International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:8, 2013

Normalization Discriminant Independent Component Analysis

Liew Yee Ping, Pang Ying Han, Lau Siong Hoe, Ooi Shih Yin, and Housam Khalifa Bashier Babiker

Abstract**—**In face recognition, feature extraction techniques attempts to search for appropriate representation of the data. However, when the feature dimension is larger than the samples size, it brings performance degradation. Hence, we propose a method called Normalization Discriminant Independent Component Analysis (NDICA). The input data will be regularized to obtain the most reliable features from the data and processed using Independent Component Analysis (ICA). The proposed method is evaluated on three face databases, Olivetti Research Ltd (ORL), Face Recognition Technology (FERET) and Face Recognition Grand Challenge (FRGC). NDICA showed it effectiveness compared with other unsupervised and supervised techniques.

Keywords**—**Face recognition, small sample size, regularization, independent component analysis.

I. Introduction

OVER the past decades, biometrics authentication technology has been widely developed for security purpose. Among these technologies, face recognition has gained considerable attention in recent years. There are numerous applications of face recognition such as in government use and commercial use, e.g. access control, surveillance purpose, banking etc. Compared with other biometrics technologies such as iris authentication, finger authentication and etc, there are some reasons for its extensive attention. Face recognition is an authentication system which does not require expensive and advance devices. Furthermore, it is contactless as the authentication process can be done without voluntary action from the users.

Normally, illumination, pose, lighting and facial expression are the variants that result system performance degradation. Many approaches have been applied on the face recognition for better improvement [1]-[5]. Generally, face analysis approaches could be classified into three categories such as holistic approach, feature-based approach and hybrid approach. Holistic approach makes use of the global information which derived from facial image pixels. The representative examples of holistic approach are Principal Component Analysis (PCA) by Kirby and Sirovich (1988) [6] and Linear Discriminant Analysis (LDA) which proposed by Peter N. Belhumeur et al. in 1997 [7]. Feature-based approach adopts local features of the face for data learning. These local features include eyes, nose, mouth, chin and head line. Elastic Bunch Graph Matching (EBGM) is one of the famous feature-based approaches which proposed by Wiskott, L et al. (1997) in face recognition [5]. Hybrid approach utilizes both holistic and feature-based approaches. The idea of this hybrid approach is that human facial shares the same basic features such as nose, mouth but challenging in distinguish the characteristics which are not present in other face component such as forehead and chin. Neither holistic nor feature-based approach could analyse these two features at a time. Hence, hybrid approach has the capability to perform this computation. One of the face recognition algorithm based on hybrid approach is proposed by Shiladitya Chowdhury et al. (2010) [8]. They were using generalized two-dimensional Fisher's Linear Discriminant method. The method adopts maximization of class separability from both the row and column directions simultaneously and yields much smaller image feature matrix.

PCA is one of the popular appearance-based techniques in the face recognition. PCA is also known as Eigenfaces. By using eigenspace decomposition of covariance matrix which derived from the training sample, PCA approximates linearly independent basis of face subspace. In PCA analysis, the whole face image is taken as a whole. This leads to the image variation from the same person (within-class scatter) could be larger than the image variation of face identity (between-class scatter). Usually, this within-class variation is induced by varying illumination, pose or facial expression of the individual. Based on this observation, PCA is enhanced through adopting discriminant criterion. This enhanced technique is known as LDA. LDA defines a projection which maximizes the between-class scatter and minimize within-class scatter. However, the decorrelation of input data is based on second-order statistics and ignoring higher order information. In face recognition, it is believed that there are some significant information contained in the higher-order relationship of image pixels. Therefore, ICA is proposed.

Independent Component Analysis (ICA) is another technique used in face recognition. ICA is a technique to separate independent component from a mixed source. It uses high order statistics as presented in [3]; ICA has two architectures, Architecture I and Architecture II. ICA Architecture I is an algorithm where the images are treated as variables and pixels are the observations. It focused on spatially localized features. For ICA Architecture II, it treats images as observation and pixels as variable. It is mainly focused on global features. There are several extensive researches on ICA face recognition [9]-[11].

This work is supported by the financial support of Telekom Research and Development Sdn Bhd. of Malaysia.

The authors are with the Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia. (email: lyping8@yahoo.com, yhpang@mmu.edu.my, lau.siong.hoe@mmu.edu.my, syooi@mmu.edu.my, me.the.fren@gmail.com).

In face recognition process, feature extraction is one of the main tasks for data learning. It attempts to search for the most appropriate representation of the data. However, face recognition often suffers from small number of available training samples as compared to dimensionality of the sample space. To deal with this situation, a regularization method is introduced. There are several research been proposed by adopting the regularization method such as [12], [13].

Inspired by these works, we propose a method called Normalization Discriminant Independent Component Analysis (NDICA). This technique forms a Laplacian matrix according to the information that obtained from the correlation coefficients between images. Then a weighting function based on the eigenvalues derived from the Laplacian matrix is defined and the input data will be regularized. The regularized input data will then been processed using independent component analysis. The experiment is tested on three face databases: Olivetti Research Ltd (ORL), Face Recognition Technology (FERET) and Face Recognition Grand Challenge (FRGC). The result shows its effectiveness.

II. Independent Component Analysis (ICA)

ICA is an unsupervised technique which separates the independent sources from a mixture. It aims to obtain basis vectors based on statistically independent variables. The general model of ICA is as in (1):

$$X = AS \tag{1}$$

where A represents unknown mixing matrix, S represents unknown source signal and X represents observed mixtures. In this case, it is assumed that the source signals are statistically independent and non-Gaussian and observed mixtures is the only information to have. In ideal condition, mixing matrix A can be inversed. Based on these assumptions, separation matrix, W, is also inverse of A. If the estimation of separation matrix is accurate, then a good approximation of source signal will be obtained.

$$U = WX = WAS \text{ and } W = A^{-1} \tag{2}$$

where U is the independent source signals.

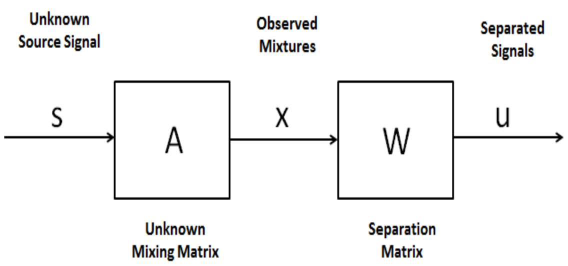

Fig. 1 Blind source separation model

There are numbers of algorithms to perform ICA. One of the famous algorithms that been studied is approached by Barlett et al. [3]. In the study, two different algorithms are presented for face recognition, ICA Architecture I and ICA Architecture II.

A. ICA Architecture I

In this approach, it is mainly for localized tasks whereby only small part of the image is taken part and this produced spatially localized features. It had been argued that this architecture performed better as it implements recognition by parts. Hence, ICA Architecture I is aims to find basis images that are statistically independent.

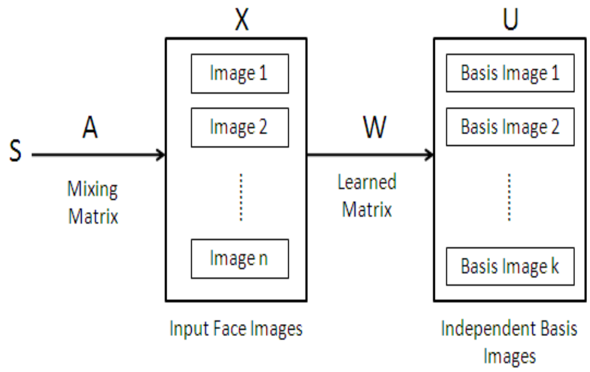

Fig. 2 Architecture I image synthesis model

B. ICA Architecture II

While ICA Architecture I aim for localized features, ICA Architecture II is aims to produce global features. This approach is to find a coefficient for the input data which is statistically independent. Hence, ICA Architecture II used Factorial Face Code to accomplish the goal.

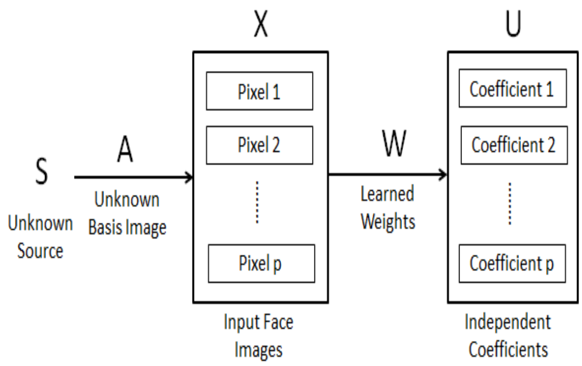

Fig. 3 Architecture II image synthesis model

In Architecture II, the pixels are treated as variables and images are observations. This results the column of $A = W^{-1}$ as a set of basis image. Column of U contains a set of independent coefficient of basis images in A for reconstructing image in X. Therefore, U is a factorial code representation.

III. Normalization Discriminant ICA (NDICA)

In real world, the facial images are derived from high dimensional and exposed to different kinds of variations such as light, illumination etc. In order to have an effective face recognition, dimensional reduction is applied to the high dimensional facial images. For better performance, one technique should be produced low variation of the same facial within-class scatter and large variation between-class scatter.

In this paper, we propose a new technique called Normalization Discriminant ICA (NDICA) to obtain most appropriate and reliable information from the data. As mentioned before, a low variation within-class scatter and large variation between-class scatter lead to good performance in recognition.

We learn the information from correlation coefficient between the images to seek data relation. These coefficients are the weight on each edge of the adjacent data pairs. In our case, we use binary to denote the weights.

W = $\{W_{jk}\}$ is weight matrix

$$W_{jk} = \begin{cases} 1 & \text{if } x_j \text{ and } x_k \text{ are belonged to the same class} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and binary Laplacian, $\mathbf{L}_{bin}$ is denoted as

$$\mathbf{L}_{bin} = \mathrm{D} - \mathrm{W} \tag{4}$$

where D = $\{\sum W_{jk}\}$. Based on the eigenvalues, v of the Laplacian scatter matrix $\mathbf{XL}_{bin}\mathbf{X}^{\mathrm{T}}$, $\mathbf{L}_{bin}$ is being used to evaluate the localized features of the data.

$$\mathbf{XL}_{bin}\mathbf{X}^{\mathrm{T}} v_i = \varphi_i v_i \tag{5}$$

in which φ_i are the eigenvalues of the corresponding eigenvector, v_i. The eigenvectors v_i forms the eigenspace, $\mathrm{V} = [v_{1,} v_{2,} \ldots v_n]$.

Fig. 4 illustrates the descending order of the plot of φ_i correspond to eigenvectors, v_i . It shows that the within-class variability is maximized when the value of φ_i that corresponding to the eigenvectors, v_i is optimized. Hence, the eigenvectors need to be regularized with the smaller weights to minimize the variation within-class. The smaller value of φ_i indicates the minimal of within-class variation.

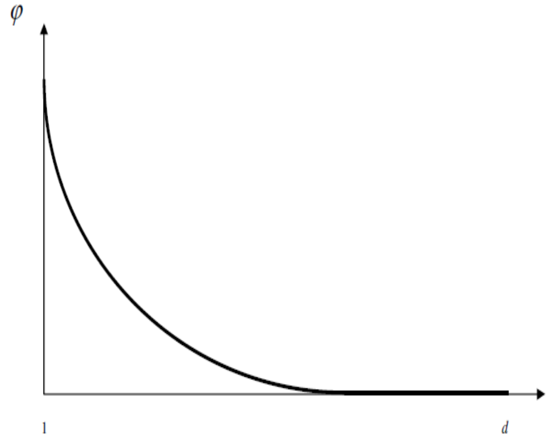

Fig. 4 Plot of φ_i

The space V is decomposed into two subspaces, *outlander subspace* which represent low locality preserving capability and *dominant subspace* which represent high locality preserving capability. Fig. 5 illustrates the decomposition of space V.

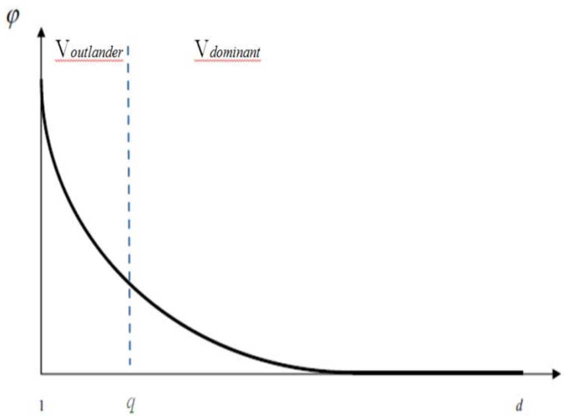

Fig. 5 Decomposition of space V

Upper quartile (Q3) and interquartile range (IQR) is applied to separate the outlanders from the data set. The outlander fence is defined as

$$\varphi_{outlander\ fence} = \alpha(Q3 + 1.5 * IQR) \tag{6}$$

where Q3 indicates the lowest 75% of the total of φ and α is scaling parameter. The end point of outlander space, q is defined as

$$\varphi_q \approx \varphi_{outlander\ fence} \tag{7}$$

Based on the locality preserving regulation model above, we defined a weight function as in (8):

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:8, 2013

$$\Omega_i = \left\{ \begin{array}{ll} 1/\sqrt{\varphi_i} & , 1 \le i \le q \\ 1/\sqrt{\varphi_q} & , q+1 \le i \le d \end{array} \right\} \quad (8)$$

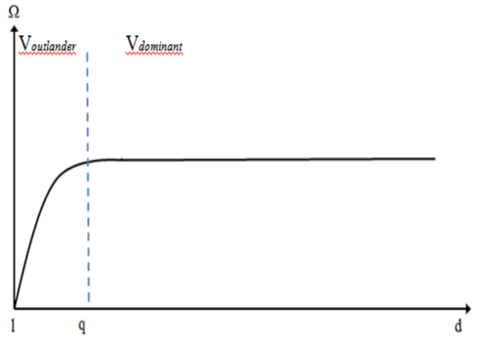

Fig. 6 Weight function of Ω_i

Fig. 6 shows the weight scaling, Ω_i increases proportionally until q and remains constant at $V_{dominant}$. Then the eigenvectors, v_i is regularized by dignified eigenvalue-based weight:

$$\widetilde{V} = [\Omega_i v_i]_{i=1}^{d} \quad (9)$$

By using the weight function and eigenvectors, the input data is regularized as in (10):

$$\widetilde{X} = \widetilde{V}^T X \quad (10)$$

where X are the input data. This regularization process produces more localized features which are better for dimension reduction.

A. NDICA I

In NDICA, with more localized features that produced previously, we adopt the features into ICA Architecture I and Architecture II. Instead of using the original input face images, X, we use $\widetilde{X}$, the input images after regularized.

In NDICA I, the algorithm is carried out where *R* be a *p* by *m* matrix where *m* is the eigenvectors of a set of n face images and *p* is the number of pixels in training sample. The input images in the row are treated as variables and the pixels in the column are observations, ICA Architecture I is performed on R^T. It aims to seek for statistically independent basis images. Therefore, regularized data, $\widetilde{X}$ are the variables and pixels are the observation.

A set of independent basis image in the rows of U is computed as

$$U = W * R^T \quad (11)$$

Let C be the PCA coefficient,

$$\begin{aligned} C &= \widetilde{X}^T * R \\ &= (\widetilde{V}^T * X)^T * R \end{aligned} \quad (12)$$

Therefore, ICA coefficient matrix is computed as

$$\begin{aligned} B &= C * W^{-1} \\ &= (\widetilde{V}^T * X)^T * R * W^{-1} \end{aligned} \quad (13)$$

B. NDICA II

In NDICA II, it is different from Architecture I as it treated pixels as variables and $\widetilde{X}$ as observations. Hence, the statistically independent coefficient can be formulated as

$$U = W * C^T \quad (14)$$

IV. Experimental Result

Three face databases, namely Olivetti Research Ltd (ORL), Face Recognition Technology (FERET) and Face Recognition Grand Challenge (FRGC) databases are adopted to evaluate the performance of the proposed technique. ORL database contains 40 individual and each of them has 10 images which vary in different environment such as lighting, facial expression or even taken under dark homogenous background. FERET database is a large database which contained images of 1199 individuals. It has seven categories of different face pose. In this work, we adopt a subset of FERET database that comprises 100 subjects with 10 images per subject. For FRGC database, it is capable for testing under uncontrolled environment. In this work, we adopt a subset of these databases. 100 individuals with 10 images of each are randomly selected from FERET database; while 150 individuals with 10 images of each are chosen from FRGC database. From each database, 5 images of each individual are used for training and the remaining 5 images of each are used for testing.

This experiment is to seek the most optimal parameter of NDICA I and II. As observed in (6), α is a free parameter that to be determined. Hence, ORL database is used to determine the best value of α. Table I shows the average error rate of NDICA I and NDICA II on the ORL database with dimension, t = 10 against different α values. The result shows that NDICA I and NDICA II have the minimal average error rate when α = 5.

TABLE I
Average Error Rate of NDICA I and NDICA II with Different Alpha Value on ORL Database

Alpha α	NDICA I	NDICA II
1	7.377	37.890
2	6.040	26.990
3	6.189	17.164
4	5.094	38.167
5	4.927	6.246

Therefore, α = 5 will be adopted on the subsequent experiment on NDICA I and NDICA II.

TABLE II
PERFORMANCE IN TERMS OF ERROR RATE ON ORL DATABASE

Method	Feature Dimension	Error Rate
PCA	20	11.76
LDA	39	6.96
ICA I	20	13.99
ICA II	7	14.4
NDICA I	20	4.73
NDICA II	8	5.5

TABLE III
PERFORMANCE IN TERMS OF ERROR RATE ON FERET DATABASE

Method	Feature Dimension	Error Rate
PCA	80	40
LDA	99	39
ICA I	50	39.2
ICA II	30	38.2
NDICA I	9	28.1
NDICA II	10	28.6

TABLE IV
PERFORMANCE IN TERMS OF ERROR RATE ON FRGC DATABASE

Method	Feature Dimension	Error Rate
PCA	110	49.44
LDA	149	40
ICA I	130	37.7
ICA II	7	36.3
NDICA I	9	23.5
NDICA II	10	25.2

From the result, we can observe that our algorithm is outperformed than other algorithms. The regularization process has improved the performance of the ordinary ICA. This proved that regularization is able to minimize within-class variation and the result is promising in the feature extraction.

For class discriminant function, LDA employs both within-class and between-class information for feature extraction. The main criterion in NDICA is regularized the input data to obtain the minimal within-class variation before utilize for feature extraction and thus showed that regularization feature does brings its effectiveness.

From the result, it is observed that unsupervised and supervised techniques such as PCA and LDA only able to identify the feature at high projection, but NDICA could determine the features at very low dimension.

V. CONCLUSION

In this study, we have explored feature extraction technique on ICA based on regularization technique which called Normalization Discriminant ICA (NDICA). Experiment had been carried out on the images which vary in face pose, illumination, facial expression etc. It shows promising result over PCA, LDA, ICA I and ICA II on ORL, FERET and FRGC databases. This proved that the proposed algorithm performed high locality preserving capability on small number of extracted features.

REFERENCES

[1] M. A. Turk and A. P. Pentland, “Face Recognition Using Eigenfaces”, *IEEE Conf. on Computer Visionand Pattern Recognition*, pp. 586-591, 1991

[2] Chelali, F.Z.; Djeradi, A.; Djeradi, R.,"Linear discriminant analysis for face recognition," *Multimedia Computing and Systems, 2009. ICMCS '09. International Conference on* , vol., no., pp.1,10, 2-4 April 2009

[3] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face recognition," presented at *SPIE Symposium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III*, San Jose, CA, 1998.

[4] Ying-Han Pang, Andrew Teoh Beng Jin, David Ngo Chek Ling, “Face authentication system using pseudo Zernike moments on wavelet subband”. *Proc. IEICE Electr. Exp.* v1 i10. 275-280

[5] Wiskott, L.; Fellous, J.-M.; Kruger, Norbert; Von der Malsburg, C., "Face recognition by elastic bunch graph matching," *Image Processing, 1997. Proceedings., International Conference on* , vol.1, no., pp.129,132 vol.1, 26-29 Oct 1997

[6] I. Sirovich and M. Kirby, “Low- dimensional procedure for the caracterization of human faces,”*Journal of Optical Society of America A*, vol. 4, no. 3, pp. 519–524, March 1987.

[7] Peter N. Belhumeur, J. P. Hespanha, David J. Kriegman, Eigenfaces vs. Fishefaces: Recognition Using Class Specific Linear Projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.19 n.7, p.711-720, July 1997.

[8] Shiladitya Chowdhury, Jamuna Kanta Sing, Dipak Kumar Basu, Mita Nasipuri, Face recognition by generalized two dimensional FLD method and multi-class support vector machines, Appl. Soft Comput. J. (2011)

[9] Keun-Chang Kwak; Pedrycz, W., "Face Recognition Using an Enhanced Independent Component Analysis Approach," *Neural Networks, IEEE Transactions on* , vol.18, no.2, pp.530,541, March 2007

[10] Dagher, I.; Nachar, R., "Face recognition using IPCA-ICA algorithm," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.28, no.6, pp.996,1000, June 2006

[11] O. Deniz, M. Castrillon, M. Hernandez, “Face recognition using independent component analysis and support vector machines”, *Pattern Recognition Letters*, v.24 n.13, p.2153 – 2157, September 2003

[12] Dao-Qing Dai; Yuen, P.C., "Face Recognition by Regularized Discriminant Analysis," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* , vol.37, no.4, pp.1080,1085, Aug. 2007

[13] J. W. Lu , K. N. Plataniotis and A. N. Venetsanopoulos "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition", *Pattern Recognit. Lett.*, vol. 26, no. 2, pp.181 -191 2005.