TWO-DIMENSIONAL LINEAR DISCRIMINANT ANALYSIS OF PRINCIPLE COMPONENT VECTORS FOR FACE RECOGNITION

P. Sanguansat, W. Asdornwised, S. Jitapunkul

Department of Electrical Engineering
Chulalongkorn University
Bangkok
Thailand

S. Marukatat

National Electronics and
Computer Technology Center (NECTEC)
Pathumthani
Thailand

ABSTRACT

In this paper, we proposed a new Two-Dimensional Linear Discriminant Analysis (2DLDA) method. Based on Two-Dimensional Principle Component Analysis (2DPCA), face image matrices do not need to be previously transformed into a vector. In this way, the spatial information can be preserved. Moreover, the 2DLDA also allows avoiding the Small Sample Size (SSS) problem, thus overcoming the traditional LDA. We combine 2DPCA and our proposed 2DLDA on the Two-Dimensional Linear Discriminant Analysis of principle component vectors framework. Our framework consists of two steps: first we project an input face image into the family of projected vectors via 2DPCA-based technique, second we project from these space into the classification space via 2DLDA-based technique. This does not only allows further reducing of the dimension of feature matrix but also improving the classification accuracy. Experimental results on ORL and Yale face database showed an improvement of 2DPCA-based technique over the conventional PCA technique.

1. INTRODUCTION

In face recognition, the linear subspace techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are the most popular ones [1, 2, 3, 4, 5]. The PCA's criterion chooses the subspace in the function of data distribution while LDA chooses the subspace which yields maximal inter-class distance while keeping the intra-class distance small. Both techniques intend to project the vector representing face image onto lower dimensional subspace, which the 2D face image matrices must be previously transformed into vectors and then concatenated into a matrix. This is the cause of three serious problems in particular approaches. First of all, the feature vectors with high dimension will leads to the curse of dimensionality. Secondly, the spatial structure information could be lost. Finally, in face recognition task, the available number of training samples is relatively very small compared to the feature dimension that is the cause of the Small Sample Size (SSS) problem.

Various solutions have been proposed for solving the SSS problem [3, 4, 5, 6, 7, 8]. Among these LDA extensions, the discriminant analysis of principle components framework [3, 4, 5] demonstrates a significant improvement when apply Linear Discriminant Analysis (LDA) over principle components from the PCA-based subspace. Since both PCA and LDA can overcome the drawbacks of each other. PCA is constructed around the criteria of preserving the data distribution. Hence, it is suited for face representation. However, in the classification tasks, PCA only give the orthonormal transformation of original features. This can be used, along with eigenvalues, to whiten the distribution. Nevertheless, this transformations do not take into account the between classes relationship. In general, the discriminatory power depends on both within and between classes relationship. LDA considers this relationship via scatter matrices analysis of within and between-class scatter matrices. Taking this information into account as in LDA allows further improvement. Especially, when there are prominent variation in lighting condition and expression. However, LDA has certain two drawbacks when directly applied to the original input space [5]. First of all, some non-face information such as image background are regarded by LDA as the discriminant information. This causes misclassification when the face of the same subject is presented on different background. Secondly, when SSS problem has occurred, the within-class scatter matrix be singular, so-called the singularity problem. Projecting the high dimensional input space into low dimensional subspace via PCA can solve these LDA problems. Nevertheless, the spatial structure information still be lost.

Recently, an original technique called Two-Dimensional Principal Component Analysis (2DPCA) was proposed [9], in which the image covariance matrix is computed directly on image matrices so the spatial information can be preserved. This yields a covariance matrix whose dimension equals to

Thanks to Thailand Graduate Institute of Science and Technology (TGIST) and the Rajadapisek Sompoj endowment and Cooperation Project Between Department of Electrical Engineering and Private Sector for Research and Development.

the width of face image. This is far smaller than the real size of face image. Therefore, the image covariance matrix can be better estimated and relieved the SSS problem. Evidently, the experimental results in [9] shown the improvement of 2DPCA over PCA on several face databases. However, like PCA, 2DPCA is prefer face representation to face recognition. For better performance of recognition task, LDA is necessary. Unfortunately, the linear transformation of 2DPCA reduces only the number of columns. The number of rows still equal to the original image. Thus, the SSS problem will be appeared when LDA is performed after 2DPCA directly. In order to overcome this problem, we propose a modified LDA, called Two-Dimensional Linear Discriminant Analysis (2DLDA),

based on the 2DPCA concept. Applying 2DLDA to 2DPCA not only can solves the SSS problem and the curse of dimensionality dilemma but also allows working directly on the image matrix in all projections. Hence, spatial structure information is maintained and the size of all scatter matrices just equal to the width of face image. Furthermore, the face image do not need to be resized, thus all information still be preserved.

In this paper, we will show that there are the biased PCAs were embedded in 2DPCA so it does not surprise that why 2DPCA is better than only one PCA. And we will show how to put the idea of 2DPCA under the LDA's criterion. Moreover, following to the previous works on this field [3, 4, 5], a combination framework of 2DPCA and 2DLDA is also evaluated. The accuracy of this proposed framework is demonstrated on real world face databases. The experimental results can promise the performance of our proposed framework.

2. 2DPCA

2.1. Basic Idea

In 2DPCA [9], the image covariance matrix $\mathbf{G}$ was defined as

$$\mathbf{G} = E[(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})], \quad (1)$$

where $\mathbf{A}$ represents the face image. This is much smaller than the size of real covariance matrix needed in PCA, therefore can be computed more accurately on small training set. Given a database of M training image matrices $\mathbf{A}_k,\ k = 1, \ldots, M$ with same dimension m by n. The matrix $\mathbf{G}$ is computed in a straightforward manner by

$$\mathbf{G} = \frac{1}{M}\sum_{k=1}^{M}(\mathbf{A}_k - \bar{\mathbf{A}})^T(\mathbf{A}_k - \bar{\mathbf{A}}), \quad (2)$$

where $\bar{\mathbf{A}}$ denotes the average image, $\bar{\mathbf{A}} = \frac{1}{M}\sum_{k=1}^{M}\mathbf{A}_k$.

Let $\mathbf{x_1}, \ldots, \mathbf{x_d}$ be d selected largest eigenvectors of $\mathbf{G}$. Each image $\mathbf{A}$ is projected onto these d dimensional subspace. The projected image $\mathbf{Y}$ is then an m by d matrix given by $\mathbf{Y} = \mathbf{AX}$ where $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x_d}]$ is a n by d projection matrix.

2.2. Comparisons of PCA and 2DPCA

The sketch for the reason that why 2DPCA is better than PCA should lie on the answer of the following question: What is happen when the inputs of PCA are the rows of each image instead of entire images?

Let $\{\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \mathbf{\Gamma}_3, \ldots, \mathbf{\Gamma}_n\}$ be the set of rows of an image $\mathbf{A}$ and $\mathbf{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Gamma}_i$ is the average of this set. The i-th row differ from the average by the vector $\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \Psi$. Therefore, the covariance matrix $\mathbf{C}$ of PCA is given by $\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Phi}_i\mathbf{\Phi}_i^T = \frac{1}{n}\mathbf{\Upsilon}\mathbf{\Upsilon}^T$, where the matrix $\mathbf{\Upsilon} = [\mathbf{\Phi}_1\ \mathbf{\Phi}_2\ \ldots\ \mathbf{\Phi}_n]$. Following Eigenface's algorithm [2], the optimal projection vectors can be determined as the eigenvectors of matrix $\mathbf{\Upsilon}^T\mathbf{\Upsilon}$. If zero mean, $\mathbf{\Psi} = 0$, the image covariance matrix $\mathbf{G}$ in Eq. (1) can be rewritten as

$$\mathbf{G} = E[(\mathbf{\Upsilon} - E\mathbf{\Upsilon})^T(\mathbf{\Upsilon} - E\mathbf{\Upsilon})] = E[\mathbf{\Upsilon}^T\mathbf{\Upsilon}] - \beta, \quad (3)$$

where $\beta = E\mathbf{\Upsilon}^T E\mathbf{\Upsilon}$. From this point, 2DPCA can be explained in a novel perspective as a collection of biased PCAs. Indeed, the number of training samples of conventional PCA is only M while the number of training samples of 2DPCA is $M \times m$. Since the dimension of $\mathbf{G}$ is $n \times n$ with $n < M \times m$, thus 2DPCA can provide the full rank image covariance matrix. This is the reason of the improvement of 2DPCA over the original PCA.

3. THE PROPOSED 2DLDA

Let $\mathbf{Z}$ be an n by q matrix. A matrix $\mathbf{A}$ is projected onto $\mathbf{Z}$ via the linear transformation, **V=AZ**. In this 2DLDA, we search for the projection matrix $\mathbf{Z}$ maximizing the Fisher's discriminant criterion [3]:

$$J(\mathbf{Z}) = \frac{tr\,(\mathbf{S}_b)}{tr\,(\mathbf{S}_w)}, \quad (4)$$

where $\mathbf{S}_w$ is the *within-class scatter matrix* and $\mathbf{S}_b$ is the *between-class scatter matrix*. The within-class scatter matrix describes how data are scattered around the means of their respective class, $\mathbf{S}_w = \sum_{i=1}^{K} Pr(\omega_i)E\left[(\mathbf{HZ})(\mathbf{HZ})^T|\omega = \omega_i\right]$, where k is the number of classes, $Pr(\omega_i)$ is the prior probability of each class, and $\mathbf{H} = \mathbf{A} - E\mathbf{A}$. The between-class scatter matrix describes how different classes, represented by their expected value, are scattered around the mixture means by $\mathbf{S}_b = \sum_{i=1}^{K} Pr(\omega_i)E\left[(\mathbf{FZ})(\mathbf{FZ})^T\right]$, where $\mathbf{F} = E[\mathbf{A}|\omega = \omega_i] - E[\mathbf{A}]$. Using the linearity of both the trace function and the expectation, $J(\mathbf{Z})$ may be rewritten as

$$\begin{aligned} J(\mathbf{Z}) &= \frac{tr(\sum_{i=1}^{K} Pr(\omega_i)E\left[(\mathbf{FZ})(\mathbf{FZ})^T\right])}{tr(\sum_{i=1}^{K} Pr(\omega_i)E\left[(\mathbf{HZ})(\mathbf{HZ})^T|\omega = \omega_i\right])} \\ &= \frac{tr\left(\mathbf{Z}^T\left(\sum_{i=1}^{K} Pr(\omega_i)E\left[\mathbf{F}^T\mathbf{F}\right]\right)\mathbf{Z}\right)}{tr\left(\mathbf{Z}^T\left(\sum_{i=1}^{K} Pr(\omega_i)E\left[\mathbf{H}^T\mathbf{H}|\omega = \omega_i\right]\right)\mathbf{Z}\right)} \\ &= \frac{tr(\mathbf{Z}^T\tilde{\mathbf{S}}_b\mathbf{Z})}{tr(\mathbf{Z}^T\tilde{\mathbf{S}}_w\mathbf{Z})}. \end{aligned} \quad (5)$$

And $\tilde{\mathbf{S}}_b$ and $\tilde{\mathbf{S}}_w$ can be evaluated as follow

$$\tilde{\mathbf{S}}_b = \sum_{i=1}^{K} \frac{n_i}{K}(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T(\bar{\mathbf{A}}_i - \bar{\mathbf{A}}) \quad (6)$$

$$\tilde{\mathbf{S}}_w = \sum_{i=1}^{K} \frac{n_i}{K}\sum_{\mathbf{A}_k \in \omega_i}(\mathbf{A}_k - \bar{\mathbf{A}}_i)^T(\mathbf{A}_k - \bar{\mathbf{A}}_i), \quad (7)$$

where n_i and ω_i, $\bar{\mathbf{A}}_i$ are the number of elements and the expected value of class ω_i, respectively. Then the optimal projection vector can be found by solving the following generalized eigenvalue problem, $\tilde{\mathbf{S}}_b\mathbf{Z} = \tilde{\mathbf{S}}_w\Lambda\mathbf{Z}$, where Λ is a diagonal matrix with eigenvalues on the main diagonal. Note that, the size of scatter matrices involved in eigen decomposition is only n by n. Thus with the limited training set, this decomposition is more reliably than the decomposition on classical covariance matrix.

4. 2DLDA OF 2DPCA

The discriminant analysis of principle components framework in [4] is applied in this section. Our framework consists of 2DPCA and 2DLDA step. From Section 2, we obtain a linear transformation matrix $\mathbf{X}$ on which each input face image $\mathbf{A}$ is projected. From this 2DPCA step, a feature matrix $\mathbf{Y}$ is obtained. This matrix $\mathbf{Y}$ is used as the input of 2DLDA step. Thus, the evaluation of within and between-class scatter matrices in this step will be slightly changed. From (6) and (7), image matrix $\mathbf{A}$ is substituted for the 2DPCA feature matrix $\mathbf{Y}$ as follows

$$\tilde{\mathbf{S}}_b^Y = \sum_{i=1}^{K} \frac{n_i}{K}(\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})^T(\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}}) \quad (8)$$

$$\tilde{\mathbf{S}}_w^Y = \sum_{i=1}^{K} \frac{n_i}{K}\sum_{\mathbf{Y}_k \in \omega_i}(\mathbf{Y}_k - \bar{\mathbf{Y}}_i)^T(\mathbf{Y}_k - \bar{\mathbf{Y}}_i) \quad (9)$$

where $\mathbf{Y}_k$ is the feature matrix of the k-th image matrix $\mathbf{A}_k$, $\bar{\mathbf{Y}}_i$ be the average of $\mathbf{Y}_k$ which belong to class ω_i and $\bar{\mathbf{Y}}$ denotes a overall mean of $\mathbf{Y}$, $\bar{\mathbf{Y}} = \frac{1}{M}\sum_{k=1}^{M}\mathbf{Y}_k$. The 2DLDA optimal projection matrix $\mathbf{Z}$ should be obtained via solving the eigenvalue problem in Section 3. Finally, the composite linear transformation, **D=AXZ**, is used to map the face image space into the classification space. The matrix $\mathbf{D}$ is 2DLDA of 2DPCA feature matrix of image $\mathbf{A}$ with dimension m by q. However, the number of 2DLDA feature vectors q cannot exceed the number of principle component vectors d. In general case ($q < d$), the dimension of $\mathbf{D}$ is less than $\mathbf{Y}$ in Section 2. Thus, 2DLDA of 2DPCA can reduce the classification time compared to 2DPCA.

A nearest neighbor classifier is used for classification as the one used in 2DPCA [9].

5. EXPERIMENTAL RESULTS

The two well-known Yale[1] and ORL[2] database are used in all experiments. The Yale database contains 165 images of 15 subjects. There are 11 images per subject, one for each of the following facial expressions or configurations: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. Each image was manually cropped and resized to 100×80 pixels. The ORL database contains images from 40 individuals, each providing 10 different images. For some subjects, the images were taken at different times. The facial expressions open or closed eyes, smiling or non smiling and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent. All images are gray scale and normalized to a resolution of 112×92 pixels.

To investigate the effect of number of feature vectors, we vary the number of principle component vectors d and the number of 2DLDA feature vectors q from 1 to 30 on both databases because the highest recognition accuracy lie in this interval. On Yale database, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used to train, and the six remaining images for test. And the first five image samples are used to train per class, and the six remaining images for test on ORL database. Fig. 1 (a) and (b) show the relationship between recognition accuracy and the number of feature vectors of 2DPCA, 2DLDA, and 2DPCA+2DLDA on Yale and ORL database respectively. It should be note that the performance of 2DPCA+2DLDA actually depends on both principle component vectors and 2DLDA feature vector, as shown in Fig. 1 (c) and (d). Therefore, for comparison reason, we plot the top recognition of Fig. 1 (c) and (d), in the direction of the number of 2DLDA feature vectors axis, in Fig. 1 (a) and (b), respectively. The results of pure 2DLDA are in agreement with pure LDA. That is the pure 2DLDA method includes the information which not useful for classification as its discriminant information. The 2DPCA+2DLDA method can achieve higher recognition rate than other methods. Table 1 and 2 show the comparisons of 3 methods on the highest recognition accuracy. In all experiments, The recognition rate of 2DPCA+2DLDA was superior to 2DPCA and 2DLDA. Especially, 2DPCA+2DLDA use only 7 feature vectors while 2DPCA use 23 principle component vectors to obtain the highest recognition rate on Yale database. In case of ORL database, 2DPCA+2DLDA obtain the highest recognition rate more than 2DPCA in same dimension.

Furthermore, since the number of feature vectors in case of 2DPCA+2DLDA must be less or equal to the number of principle component vectors of 2DPCA in the first step, so

[1]The Yale database is available for download from http://cvc.yale.edu

[2]The ORL database is available for download from http://www.cam-orl.co.uk

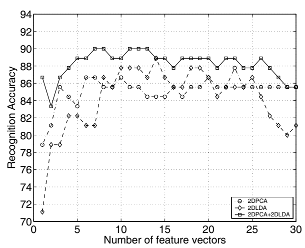
Table 1. The top recognition accuracy comparisons of 2DPCA, 2DLDA and 2DPCA+2DLDA on Yale database

Method	Accuracy (%)	d	q	Dimension	Training time	Classification time	Overall time
2DPCA	87.78	23	N/A	100×23	1 (0.17 s)	1 (3.14 s)	1 (3.31 s)
2DLDA	88.89	N/A	14	100×14	0.94 (0.16 s)	0.62 (1.94 s)	0.63 (2.10 s)
2DLDA+2DPCA	90	21	7	100×7	1.14 (0.24 s)	0.32 (1.01 s)	0.38 (1.25 s)

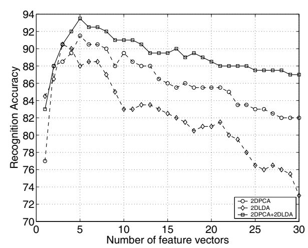
Table 2. The top recognition accuracy comparisons of 2DPCA, 2DLDA and 2DPCA+2DLDA on ORL database

Method	Accuracy (%)	d	q	Dimension	Training time	Classification time	Overall time
2DPCA	91.5	5	N/A	112×5	1 (0.33 s)	1 (4.33 s)	1 (4.688 s)
2DLDA	90.5	N/A	3	112×3	1.51 (0.50 s)	0.64 (2.75 s)	0.7 (3.281 s)
2DLDA+2DPCA	93.5	14	5	112×5	1.94 (0.64 s)	1 (4.33 s)	1.07 (5.219 s)

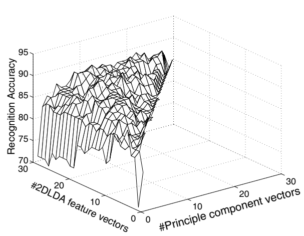
the time consumed for classification can be reduced.



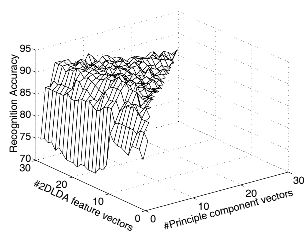
(a) 3 methods on Yale



(b) 3 methods on ORL



(c) 2DPCA+2DLDA on Yale



(d) 2DPCA+2DLDA on ORL

Fig. 1. Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA.

6. CONCLUSIONS

In this paper, we proposed a Two-Dimensional Linear Discriminant Analysis (2DLDA) for face recognition based on 2DPCA concept. All scatter matrices in 2DLDA are much smaller and compute eigenvectors easier than LDA techniques. The SSS problem of within-class scatter matrix is relieved within the 2DLDA framework. By using 2DLDA of principle component vectors framework, 2DLDA can improved a performance over 2DPCA in face recognition task, as shown in all our experimental results on the well-known face databases.

7. REFERENCES

[1] L. Sirovich and M. Kirby, "Low-dimensional procedure for characterization of human faces," *J. Optical Soc. Am.*, vol. 4, pp. 519–524, 1987.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 19, pp. 711–720, July 1997.

[4] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principle components for face recognition," Japan, 1998, IEEE 3rd Inter. Conf. on Automatic Face and Gesture Recognition.

[5] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical performance analysis of linear discriminant classifiers," in *CVPR*. 1998, pp. 164–171, IEEE Computer Society.

[6] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[7] J. Lu, K. Plataniotis, and A. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recogn. Lett.*, vol. 24, no. 16, pp. 3079–3087, 2003.

[8] R. Huang, Q. Liu, H. Lu, and S. Ma, "Solving the small sample size problem of LDA," *Pattern Recognition*, vol. 3, pp. 29–32, 2002.

[9] J. Yang, D. Zhang, A. Frangi, and J. Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 26, pp. 131–137, Jan. 2004.