	87.5 (16)	89.29 (8,12)	85.71	**88.40**
	COS	82.14 (10)	87.5 (4,18)	91.07 (26)	94.64 (10,14)	83.92 (34)	85.71 (30.26)	89.29 (7)	94.64 (16,8)	86.60	**90.62**
50	EUC	83.92 (13)	89.29 (12,14)	87.5 (21)	91.07 (10 19)	85.71 (20)	87.5 (24,37)	87.5 (16)	91.07 (8,11)	86.16	**89.73**
	COS	82.15 (10)	89.29 (4,22)	91.07 (27)	96.43 (34,17)	85.71 (40)	85.71 (6,32)	89.28 (7)	94.64 (16,8)	87.05	**91.51**
100	EUC	83.92 (13)	91.07 (4, 29)	87.5 (22)	91.07 (4, 18)	83.92 (18)	85.71 (2, 41)	87.5 (16)	91.07 (6,24)	85.71	**89.73**
	COS	82.15 (10)	92.86 (4,45)	91.07 (35)	98.21 (34,27)	82.14 (14)	85.71 (6, 85)	89.29 (7)	92.86 (8, 20)	86.16	**92.41**

Table 2. Emotion recognition with ICA Arch-I features

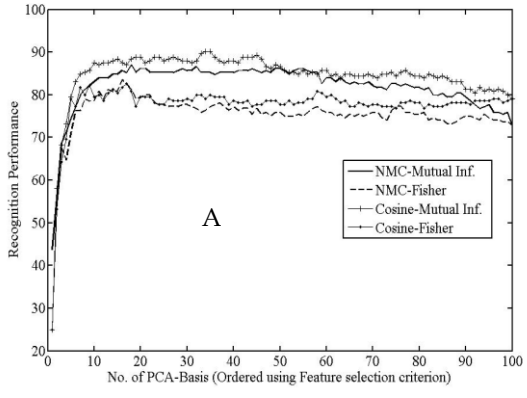
Training Set		{1}		{2}		{3}		{4}		Average	
# of features	FSC Dist.	FS	MI	FS	MI	FS	MI	FS	MI	FS	MI
20	EUC	83.92 (18)	91.07 (40,17)	83.93 (14)	89.29 (22,15)	91.07 (16)	91.07 (34,15)	91.07 (15)	89.29 (6,17)	87.5	**90.17**
	COS	83.92 (11)	89.28 (2,12)	85.71 (15)	83.93 (28.17)	89.29 (16)	87.5 (24,15)	91.07 (14)	83.93 (2,16)	87.5	**86.16**
30	EUC	82.14 (26)	83.93 (4,23)	91.07 (17)	92.86 (6,27)	85.71 (26)	83.93 (2,30)	82.14 (27)	83.93 (14,28)	85.26	**86.16**
	COS	82.14 (29)	83.92 (10,22)	89.29 (28)	91.07 (12,25)	83.93 (16)	89.29 (4,18)	78.57 (23)	92.86 (36,18)	83.48	**89.28**
40	EUC	83.93 (7)	89.28 (2,23)	80.35 (20)	83.93 (4,24)	82.14 (35)	89.29 (18,28)	92.85 (19)	87.5 (14,30)	84.82	**87.5**
	COS	83.93 (25)	87.5 (2,19)	87.5 (24)	87.5 (34,39)	83.93 (24)	89.29 (2,33)	92.85 (19)	91.07 (4,27)	87.05	**88.84**
50	EUC	89.29 (27)	87.5 (2,20)	80.35 (24)	85.71 (12,24)	85.71 (29)	87.5 (16,28)	89.29 (35)	92.86 (20,22)	86.16	**88.39**
	COS	87.5 (36)	91.07 (20,16)	87.5 (39)	87.5 (12,30)	85.71 (47)	87.5 (2,38)	91.07 (38)	92.86 (38,23)	87.95	**89.73**
100	EUC	85.71 (22)	91.07 (28,26)	83.93 (51)	87.5 (30,35)	83.93 (55)	91.07 (22,34)	85.71 (49)	89.28 (36,71)	84.82	**89.73**
	COS	87.5 (46)	89.29 (24,64)	87.5 (32)	89.29 (4,89)	85.71 (78)	91.07 (20,50)	92.86 (63)	92.86 (2,67)	88.39	**90.62**

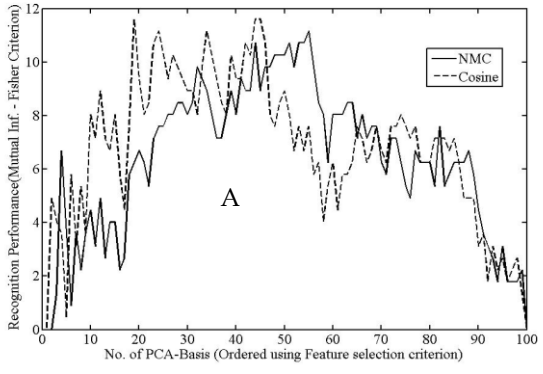
Table 3. Emotion recognition with ICA Arch-II features

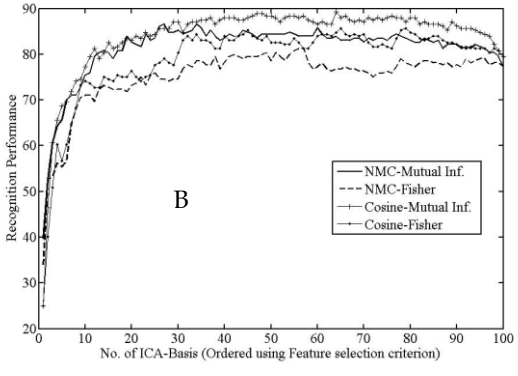
Training Set		{1}		{2}		{3}		{4}		Average	
# of features	*FSC* Dist.	*FS*	*MI*	*FS*	*MI*	*FS*	*MI*	*FS*	*MI*	*FS*	*MI*
20	**EUC**	89.29 (3)	91.07 (4,4)	91.07 (3)	92.86 (14,8)	96.43 (9)	96.43 (12,9)	92.86 (3)	96.43 (38,11)	92.41	**94.19**
	COS	87.5 (3)	91.07 (8,8)	96.49 (6)	94.64 (8,7)	94.64 (9)	96.43 (2,11)	94.64 (8)	96.43 (26,11)	93.30	**94.64**
30	**EUC**	82.14 (11)	85.71 (2,15)	92.86 (3)	94.64 (18,9)	89.29 (7)	91.07 (14,13)	91.07 (9)	91.07 (14,10)	88.84	**90.62**
	COS	82.14 (10)	87.50 (16,12)	92.86 (26)	96.43 (30,7)	89.29 (6)	87.5 (2,7)	91.07 (9)	91.07 (2,9)	88.84	**90.62**
40	**EUC**	85.71 (30)	85.71 (26,16)	83.93 (35)	87.5 (2,36)	83.93 (25)	83.92 (4,28)	89.29 (17)	89.28 (34,21)	85.71	**86.60**
	COS	83.93 (11)	83.92 (26,16)	83.93 (32)	87.5 (2,30)	82.14 (32)	82.14 (2,29)	87.5 (22)	89.29 (14,22)	84.37	**85.71**
50	**EUC**	89.29 (27)	83.93 (4,30)	82.14 (28)	82.14 (16,44)	83.93 (36)	85.71 (2,31)	89.29 (24)	83.93 (8,23)	86.16	**83.93**
	COS	85.71 (29)	87.5 (20,24)	78.57 (21)	82.14 (12,41)	87.5 (42)	87.5 (2,30)	85.71 (24)	87.5 (8,23)	84.37	**86.16**
100	**EUC**	75 (97)	78.57 (2,89)	76.79 (56)	67.85 (26,26)	67.86 (17)	64.29 (8,18)	73.21 (38)	73.21 (1,89)	73.21	**70.98**
	COS	76.78 (74)	76.78 (36,55)	83.92 (30)	82.14 (2,64)	73.21 (94)	73.21 (2,71)	78.57 (66)	78.57 (2,100)	78.12	**77.67**

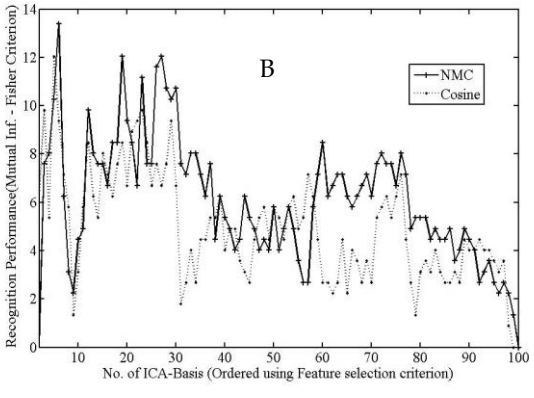
Table 4. Emotion recognition with NMF features

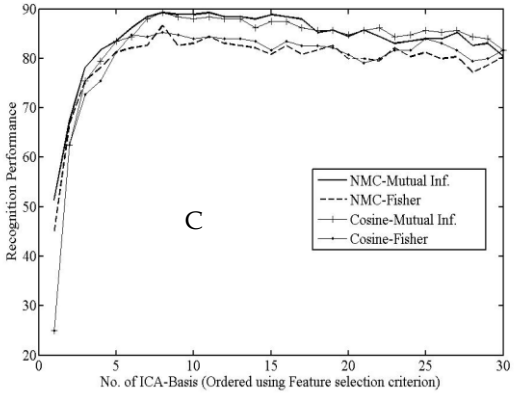
Training Set		{1}		{2}		{3}		{4}		Average	
# of features	*FSC* Dist.	*FS*	*MI*	*FS*	*MI*	*FS*	*MI*	*FS*	*MI*	*FS*	*MI*
20	**EUC**	96.43 (5)	96.43 (4,5)	85.71 (14)	92.85 (26,14)	96.43 (9)	92.85 (2,11)	92.86 (11)	96.43 (20,10)	92.85	**94.64**
	COS	91.07 (16)	92.86 (2,14)	89.29 (10)	92.86 (18,11)	94.64 (9)	94.64 (6,10)	89.29 (8)	92.86 (18,10)	91.07	**93.30**
30	**EUC**	87.5 (7)	92.86 (10,10)	94.64 (19)	91.07 (4,20)	87.5 (11)	91.07 (6,20)	91.07 (12)	87.5 (6,15)	90.17	**90.62**
	COS	89.29 (7)	92.85 (10,10)	92.86 (19)	92.85 (2,23)	87.5 (21)	91.07 (6,232)	89.29 (12)	91.07 (10,12)	89.73	**91.96**
40	**EUC**	89.29 (18)	91.07 (14,23)	85.71 (16)	91.07 (6,27)	83.93 (15)	87.5 (4,19)	83.93 (12)	87.5 (32,19)	85.71	**89.28**
	COS	87.5 (15)	92.86 (4,12)	87.5 (25)	91.07 (6,20)	85.71 (15)	87.5 (4,16)	82.14 (12)	87.5 (14,15)	85.71	**89.73**
50	**EUC**	82.14 (17)	85.71 (6,22)	92.86 (32)	91.07 (18,46)	82.14 (36)	89.29 (10,27)	75 (32)	82.14 (4,18)	83.04	**87.05**
	COS	85.71 (13)	87.50 (12,24)	91.07 (28)	92.86 (12,44)	83.93 (9)	91.07 (38,15)	78.57 (36)	82.14 (4,19)	84.82	**88.39**
100	**EUC**	91.07 (27)	89.29 (10,25)	83.93 (63)	91.07 (2,76)	89.29 (54)	87.5 (4,81)	89.29 (31)	89.29 (4,53)	88.39	**89.29**
	COS	89.29 (27)	94.64 (10,25)	89.29 (98)	89.29 (2,73)	87.5 (64)	87.5 (24,73)	91.07 (30)	87.5 (4,39)	89.29	**89.73**

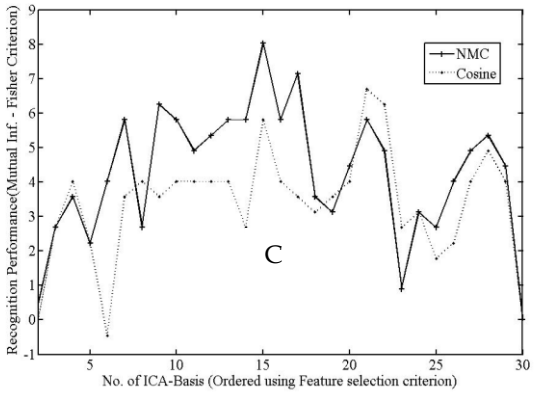












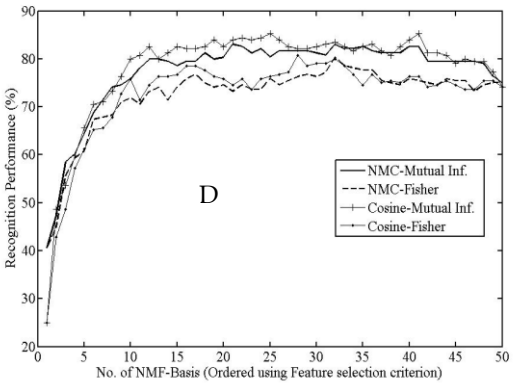


Figure 2: Comparison among feature selection methods (Mutual Information and Fisher Criteria) with NMC (Nearest mean classifier) and Cosine (cosine similarity measure) classifier for case: A. Features extracted by PCA, B. Features extracted by ICA Arch-I, C. Features extracted by ICA Arch-II, and D. Features extracted by NMF

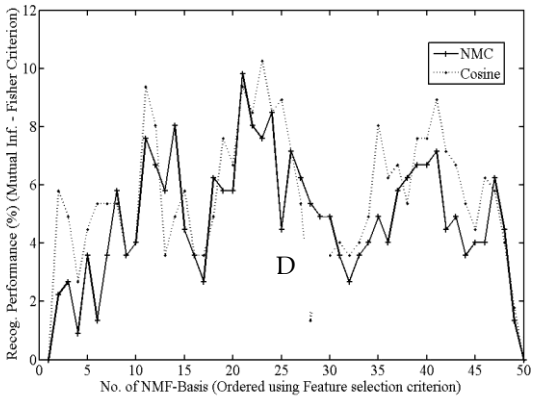


Figure 3: Improvement of performance by Mutual information over Fisher criterion feature selection method with NMC (Nearest mean classifier) and Cosine (cosine similarity measure) classifier for case: A. Features extracted by PCA, B. Features extracted by ICA Arch-I, C. Features extracted by ICA Arch-II, and D. Features extracted by NMF

table 1, 83.93 (14) means: M = 14 features with top values of fisher score gives recognition performance of 83.93% for training set1. We get recognition performance of 87.5% with L = 14 features (features are the ones with maximum mutual information value) and N = 24 bins are used to estimate the probability distribution function (pdf) of the PCA extracted features. Similar explanation is also applicable for other cases.

From the tables it has been observed that in maximum cases the feature selection by mutual information is better than Fisher criteria. The two criterions are different, hence, they need not be compared, and however, the important factor is the decay in the values of the two scores. Figure 2 shows the emotion recognition performance for four different approaches (PCA, ICA-I, ICA-II and NMF) of dimension reduction or feature extraction methods. Two different feature selection criteria (Mutual Information (MI) and Fisher Selection criteria (FS)) and two different classifiers (NMC and cosine similarity) have been used for each feature extraction process. The X-axis of figure indicates the number of input features to the classifier, which are selected by MI or FS followed by one of the feature extracted methods, and the Y-axis represents the recognition rate. Figure 1.A. is the combination of PCA+(MI or FS)+(NMC or Cosine), 1.B. is the combination of ICA Arch-I+(MI or FS)+(NMC or Cosine), 1.C. is the combination of ICA Arch-II+(MI or FS)+(NMC or Cosine), and 1.D. is the combination of NMF+(MI or FS)+(NMC or Cosine).

Figure 3 represents the emotion recognition performance between feature selection methods of MI and FS for both classifiers. 2.A. PCA+(MI-FS)+ (NMC or Cosine), 2.B. ICA Arch-I+ (MI-FS) +(NMC or Cosine), 2.C. ICA Arch-II +(MI-FS)+(NMC or Cosine) and 2.D. NMF+ (MI-FS) + (NMC or Cosine).

6 Conclusions and Future Works

Feature extraction algorithms like PCA, ICA, and NMF are evaluated in case of emotion. Feature selection based on Fisher criterion and information gain was applied to find efficient features from given basis. Information Gain criterion shows improved performance with ICA, NMF and PCA feature when applied to Nearest Mean Classifier with Euclidean and cosine distance measure. Cosine distance measure performs better with PCA features in contrast to the better performance of Euclidean distance measure when applied to ICA features. When SVM is used as classifier, slight improvement is seen when features are selected based on information gain compared to the feature selection method based on fisher score.

Architecture 1 for image representation of ICA will be implemented and compared with the architecture 2 as proposed in [1]. Architecture 1 finds statistically independent basis images, in comparison to architecture 2 which finds factorial code representation consisted of independent coefficients.

Acknowledgment

This research was supported as the Brain Neuroinformatic Research Program by Korean Ministry of Commerce, Industry, and Energy.

References

[1] W M. S. Bartlett, J. R Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis", *IEEE Transactions on Neural Networks*, 13(6) pp. 1450-64, 2002.

[2] C. S. Dhir, N. Iqbal, and S.-Y. Lee, "Efficient feature selection based on information gain criterion for face recognition", Accepted for publication in the *proceedings of IEEE Int. conf. on Information Acquisition 2007*.

[3] IMDB, Intelligent Multimedia Laboratory, "Asian face image database PF01", *Technical report POSTECH*, Korea, 2001. (http://nova.postech.ac.kr/)

[4] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization", *Advances in Neural Information Processing*, vol. 13, MIT Press, 2001.
[5] A. Hyvarinen, J. Karhunen, and E. Oja, "Independent Component Analysis", John Wiley & sons, Inc., 2001,
[6] M. A. Turk and A.P Pentland. "Eigenfaces for Recognition", *Cognitive Neuroscience*, 3(1):71–86, 1991.
[7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, No. 7, pp. 711-720, July 1997.
[8] G. Wang, F. H. Lochovsky, and Q. yang, "Feature selection with conditional mutual information MaxMin in text categorization", *Proc. Int. conf. on information and knowledge management*, pp. 342-349, 2004.
[9] http://cvc.yale.edu/projects/yalefaces/yalefaces.html
[10] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", *IEEE Trans. Pattern analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.
[11] A. Hyvärinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, 10 (3): pp. 626-634, 1999.
[12] http://www.cis.hut.fi/projects/ica/fastica/
[13] B. A Draper et al., "Recognizing faces using PCA and ICA," *Computer Vision and Academic understanding, vol.* 91, pp. 115 – 137, 2003.
[14] http://svmlight.joachims.org/
[15] Nicu Sebe, Michael S. Lew, Ira Cohen, Ashutosh Garg, Thomas S. Huang, "Emotion Recognition Using a Cauchy Naive Bayes Classifier", *16th International Conference on Pattern Recognition (ICPR'02)*, vol. 1, pp. 100, 2002.

Paresh Chandra Barman has completed his M.Sc. degree from the University of Rajshahi-Bangladesh in 1995 and Ph.D. degree from KAIST, Republic of Korea in 2008. He is working as an Associate Professor and Chair of the the department of Information and Communication Engineering, Islamic University, Kushtia, Bangladesh.

Chandra Shekhar Dhir has completed his M.Sc. degree from KAIST, Republic of Korea in 2006 and currently he is pursuing his Ph.D.degree in the same university.

Soo-Young Lee is working as a Professor in KAIST, Republic of Korea.