World Applied Sciences Journal 34 (6): 688-694, 2016
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2016.34.6.422

Face Recognition Based on Modified Discriminant Independent Component Analysis

[1]*Maryam Mollaee and* [2]*Mohammad Hossein Moattar*

[1]Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran
[2]Department of Software Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran

Abstract: This paper proposes a novel feature extraction method for face recognition problem. There exist several successful methods for face recognition. The aim of recognition is to approximately estimate the components from the raw image. Components plays an important role in face recognition systems. Consequently, these components are used for extraction of face image features. However, these features may not be appropriate for classification, since the ICA method does not consider the class information. For the purpose of optimizing the performance of ICA, the discriminant ICA (dICA) method, which is a combination of ICA and LDA methods, is utilized for face recognition in this study. Improved performance of our method comparing with common methods like LDA and ICA is shown using the Yale data. We have also proposed particle swarm optimization (PSO)-dICA method to improve the dICA performance, in which PSO is used instead of the gradient approach for learning dICA. The results of PSO-dICA method confirm our idea in classification experiments compared to other methods. Using proposed method on Yale B dataset, gives an average classification accuracy of 92.169% compared with an accuracy of 91.322% using when dICA and accuracy of 89.77% compared with ICA and accuracy of 86.18% using PCA and also accuracy of 84.76% using LDA.

Key words: Discriminant independent component analysis · Feature extraction · Particle Swarm Optimization · Linear discriminant analysis

INTRODUCTION

Over the past twenty years, face recognition has become a subject of scientific research. Face recognition is a combination of computer science, engineering and statistic techniques [1]. Dimensionality reduction has been developed to improve speed and accuracy of face recognition systems. In fact, this methods reduce the number of dimensions of original image using mapping into a subspace with lower dimensional. This is essential for face recognition because only face pattern is necessary for identification, while the original image is high dimensional with background and unneeded texture. Features that are transformed into such a subspace have more prominent and useful information than original image and make identification easier and more effective [2]. In [3], authors proposed the application of linear discriminant analysis (LDA) for face recognition. LDA method maximizes the intra-class scatter to inter-class scatter ratio for finding a linear mapping. However, LDA has the small sample size problem.

Principal component analysis (PCA) has been used successfully to extract facial features from images. Kirby *et al*. [4] represented human faces as a linear combination of weighted eigenvectors using PCA. Also Yang *et al.* [5] proposed two-dimensional PCA (2DPCA) [6]. However, PCA has poor discriminatory power and high computational. To eliminate these shortcomings, Bartlett *et al*. [7] proposed the use of independent component analysis (ICA) for face recognition which is a generalization of PCA. PCA considers only second-order statistical dependence and makes data uncorrelated while ICA also takes higher-order statistics into account and makes data independent as possible. In face recognition, all methods aim to estimate the components of raw images. These components are then used to extract the features of face images which are used for face classification. Xuemei Wang [8] and J. Yang *et al.* [9] applied kernel ICA for face recognition. Yanhong Zhou *et al.* [10] proposed a method based on the combination of principal component analysis (PCA) and independent component analysis (ICA) for face

Corresponding Author: Maryam Mollaee, Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

recognition. Fei *et al*. [11] proposed a two-stage method for face recognition where firstly, facial data is projected by independent features and then LDA is performed. But projection of independent features using LDA to a discriminant feature space may cause that the features not to be independent.

In order to overcome these serious shortcomings a linear, single-stage semi-supervised discriminant independent component analysis proposed by Lee, *et al.* [12, 13] in which Fisher linear discrimination and Negentropy are simultaneously maximized and the latent independent factors are extracted from multivariate data. This approached is called Discriminant ICA (dICA). dICA method is the new schema of projection that combines the power of supervised and unsupervised learning.

dICA can be more effective compared with the classifications developed by ICA, PCA, LDA methods, the dICA classification enjoys lower computational complexity, a better representation and fewer data reconstruction errors [13]. dICA is a semi-supervised method to transform multivariate data into a feature space with low dimension in which the features are statistically independent and try to capture most discriminative directions [12, 13].

However because of using the Gradient method in dICA, it is prone to falling into local optima. The Discriminant Independent Component Analysis based on PSO is proposed to overcome this shortcoming [14]. Yale B dataset is used for evaluation in this study. Face samples are viewed as a set of components and independent and distinct faces are extracted using dICA method.

The remainder of this paper is organized as follows: Next section, presents the fundamental of dICA and PSO optimization algorithm. In Section 3 the proposed method is introduced. In Section 4, the experimental results are presented and Section 5 summarizes the conclusions.

Basic Concept

Particle Swarm Optimization: In recent years, heuristic optimization methods are utilized for achieving results with better performance compared with random methods in optimization problems. These methods develop absolutely competitive solutions when appropriately performed. They also have very good experimental convergence features comparing with conventional searching algorithms.

One of these methods is PSO method in which a group of random particles are initialized and an optimum solution is searched for in several iterations .During each iteration level every particle is update by the best local and the best general. This method needs few initial parameters and also, it is easily controlled. X_i is the position of i^{th} particle in the searching space that is considered as a solution. The cost value is calculated using cost function and X value. The best position calculated by itself, p_{besti} and the best position calculated with total population, g_{besti}, which are calculated as follows [15-17].

$$p_{besti}(t+1) = \begin{cases} p_{besti}(t) if \quad f(X_i(t+1)) < \mathrm{f}(p_{besti}(t)) \\ X_i(t+1) if \quad f(X_i(t+1)) \geq \mathrm{f}(p_{besti}(t)) \end{cases} \tag{1}$$

$$g_{besti}(t) = \max\left\{ f(p_{best1}(t), f(p_{best2}(t) \ldots . f(p_{bestM}(t)) \right\} \tag{2}$$

Position and speed of each particle in the next iterations are calculated as follows:

$$x_i(t+1) = x_i(t) + v_i(t+1), \mathrm{v}_{ij}(t) \in \left[-v_{\max}, v_{\max} \right] \tag{3}$$

$$v_{i,j}(t+1) = w_i^t v_i(t) + c_1 r_1(t)(pbest_i(t) - x_i(t)) + c_2 r_2(t)(pbest_i(t) - x_{i,j}(t)) \tag{4}$$

Independent Component Analysis: ICA is a high dimension statistical analysis technique for increasing the possibility of discovering hidden agents, which are placed in the set of stochastic variables as well as signals. To describe the method, assume that N statistic independent sources exist, as shown by vector S. also, x observations are different combinations of N, which are written compactly.

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ . \\ . \\ x_n(t) \end{pmatrix} = A \begin{pmatrix} s_1(t) \\ s_2(t) \\ . \\ . \\ s_n(t) \end{pmatrix} \tag{5}$$

A represents the unknown matrix. For simplicity, unmixing model can be as follows [18]:

$$Y = W * X \tag{6}$$

The aim of ICA method is to find a separator matrix called w so as the components of y=wx have maximum possible statistic independency. Generally, w optimization is based on some kind of independency criterion [19, 20]. In this study, the maximum of negentropy is used.

Bartlett and Liu [7] were the first to utilize ICA for face recognition and found out that it is better than PCA when cosines distance is used as similarity criterion. Face samples were observed as a set of combinations and independent faces were extracted using ICA method. ICA aims to find a linear transformation for input data to make them as independent as possible, while PCA aims to find uncorrelated variables from input data. Therefore, ICA is considered as an extension of PCA method and it combines the learning ability with an observer and without it.

A set takes N learning face image every one of which is a column vector consisting of m pixels of a face image. The aim of demonstrating the sub-space based on ICA is to extract the features that are the most different and the feature extraction takes place by mapping the test image to this sub-space. Every face image is considered as an observed variable and it is assumed as a combination of hidden independent components.

Discriminant Independent Component Analysis: In dICA method, multivariate data with lower dimension and independent features can be obtained by maximizing Negentropy. In dICA, the Fisher criterion and the sum of the marginal Negentropy of the extracted independent features are maximized simultaneously. Therefore, dICA combines representation model with discriminant model in order to develop a better classification [12, 13].

Extraction of Independent Features by Negentropy Maximization: Negentropy is a good statistical estimate of non-Gaussian random variables. An approximation of the marginal Negentropy can be as following:

$$J(\mathrm{y}_i) \approx k_1(E(\mathrm{G}^1(\mathrm{y}_i)))^2 + k_2(E(\mathrm{G}^1(\mathrm{y}_i)) - E(\mathrm{G}^2(\gamma_i)))^2 \quad (7)$$

where G1 and G2 denote non-quadratic odd and even functions. The followings are some of the usual choices for a random vector with a symmetrical (normal) distribution:

$$G^1(y_i) = \mathrm{y}_i{}^3 \quad (8)$$

$$G^2(\mathrm{y}_i) = \frac{1}{a_1}\log\cosh a_1\,\mathrm{y}_i, 0 < a_1 \le 1 \quad (9)$$

$$G^2(\mathrm{y}_i) = -\exp(\frac{\mathrm{y}_i{}^2}{2}) \quad (10)$$

Maximization of the sum of the marginal Negentropy with the unit covariance can be achieved through a Lagrange formula is defined as.

$$\tilde{L}(W) = \sum_{i=1}^{k}[E(G(\mathrm{w}_i^T z)) - E(G(\upsilon))]^2 + \sum_{i=1}^{R}\beta_i(\mathrm{w}_i^T w_i - 1) \quad (11)$$

Features can be obtained by maximizing the target function in Equation 11. Optimization problem that maximizes the functional criterion of classification performance and Negentropy of the extracted independent features simultaneously can be defined through a Lagrange formula in the following form:

$$\hat{L} = \tilde{L}(W) + k\phi(W, \mathrm{Z}, \mathrm{C}) \quad (12)$$

where K is a constant and ϕ denotes the function measuring the efficiency of the classification of the features of Y with the given C and $\tilde{L}(W)$ the same as Equation 11. Learning rule that is formed as:

$$\Delta w_i = \eta(\gamma_i(E(Z_g(w_i^T Z)) + k\frac{\partial\phi(W,Z,C)}{\partial w_i} + 2\beta_i w_i) \quad (13)$$

$$\beta_i = -\frac{1}{2}\gamma_i E(y_i g(y_i)) \quad (14)$$

$$\gamma_i = 2\left(-\frac{\sum_{n=1}^{N}\exp\left(-\frac{y_{in}^2}{2}\right)}{N} + \frac{1}{2}\right) \quad (15)$$

Symmetric orthogonalization of the matrix is done as follows:

$$W \leftarrow (WW^T)^{-1/2}W \quad (16)$$

Functional Measure of Classification Performance: The function measuring the classification performance is defined as:

$$\phi(W,Z,C) = \sum_{i=1}^{R}\log\frac{w_i^T S_B w_i}{w_i^T S_W w_i} = \sum_{i=1}^{R}\log\frac{\sum_{c=1}^{C}N_c(\mu_{ic} - \mu_i)^2}{\sum_{c=1}^{C}N_c\sigma_{ic}^2} \quad (17)$$

$$\mu_{ic} = \frac{1}{N_c}\sum_{n\in Class\ c} y_{in}\ , \sigma_{ic} = \frac{1}{N_c}\sum_{n\in Class\ c}(y_{in} - \mu_{ic})^2 \quad (18)$$

In order to maximize Equation 17, the gradient of 17 is derived with respect to W, as [12, 13]:

dICA Algorithm to Obtain Features [13]

Step 1: Data X is centered as. $X \leftarrow X - E(X), X \in R^p$
Step 2: To obtain orthonormal features the centered data is Whitened, $Z \in R^L$.
Step 3: Initialize W as: $W=W_0 \,\|W_0\|=I_R$.
Step 4: dICA features are obtained from $Y = W^T Z, Y \in R^R$.
Step 5: W is Updated according to Equation 13.
Step 6: Symmetric orthogonalization of W is done as: $(WW^T)^{-\frac{1}{2}} W$
Step 7. If the sum of Negentropy of y_i and discriminant function $\phi(W, Z, C)$ are not converged, go back to step 4 ; else stop iteration.

Proposed Method: In this paper, we present a framework on face recognition with the aim of classification. The framework has two main stages: a feature extraction stage and a classification stage. In the first stage transform the original image data into a less dimensional space. In the second stage support vector machines (SVM) [21] is applied for classification. In this study, PSO method is chosen for dICA search algorithm, since it has two benefits: efficient calculations and easy implementation. In this method, a fitness function is considered, then iterations are performed in the search space and the position and the velocity are updated in each level. Finally, optimized solution is acquired when best convergence is reached.

In the proposed approach, Equation 12 is considered as the PSO fitness function which maximizes Negentropy and the Fisher index at the same time. In fact, the optimization problem is to find the values of particle so as to maximize the fitness function. The position of particles gives us the separating matrix W. This step is simpler than the calculation of the derivation in the Gradient method.

PSO-dICA Algorithm:

Step 1. Center the data X: $X \leftarrow X - E(X), X \in R^p$. Whiten the centered data to obtain orthonormal features, $Z \in R^L$
Step 2: Initialize the particles (position and velocity of particle) randomly.
Step 3: Evaluate the fitness of each the particle according to Equation 12.
Step 4: Calculate local best position according to Equation 1.
Step 5: Calculate the global best position according to Equation 2.
Step 6: Update the velocity and position according to Equations 3 and 4.
Step 7: If maximum iteration number is not satisfied, go back to Step3; else W is obtained.
Step 8: Obtain features according to Equation 6.

Experimental Results: In this section we evaluate the performance of the proposed method on Yale B face database and performance is compared with the performance of other methods in order to prove its effectiveness. Recognition performance is evaluated by SVM classifier.

Table 1 shows The PSO parameters. In PSO, an n×n matrix (W) is the unmixing matrix of ICA method. Each one of its columns is equivalent to w_i vectors. w_i vectors are superposed with different independent components and are orthogonal to each other. To prevent convergence of different vectors to equal maximum, w_i vectors are orthogonalized after every iteration according to equation 16.

The number of the features which are extracted is equal to the number of ws i.e. n. According to symmetrical or orthogonalization method, values of w_i are estimated altogether instead of being estimated one by one. By improving particle position, the point for which fitness function has the least cost is obtained and this value is the solution of the problem i.e. W matrix.

Data Set: A subset of Yale B data set [2] is used to demonstrate the effectiveness of PSO-ICA method. It contains 15 persons with 30 different images for everyone. Images are captured with different light resource direction in different times and face states and details .In this research images have changed in size to 32×32 pixel for better performance. Training and test data include 315 and 135 samples, respectively.

Figure 1 shows sample images of a subject from Yale B database. All Images are resized to 32×32 and converted from RGB to grayscale images. From this database, 70% of images randomly chosen from a person as training set and the remaining images (30%) for testing set.

Evaluation Measures: The most popular Measure is used to evaluate classification is accuracy of classification as in Equation 19.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{19}$$

Table 1: Parameters used in the experiments

PSO parameter		Gradient
Iteration	160	2000
Max	0.5	
Min	-0.5	

Table 2: Experimental results on Yale Face B database using SVM classifier with RBF kernel

Feature Extraction Method	Number of Features	PCA	ICA	dICA	PSOdICA
Accuracy (%)	10	71.429	84.444	85.926	87.407
	20	86.666	93.333	93.333	93.333
	30	87.619	87.40	94.074	92.592
	40	89.524	88	88.889	91.111
	50	90.476	93.333	93.333	94.074
	60	91.42	91.555	91.111	93.333
Average accuracy	70	91.42	90.370	92.592	93.333

Table 3: Average experimental results on Yale Face B database using SVM classifier and RBF kernel

Feature Extraction Method	PCA	LDA	ICA	dICA	PSO-dICA
Accuracy (%)	86.18	84.762	89.776	91.322	92.169

Table 4: Elapsed time for feature extraction in seconds

Feature Extraction Method	Number of Features	dICA	PSO-dICA
Elapsed Time for each image (s)	60	4.17492	1.13618
	70	5.42400	1.41831

Fig. 1: Sample images of a subject from Yale B database

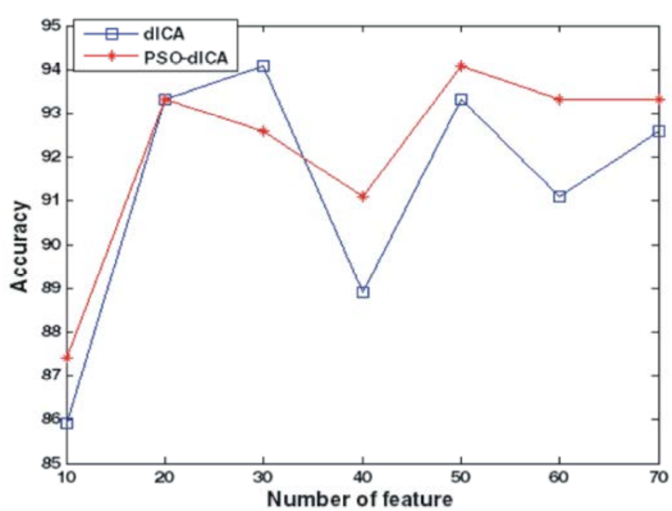

Fig. 2: Comparing the accuracy of PSO-dICA vs dICA using different number of features

Comparative Results: The results of the classification SVM are depicted in Table 2. It can be observed that PSO-dICA method improved the classification performance compared to other methods and dICA has higher accuracy compared with ICA, PCA and LDA.

Figure 2 illustrated the comparative results for the two methods (PSO-dICA & dICA) based on accuracy obtained. Each plot corresponds to a particular number of features. As shown, in most cases the PSO-dICA accuracy is higher than dICA. In general, increasing the number of features extracted led an improvement in accuracy.

The average accuracy on Yale data set for SVM is reported in Table 3,which presents the accuracy average of PSO-dICA versus the other methods. The proposed method has average accuracy of 91.70 % while the average accuracy of dICA is 91.11%. The results for the PCA and LDA methods are also included.

The average accuracy of classifiers are computed for 6 runs of feature extraction methods with various number of feature (i.e 10, 20, 30, 40, 50, 60) and c-1 (c is the number of classes i.e 15) for LDA.

In PSO-dICA, using PSO for learning dICA lead to more accurate and faster convergence than other existing method. PSO can quickly and effectively achieve optimal resolution to the dICA and avoids the local optima condition. Also PSO-dICA tries to capture most discriminative directions. In Table 4, it could be observed that the proposed PSO-dICA method is faster than dICA method in feature extraction phase.

CONCLUSION

In this paper, dICA classifier is optimized using PSO method for improved face recognition accuracy. Also, a two level approach is utilized, in which feature extraction is performed first by dICA which makes the input data as independent as well as discriminant as possible and then the dICA output is classified. For the purpose of evaluating dICA method, Yale B database is used. Comparing the results by the presented method in this paper with conventional methods such as LDA proves the effectiveness of our method in face recognition. Also, the accuracy is increased significantly in PSO-dICA method comparing with LDA method. The proposed method decreases the dimension while increasing the performance of the classifier. Issues like this produces some challenges in using this method. In the future, we will study 3D face images in our recognition system.

REFERENCES

1. Givens, G.H., J.R. Beveridge, P.J. Phillips, B. Draper, Y.M. Lui and D. Bolme, 2013. Introduction to face recognition and evaluation of algorithm performance. Computat Statist Data Anal, 67: 236-247.
2. Lei, J. and C. Lu, 2011. Enhancement of components in ICA for face recognition. In: 9th International Conference on Software Engineering Research, Management and Applications; 10-12 August 2011; Baltimore, MD, USA: IEEE, pp: 33-38.
3. Lu, J., K.N. Plataniotis and A.N. Venetsanopoulos, 2003. Face recognition using LDA-based algorithms. IEEE Trans Neural Netw, 14: 195-200.
4. Kirby, M. and L. Sirovich, 1990. Application of the Karhunen–Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal, 12: 103-108.
5. Ren, H. and H. Ji, 2014. Nonparametric subspace analysis fused to 2DPCA for face recognition, Optik - Int J. Light Elect Opt., 125: 1922-1925.
6. Mohammed, A.A., R. Minhas, Q.M. Jonathan Wu and M.A. Sid-Ahmed, 2011. Human face recognition based on multidimensional PCA and extreme learning machine. Pattern Recogn, 44: 2588-2597.
7. Bartlett, M.S., J.R. Movellan and T.J. Sejnowski, 2002. Face recognition by independent component analysis. IEEE Trans Neural Netw, 13: 1450-1464.
8. Wang, X., 2012. Supervised manifold learning and kernel independent component analysis applied to the face image recognition. In: Fifth International Conference on Intelligent Computation Technology and Automation; 12-14 January 2012; Zhangjiajie, Hunan, China, pp: 600-603.
9. Yang, J., X. Gao, D. Zhang and J. Yang, 2005. Kernel ICA: an alternative formulation and its application to face recognition. Pattern Recog, 38: 1784-1787.
10. Yanhong, Z., C. Shukai, W. Dong, Z. Huiyang and Z. Liqiang, The study of face recognition based on hybrid principal components analysis and independent component analysis. In: the Electronics, Communications and Control (ICECC); 9-11 September 2011; Ningbo World Hotel Ningbo, China. pp: 2964-2966.
11. Long, F., J. He, X. Ye, Z. Zhuang and B. Li, 2006. Discriminant independent component analysis as a subspace representation. J. Electron (China), 23: 103-106.
12. Dhir, C.S. and S.Y. Lee, 2011. Discriminant independent component analysis. IEEE Trans Neural Netw, 22(6): 845-857.

13. Dhir, C.S., J. Lee and S.Y. Lee, 2011. Extraction of independent discriminant features for data with asymmetric distribution. Knowl. Inf. Syst., Dol 10. 1007/s 10115-011-0381-9.
14. Mollaee, M., M.H. Moattar and S.J. Seyyed Mahdavi, 2014. Feature extraction based on modified discriminant independent component analysis via particle swarm optimization. Int J. Softw Engin Appl., 8: 91-102.
15. Kennedy, J. and R.C. Eberhart, 1995. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Network IV; 27 November 1995; Perth, WA., pp: 1942-1948.
16. Kennedy, J and R.C. Eberhart, 1997. A discrete binary version of the particle swarm algorithm. In: Proceedings of the 1997 Conference on System, Man and Cybernetics, IEEE Service Center, Piscataway, NJ, pp: 4104-4109.
17. Zhang, Y. and Y. Zhang, 2010. Fault detection of non-Gaussian processes based on modified independent component analysis. Chem. Eng. Sci., 65: 4630-4639.
18. Hyvärinen, A., J. Karhunen and E. Oja, 2001. Independent component analysis. New York: John Willy.
19. Hyvarinen, A., 1992. Survey on independent component analysis. Neural Comput Surv., 2: 94-128.
20. Hyvarinen, A. and E. Oja, 2000. Independent component analysis: algorithms and applications. Neural Netw, 13: 411-430.
21. Burges, J.C., 1999. A tutorial on support vector machines for pattern recognition. Boston: Kluwer Academic Publishers.