Revised R-LDA based ANN for Small Sample Size (SSS) Problem of Face Recognition

Lee Hui Kueh[1.1], Sang-Hyun Byun[2] and John-Tark Lee[1.2]

[1]*Electrical Department, Dong-A University, 840 Hadan2-dong, Saha-Gu, Busan 604-714, Korea.*
{Lee Hui Kueh, Sang-Hyun Byun, John-Tark Lee} leehkueh@naver.com

Abstract

A face recognition (FR) system is automatically identifying or verifying a personal face acquired from a digital camera or a image generation device. In order to do this, facial features from the acquired image should be extracted and compared with a facial database. All FRs face an obstacle related to the viewing angle of the face including poor lighting and low resolution. Because of those problems, its recognition rate substantially decreases. In this paper, a newly weighted regularization parameter based FR system which can improve recognition rate under certain environmental constraints is proposed. This approach is based on the conventional regularized linear discriminant analysis (R-LDA) and includes Artificial Neural Network (ANN) which can improve face recognition rate with a prominent classification ability. The revised R-LDA algorithm is attempted to address the Small Sample Size (SSS) problem that encountered in all FRs and the ANN is useful to detect the frontal views of faces. This algorithm has been tested over 350 images (35 classes) of Olivetti Research Lab (ORL) database using MATLAB. Its test results give us recognition rates of above 95%. In addition, it is also tested on the mirror and combination of the ORL database and the recognition performances are shown that the system is fairly robust and has the performance of more than 90%.

Keywords: *Face Recognition, Regularized Linear Discriminant Analysis, Small Sample Size, Artificial Neural Network.*

1. Introduction

A FR system is a process of automatically identifying or verifying a personal face in a system. FR is not limited on the fields of security systems, card verification and identification in bio mimetic systems; its demand is increasing due to the increased penetration of technologies, such as smart phone, the internet and some products of security oriented application. However, FR is not perfect and has one obstacle related to the viewing angle of the face including poor lighting and low resolution. Because of those problems, its recognition rate substantially decreases.

We proposed a RF system using a newly weighted regularization parameter based on the conventional R-LDA to address the SSS problem that encountered in all FRs and ANN to detect the frontal views of faces in order to improve the face recognition rate with prominent classification ability.

The R-LDA is a holistic approach [1], which considers the information close to null space of the within-class scatter matrices is more significant for discriminant tasks and takes the eigenvectors of the null spaces as the most significant feature bases. It uses

---

[1.2] Corresponding author jtlee@dau.ac.kr

the regulation parameter to control the trade-off between variance and bias that encounter in the eigenvalues estimates of the within-class scatter matrices due to the insufficient of training samples that all FRs faces.

ANN is a powerful network that has been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision and control systems. Here, the ANN is used for classification [2] of face images and their detections.

The remaining of the paper is organized as follows. Section 2 describes the R-LDA algorithm and Section 3 introduces the ANN. Section 4 shows the experiments and results and finally Section 5 presents the conclusions and final remarks.

2. R-LDA

Given a set of Z training face images$\{Z_i\}_{i=1}^{C}$, containing C classes with each class $\{Z_{ij}\}_{j=1}^{C_i}$ consisting of a number of localized face images Z_{ij}, a total of $N = \sum_{i=1}^{C} C_i$ face images are available in the data set. The size of each image is represented as a column vector of length $J(= I_W \times I_h)$ by lexicographic ordering, i.e. $Z_{ij} \in R^J$, where R^J denotes the J-dimensional real space.

The proposed R-LDA introduced a weighted regulation parameter into the original LDA as follows:

$$\Psi = \arg\max_{\Psi} \frac{|\Psi^T S_b \Psi|}{\left|\frac{\eta}{2}(\Psi^T S_b \Psi) + \left(1 - \frac{\eta}{2}\right)(\Psi^T S_w \Psi)\right|} \tag{1}$$

where the $0 \leq \eta \leq 1$ is the regulation parameter, S_b is the between-class scatter matrice and S_W is the within-class matrice. Both matrices are respectively defined as:

$$S_b = \frac{1}{N}\sum_{i=1}^{C} C_i(\bar{Z}_i - \bar{Z})(\bar{Z}_i - \bar{Z})^T = \sum_{i=1}^{C} \Phi_{b,i}\Phi_{b,i}^T = \Phi_b\Phi_b^T \tag{2}$$

$$S_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{C_i}(Z_{ij} - \bar{Z}_i)(Z_{ij} - \bar{Z}_i)^T \tag{3}$$

where $\Phi_{b,i} = \left(\frac{C_i}{N}\right)^{1/2}(\bar{Z}_i - \bar{Z})$, $\Phi_b = [\Phi_{b,1}, \cdots, \Phi_{b,c}]$, $\bar{Z}_i = \frac{1}{N}\sum_{j=1}^{C_i} Z_j$ denotes the mean of the class Z_i and $\bar{Z}$ is the total mean.

The R-LDA seeks a set of $M \gg J$ feature basis vectors, denoted as $\{\psi_m\}_{m=1}^{M}$, in such the way that the optimal transformation of equation (1) preserves the given cluster structure which would maximize trace (S_b) and minimize trace (S_W). Trace (S_b)measures the closeness of the vectors within classes and trace (S_W) measures the separation of the vectors between classes. The solution to equation (1) is equivalent to the following generalized eigenvalue problem,

$$S_b\psi_m = \lambda_m S_W\psi_m, m = 1 \cdots m \tag{4}$$

Assuming S_W is non-singular, the feature basic vectors Ψ sought in equation (1) correspond to the first M eigenvectors with the largest eigenvalues of the matrice $(S_W^{-1}S_b)$. The M-dimensional representation is then obtained by projecting the original face images onto the subspace spanned by the M eigenvectors. As mentioned in the introduction, refer to [3-5], the null space of S_W may contain significant discriminant information if the projection of S_b is not zero. It means that the ratio of equation (1) can reach maximum if only when $|\Psi^T S_b \Psi| = 0$ and $|\Psi^T S_W \Psi| \neq 0$. In the other hand, there is no significant information to be lost if S_b is discarded. Due to the SSS problem, the performance of this algorithm is deteriorated rapidly by two factors, bias and variance. Bias is produced by the biased estimates of eigenvalues of

S_W (equation (3)), which draw outs the largest ones (biased high) and the smallest ones (biased low). The variance is caused by the unstable estimation of null space S_W, that may rise up highly. Both are so-called ill-posed situations and are determined by the degree of the SSS problem.

The regulation parameter, is added into equation (1) in order to control the these two factors. It can decrease the larger eigenvalues and increase the smaller eigenvalues, thereby counteracting the biasing and stabilizing the smallest eigenvalues. Add to this, η is set as $0 < \eta \leq 1$ in this paper. The steps of R-LDA issummarized as in Figure 1 (according to steps of the D-LDA of Yu and Yang [6])

Algorithm : Revised Regularized LDA

Input : Given a training set, $Z = \{Z_i\}_{i=1}^{C}$ with C classes with each class containing $Z_i = \{Z_{ij}\}_{j=1}^{C_i}$ face images, where $Z_{ij} \in R^J$, and the regulation parameter η.

Output : An M −dimensional LDA subspace spanned by Ψ, a $J \times M$ matrices with $M \ll J$.

1. Express $S_b = \Phi_b \Phi_b^T$, with $\Phi_b = [\Phi_{b,1}, \cdots, \Phi_{b,c}]$, $\Phi_{b,i} = \left(\frac{C_i}{N}\right)^{1/2} (\bar{Z}_i - \bar{Z})$, $\bar{Z}_i = \frac{1}{C_i} \sum_{j=1}^{C_i} Z_{ij}$, and $\bar{Z} = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{C_i} Z_{ij}$.
2. Find the m eigenvectors of $\Phi_b \Phi_b^T$, with non-zero eigenvalues, and denote them as $E_m = [e_1, \cdots, e_m]$.
3. Calculate the first m most significant eigenvectors (U_m) of S_b and their corresponding eigenvalues (Λ_b) by $U_m = \Phi_b E_m$ and $\Lambda_b = U_m^T S_b U_m$.
4. Let $H = U_m \Lambda_b^{1/2}$. Find the eigenvectors of $H^T S_w H$, $P = [p_1, \cdots, p_M]$ and sort it in the increasing eigenvalues order.
5. Choose the first $M (\leq m)$ eigenvectors P. Let P_M and Λ_w be the chosen eigenvectors and their eigenvalues, respectively.
6. Return $\Psi = \dfrac{H P_M \left(\left(\frac{\eta}{2} / \left(1 - \frac{\eta}{2}\right) \right) I + \Lambda_w \right)^{-1/2}}{1 - \frac{\eta}{2}}$.

Figure 1. The Pseudo Code for Computation of the Proposed R-LDA

3. ANN

ANNs provide a great alternative to the other conventional classifiers, feature extraction, decision making and pattern recognition. In this paper, the ANN has Feed Forward (FF) architecture within an input layer, a hidden layer and an output layer is used. The input layer is the vector constituted by $N \times M$ units of neuron ($N \times M$ pixel input face images). The input layer, I, is fully connected to the hidden layer, H, which has M neurons and is connected to the output layer, O, where the numbers of the output neurons are the classes of the face images. The output layer units are contained by K neurons which are active to 1 and -1 if the face is one class of face images. Figure 2 shows the architecture of the proposed ANN.

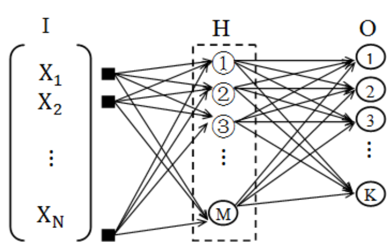

Figure 2. The Architecture of the Proposed ANN System

4. Experiments and Results

The experiments are divided into two parts. The first part is to obtain the all projection of the original face images of ORL database [7] onto a subspace spanned by the eigenvectors using the proposed R-LDA. The aim is to calculate the face descriptors of the face images. Then, the obtained face descriptors are used as the input vector for ANN experiments in the second part of the experiments. The recognition rates are obtained here. In addition, the experiments for mirror face images and combination face images of the original face images are also carried out in this section.

4.1. Database

The ORL database is used to examine our method. The images were taken with the persons in a frontal position against a dark background, where there are small variations of the background gray level also. The images present variations in facial expression, in facial position (there are slightly rotated faces) and in some other details like glasses and without glasses. The resolution of images is (112×92) pixels with 256 gray levels in tiff file. We partitioned the database into training set and testing set, denoted as G and Q, respectively. There is no overlapping between the G and Q.

4.2. Experiments for the Proposed R-LDA

To achieve the goal of the first part experiments, a pair (η, M) is required as the testing grid is applied to the experiments. The testing grid is defined as the outer product of $\eta = [0.001: 0.01: 1]$and $M = [20: 1: 34]$, where$[\min: \text{step}: \max]$ means the η and M to be increasing from min to max by step. For example, $M = [20: 1: 34]$ and $\eta = [0.001: 0.01: 1]$ denotes that a spaced vector is increasing from 20 elements to 34 elements over all the possible regularization parameter, η. The η is set greater than 0 to prevent the $H^T S_b H$ to be singular in the R-LDA experiments.

As mentioned in the introduction, FR is affected by SSS problem. Therefore, the strength of regularization parameter is significantly influenced by the number of available training samples for each person in the training set, G. Five experiments are carried in this section. They are examined according to the number of available training samples for each subject, regard as L. The ranging of L is set from $L = 1$ to $L = 5$. For a particular L, the training set is composed of $G = (L \times 35)$ samples and L images per person are chosen randomly. The rest of the images are set such as the testing set that is a composition of $Q = (350 - (L \times 35))$. Each training set is test on the proposed algorithm with the optimal pair of testing grid, which is denoted as (η^*, M^*). The optimal pair gives the best correct recognition rate (CRR) for each experiment.

4.3. Experiments for ANN

The ANN uses pairs of input and target vectors. The output vector of each input vector is compared with the target vector. The random weights and biases are assigned to the network in order to minimize the difference from the comparison with output vector and target vector. The weights and biases are updated every iteration to minimize the mean square error between output vector and target vector. If the error is minimum than the set error, the training process will stop. Otherwise, the weights and biases need to be updated.

In this paper, the input vectors are the eigenvectors obtained from R-LDA for each L. The target vectors are the vectors constructed by gray code, which each code represents one person. During the training task, the selection of the parameters is necessary in order to ensure an efficient operation of the network. Five training task experiments are conducted. Theirs initial weights and biases of the hidden layer and output layer are set to random number between -0.5 and 0.5. Besides, the training rate is set to 0.001 and the minimum error is set to 0.01.

To test the trained neural networks, as mentioned the beginning of Section 4, three testing sets are used to evaluate the recognition performance of the proposed FR system. Its sensitivities are measured in the CRR. Add to this, in order to enhance the accuracy of the recognition rate, the testing set is inserted with noise level from range 0 to 0.5. The CRRs of the testing sets are taken from the overage of five results of the experiments for each testing set.

4.4. Recognition Performance

The proposed RF system is analyzed by varying the number of available training samples for each subject used for feature extraction. The recognition performance is shown in Table 1.

Table 1. The Comparisons of the best found CRR(%) obtained on the test set and their corresponding parameter values.

L		1	2	3	4	5
R-LDA NN Original face images	AVE CRR	97.14	97.86	97.95	97.64	98.14
	(M^*)	34	34	27	27	27
	(η^*)	1	1	0.100	0.011	0.001
R-LDA NN Mirror face images	AVE CRR	94.21	95.56	96.03	97.14	97.64
	(M^*)	34	34	27	27	27
	(η^*)	1	1	0.100	0.011	0.001
R-LDA NN Combination face images	AVE CRR	93.14	95.14	96.14	96.56	97.03
	(M^*)	34	34	27	27	27
	(η^*)	1	1	0.100	0.011	0.001

The results derived from the proposed algorithm are compared with other techniques which are PCA, LDA, and PCA-NN [8]. The comparison of average CRR(%) results has been tabulated in Table 2.

Table 2. The CRR(%) performance of conventional PCA, LDA, PCA-NN versus proposed R-LDA NN.

L	PCA	LDA	PCA-NN	R-LDA NN
1	74.29	74.60	96.14	97.14
2	82.86	83.21	96.86	97.86
3	91.84	94.29	96.95	97.95
4	92.86	96.67	97.14	97.64
5	97.14	98.28	97.30	98.14

5. Conclusion

The paper presents a FR system using R-LDA and ANN algorithm. It can be seen from Table 1 that the mirrored face images cannot provide only additional training samples but also complementally information which enhance the performance of recognition rate. In the Table 2 one can see that the system is robust against the SSS problem and outperforms the traditional methods. From their test results it was showed that the proposed FR system has the acceptance recognition rates above 95%.

Acknowledgements

This work was supported by the Dong-A University research fund.

References

[1] J. Lu, K. Plataniotis and A. Venetsanopoulus, "Regularized Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application Problem to Face Recognition", Pattern Recognition Letter **(2005)**, vol. 26, no. 2, pp. 181-191.

[2] S. A. Nazeer, N. Omar and M. Khalid, "Face Recognition using Artificial Neural Networks Approach", International Conference on Signal Processing, Communication Networking (ICSN'70), **(2007)** Feb 22-24, pp. 420-425.

[3] K. Liu, Y. Cheng and J. Yang, "A Generalized Optimal Set of Discriminant Vectors", Pattern Recognition, **(1992a)**, pp. 731-739.

[4] K. Liu, Y. Cheng, J. Yang and X. Liu, "An Artificial Algorithm for Foleysammon Optical Set of Discriminant Vectors by Algebraic Method", Int. J. Pattern Recog. Artif.Intell.6, **(1992b)**, pp. 817-829.

[5] K. Liu, Y. Cheng and J. Yang, "Algebraic Feature Extraction for Image Recognition based on Optimal Discriminant Criterion", Pattern Recognition 26, **(1993)**, pp. 903-911.

[6] H. Yu and J. Yang, "A Direct LDA Algorithm for High-dimensional Data with Application to Face Recognition", Pattern Recognition, vol. 34**,(2001)**, pp. 2067-2070.

[7] ORL. Website: http://www.cl.cam.ac.uk/research/DTG/attarchive/.

[8] A. Mayank, J. Nikuni, K. Manish and A. Himanshu, "Face Recognition using Eigen Faces and Artificial Neural Network", International Journal of Computer Theory and Engineering, vol. 2, No. 4, **(2010)** August, pp. 1793-8201.

Authors

Lee Hui Kueh

She received her B.S. degree in Electronic Telecommunication Engineering from University of Malaysia Sarawak in 2001 and M.S. degree in Electrical Engineering from Dong-A University in 2007. From March 2007 till present, she is studying her PhD. Degree in Electrical Engineering in Dong-A University. She is interested in face recognition, image processing, intelligent theory and ANN.

Sang-Hyun Byun

He received his B.S degree in Engineering and Mechatronics Engineering from University of DONG-EUI in 2012. He is interested in face recognition, Networking, security and intelligent theory.

John-Tark Lee

He received his B.S, and M.S degrees, both in Electrical Engineering from Dong-A University, in 1979 and 1981, respectively. In 1988, he received his Ph.D degree in Electrical Engineering fr om Chung-Ang University. From 1983 to 1985, he was with LG Electronic Co. Ltd. and since 1985, he joined the Dong-A University until present as a professor of Electrical Engineering Department. His research interests are Fuzzy theory. Neuro-Science, GA and intelligent controller design including stability analysis.