IEICE TRANS. INF. & SYST., VOL.E94–D, NO.4 APRIL 2011

LETTER

Non-iterative Symmetric Two-Dimensional Linear Discriminant Analysis

Kohei INOUE[†a)], **Kenji HARA**[†], *and* **Kiichi URAHAMA**[†], *Members*

SUMMARY Linear discriminant analysis (LDA) is one of the well-known schemes for feature extraction and dimensionality reduction of labeled data. Recently, two-dimensional LDA (2DLDA) for matrices such as images has been reformulated into symmetric 2DLDA (S2DLDA), which is solved by an iterative algorithm. In this paper, we propose a non-iterative S2DLDA and experimentally show that the proposed method achieves comparable classification accuracy with the conventional S2DLDA, while the proposed method is computationally more efficient than the conventional S2DLDA.
key words: *symmetric two-dimensional linear discriminant analysis, face recognition, dimensionality reduction*

1. Introduction

Linear discriminant analysis (LDA) is one of the well-known schemes for feature extraction and dimensionality reduction of labeled data. Recently, LDA, which was originally formulated for discriminating labeled vectors [1], has been extended to two-dimensional LDA (2DLDA) for discriminating labeled matrices such as images. Yang et al. [2] proposed a 2DLDA which performs the uncorrelated image matrix-based LDA (IMLDA) proposed by Yang et al. [3] twice: the first and the second times are in horizontal and vertical directions, respectively. Ye et al. [4] also proposed another 2DLDA which attempts to optimize two transformation matrices L and R for reducing dimensions of rows and columns, respectively. In Ye's 2DLDA [4], L and R are optimized with different objective functions by an alternating algorithm. However, the optimization with L may not be optimal for the objective function for R, and vice versa. Ye et al. [4] demonstrated that the accuracy curves for face recognition are stable with respect to the number of iterations, which led them to adopt a single-iteration renewal procedure. We proposed a non-iterative 2DLDA [5] which is a combination of Yang's 2DLDA [2] and Ye's one [4]. Luo et al. [6] pointed out the ambiguity problem on the objective function for 2DLDA, i.e., there are several choices for the objective function for 2DLDA, and derived a reasonable objective function for 2DLDA called symmetric 2DLDA (S2DLDA). However, S2DLDA is no longer reduced to any generalized eigenvalue problems. Therefore, they used a gradient ascent approach. However, their algorithm does not necessarily increase the objective function value monotonically as demonstrated in their paper [6], and therefore it is difficult to determine the number of iterations of the renewal procedure.

This work is motivated by the above observation of the state-of-the-art 2DLDA techniques. That is, Luo's solution to the ambiguity problem in 2DLDA, i.e., S2DLDA, caused other problems that S2DLDA is no longer reduced to any generalized eigenvalue problems and therefore Luo et al. [6] adopted a gradient ascent approach which has the difficulty in determining the optimal number of iterations. If we can reformulate S2DLDA in a way that the reformulation is reduced to a generalized eigenvalue problem, then we will overcome the above problems.

In this paper, we propose a non-iterative S2DLDA which can be reduced to a generalized eigenvalue problem and therefore has an analytical solution. Experimental results show that the proposed method achieves comparable classification accuracy with S2DLDA [6] and is computationally efficient in the training stage.

The rest of this paper is organized as follows: Section 2 summarizes S2DLDA [4]. Section 3 proposes a non-iterative S2DLDA. Section 4 shows experimental results. Finally, Sect. 5 concludes this paper.

2. Symmetric 2DLDA

Let $A_i \in \mathbb{R}^{r\times c}$ for $i = 1, \ldots, m$ be m images categorized into n classes $\Pi_1, \ldots, \Pi_n$. Let $M_j = \frac{1}{n_j}\sum_{A_i\in\Pi_j} A_i$ be the mean of the jth class for $j = 1, \ldots, n$ where $n_j = |\Pi_j|$ is the number of elements in Π_j, and let $M = \frac{1}{m}\sum_{j=1}^{n}\sum_{A_i\in\Pi_j} A_i$ be the global mean. In 2DLDA, the dimensions of A_i are reduced by

$$B_i = L^T A_i R \in \mathbb{R}^{\tilde{r}\times\tilde{c}},$$

where $L \in \mathbb{R}^{r\times\tilde{r}}$ for $\tilde{r} \le r$ and $R \in \mathbb{R}^{c\times\tilde{c}}$ for $\tilde{c} \le c$. In symmetric 2DLDA (S2DLDA), L and R are obtained by solving the following maximization problem:

$$\max_{L,\,R}\ \mathrm{tr}\left[\left(R^T S_w^L R\right)^{-1} R^T S_b^L R\right] + \mathrm{tr}\left[\left(L^T S_w^R L\right)^{-1} L^T S_b^R L\right], \tag{1}$$

where tr denotes the matrix trace, S_w^R and S_w^L are the within-class scatter matrices defined by

$$S_w^R = \sum_{j=1}^{n}\sum_{A_i\in\Pi_j}(A_i - M_j)RR^T(A_i - M_j)^T,$$

Manuscript received October 5, 2010.
Manuscript revised December 20, 2010.
[†]The authors are with the Faculty of Design, Kyushu University, Fukuoka-shi, 815–8540 Japan.
a) E-mail: k-inoue@design.kyushu-u.ac.jp
DOI: 10.1587/transinf.E94.D.926

$$S_w^L = \sum_{j=1}^{n} \sum_{A_i \in \Pi_j} (A_i - M_j)^T L L^T (A_i - M_j),$$

respectively, and S_b^R and S_b^L are the between-class scatter matrices defined by

$$S_b^R = \sum_{j=1}^{n} n_j (M_j - M) R R^T (M_j - M)^T,$$

$$S_b^L = \sum_{j=1}^{n} n_j (M_j - M)^T L L^T (M_j - M),$$

respectively. Let J be the objective function in (1). Then J can be rewritten as follows:

$$\begin{aligned} J &= \mathrm{tr}\begin{bmatrix} \left(R^T S_w^L R\right)^{-1} R^T S_b^L R & 0 \\ 0 & \left(L^T S_w^R L\right)^{-1} L^T S_b^R L \end{bmatrix} \\ &= \mathrm{tr}\left(\begin{bmatrix} \left(R^T S_w^L R\right)^{-1} & 0 \\ 0 & \left(L^T S_w^R L\right)^{-1} \end{bmatrix} \begin{bmatrix} R^T S_b^L R & 0 \\ 0 & L^T S_b^R L \end{bmatrix}\right) \\ &= \mathrm{tr}\left(\begin{bmatrix} R^T S_w^L R & 0 \\ 0 & L^T S_w^R L \end{bmatrix}^{-1} \begin{bmatrix} R^T S_b^L R & 0 \\ 0 & L^T S_b^R L \end{bmatrix}\right) \\ &= \mathrm{tr}\left[\left(\begin{bmatrix} 0 & L \\ R & 0 \end{bmatrix}^T \begin{bmatrix} S_w^R & 0 \\ 0 & S_w^L \end{bmatrix} \begin{bmatrix} 0 & L \\ R & 0 \end{bmatrix}\right)^{-1} \right. \\ &\qquad \left. \begin{bmatrix} 0 & L \\ R & 0 \end{bmatrix}^T \begin{bmatrix} S_b^R & 0 \\ 0 & S_b^L \end{bmatrix} \begin{bmatrix} 0 & L \\ R & 0 \end{bmatrix}\right] \\ &= \mathrm{tr}\left[\left(U^T S_w U\right)^{-1} U^T S_b U\right], \end{aligned}$$

where

$$U = \begin{bmatrix} 0 & L \\ R & 0 \end{bmatrix}, \quad S_w = \begin{bmatrix} S_w^R & 0 \\ 0 & S_w^L \end{bmatrix}, \quad S_b = \begin{bmatrix} S_b^R & 0 \\ 0 & S_b^L \end{bmatrix}.$$

Since both S_w and S_b contain L and R, (1) has no analytical solution. Therefore, Luo et al. [6] used a gradient ascent approach. However, their algorithm does not necessarily increase the objective function value monotonically, and therefore it is difficult to determine the number of iterations in their approach.

3. Non-iterative Symmetric 2DLDA

We define the row-row within-class and between-class scatter matrices [5] as

$$S_w^r = \sum_{j=1}^{n} \sum_{A_i \in \Pi_j} (A_i - M_j)(A_i - M_j)^T,$$

$$S_b^r = \sum_{j=1}^{n} n_j (M_j - M)(M_j - M)^T,$$

respectively. We also define the column-column within-class and between-class scatter matrices [5] as

$$S_w^c = \sum_{j=1}^{n} \sum_{A_i \in \Pi_j} (A_i - M_j)^T (A_i - M_j),$$

$$S_b^c = \sum_{j=1}^{n} n_j (M_j - M)^T (M_j - M),$$

respectively. Then we form a combined within-class scatter matrix

$$\tilde{S}_w = \begin{bmatrix} S_w^r & 0 \\ 0 & S_w^c \end{bmatrix}$$

and a combined between-class scatter matrix

$$\tilde{S}_b = \begin{bmatrix} S_b^r & 0 \\ 0 & S_b^c \end{bmatrix}.$$

Now we formulate a non-iterative S2DLDA as follows:

$$\max_{\tilde{U}} \ \mathrm{tr}\left[\left(\tilde{U}^T \tilde{S}_w \tilde{U}\right)^{-1} \tilde{U}^T \tilde{S}_b \tilde{U}\right]. \quad (2)$$

Since S_w^r, S_b^r, S_w^c and S_b^c do not contain L and R and therefore $\tilde{S}_w$ and $\tilde{S}_b$ do not contain L and R either, the solution to (2) is given by the principal eigenvectors of $\tilde{S}_w^{-1}\tilde{S}_b$ as in the conventional LDA [1]. Therefore, we can avoid using a gradient ascent approach which has the difficulty in determining the number of iterations.

The procedure of the non-iterative S2DLDA is summarized in Table 1. Since the projection matrices L and R are intermingled in $\tilde{U}$, we separate them in steps 5-1 and 5-2 in Table 1 and express them as $\tilde{L}$ and $\tilde{R}$. Since $\tilde{S}_w$ and $\tilde{S}_b$ are block diagonal, $\tilde{S}_w^{-1}\tilde{S}_b$ is also block diagonal. Therefore, the kth eigenvector $\tilde{u}_k$ of $\tilde{S}_w^{-1}\tilde{S}_b$ has the form $\tilde{u}_k = [\tilde{u}_{k,1}, \ldots, \tilde{u}_{k,\tilde{r}}, 0, \ldots, 0]^T = [x^T, 0, \ldots, 0]^T$ or $\tilde{u}_k = [0, \ldots, 0, \tilde{u}_{k,\tilde{r}+1}, \ldots, \tilde{u}_{k,K}]^T = [0, \ldots, 0, y^T]^T$. In step 5-2, we select a nonzero subvector x or y by comparing their Euclidean norm.

Table 1 The proposed algorithm.

Algorithm: Non-iterative symmetric 2DLDA
1. Compute S_w^r, S_b^r, S_w^c and S_b^c.
2. Compute $\tilde{S}_w$ and $\tilde{S}_b$.
3. Compute the first K principal eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K$ of $\tilde{S}_w^{-1}\tilde{S}_b$ and the corresponding eigenvectors $\tilde{u}_1, \ldots, \tilde{u}_K$.
4. Initialize $\tilde{L} = []$ and $\tilde{R} = []$.
5. For $k = 1, \ldots, \tilde{K}$, do the following:
5-1. Let $x = [\tilde{u}_{k,1}, \ldots, \tilde{u}_{k,\tilde{r}}]^T$ and $y = [\tilde{u}_{k,\tilde{r}+1}, \ldots, \tilde{u}_{k,K}]^T$, where $\tilde{u}_{k,l}$ is the lth element of $\tilde{u}_k$.
5-2. If $\lVert x \rVert \geq \lVert y \rVert$, then renew $\tilde{L} \leftarrow [\tilde{L}, x]$, or else $\tilde{R} \leftarrow [\tilde{R}, y]$.
6. Reduce the dimension of A_i by $\tilde{B}_i = \tilde{L}^T A_i \tilde{R}$.

4. Experimental Results

In this section, we evaluate the performance of the proposed non-iterative S2DLDA compared with the conventional S2DLDA [6]. We use the ORL face image database [7] in our experiments. Figure 1 shows some face images in this database, which contains the face images of 40 persons. There are 10 different images per person. We use 3 images per person for training, and the remaining 7 images for testing. The size of each image is 112×92 pixels, i.e., $r = 112$ and $c = 92$. We set $\tilde{r} = 10$ and $\tilde{c} = 6$ by leave-one-out

Fig. 1 ORL face images.

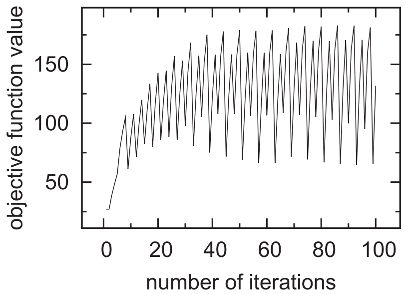

Fig. 2 Objective function value for symmetric 2DLDA.

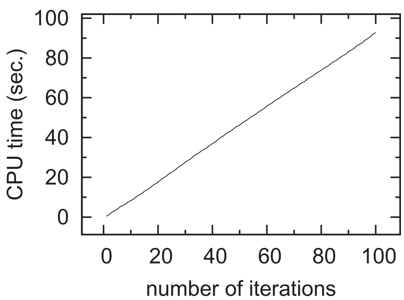

Fig. 3 CPU time for symmetric 2DLDA.

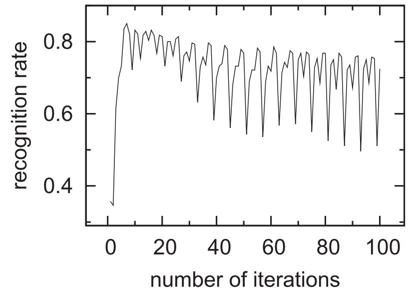

Fig. 4 Recognition rate for symmetric 2DLDA.

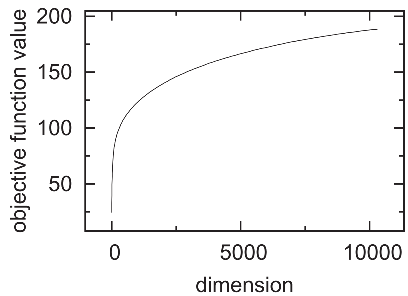

Fig. 5 Objective function value for the proposed method.

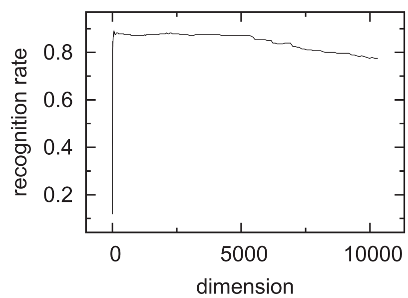

Fig. 6 Recognition rate for the proposed method.

cross-validation in non-iterative S2DLDA. Figure 2 shows the variation in the objective function value J for S2DLDA, where the horizontal axis denotes the number of iterations and the vertical axis denotes the value of J. As demonstrated in [6], the value of J fluctuates with the increase in the number of iterations. That is, the gradient ascent approach by Luo et al. [6] does not guarantee that J increases monotonically. One of the reasons that J fluctuates in their approach is that L and R are normalized once every $c = 3$ iterations where c is a parameter used in their paper [6]. Figure 3 shows the CPU time for training in S2DLDA. The CPU time linearly increases with the number of iterations. We performed our experiments using MATLAB on a Pentium 4 CPU 3.40 GHz machine with 2.00 GB RAM. Figure 4 shows the recognition rate for S2DLDA, where we classified the test set with the nearest-neighbor rule [8]. The highest recognition rate 0.85 is obtained at the 7 iterations.

Next, we show the results for non-iterative S2DLDA. Figure 5 shows the maximum objective function value in (2), where the horizontal axis denotes the dimension of $\tilde{B}_i = \tilde{L}^T A_i \tilde{R}$ or the number of the elements in $\tilde{B}_i$. The CPU time for training in non-iterative S2DLDA is 0.218 seconds, which is shorter than that (0.469 seconds) for one iteration in S2DLDA. Figure 6 shows the recognition rate for non-iterative S2DLDA. The highest recognition rate 0.889 is obtained when the number of dimensions is $54 = \tilde{r} \times \tilde{c}$ where $\tilde{r} = 9$ and $\tilde{c} = 6$. Thus, the proposed non-iterative S2DLDA achieves comparable recognition rate with the conventional S2DLDA, and the CPU time for training in non-iterative S2DLDA is shorter than that in S2DLDA.

Figure 7 shows a confusion matrix for the proposed method. The highest rate in the off-diagonal elements is 0.8 which is the (17, 36) element, i.e., the 17th person is likely to be confused with the 36th person. Figures 8 and 9 show the face images of the 36th person for training and that of the 17th person for testing, respectively.

We compared the performance of the proposed method with that of the conventional methods: no dimension reduction (baseline), PCA+LDA, 2DLDA [4], S2DLDA [6] and image shrinkage by a photo-retouch software (Adobe Photoshop CS). We excluded the conventional LDA from

Table 2 Classification accuracy.

	baseline	PCA+LDA	2DLDA	S2DLDA	photo-retouch	proposed
ORL	0.857	0.768	0.829	0.811	0.864	0.889
IFD	0.778	0.750	0.534	0.773	0.824	0.818
UMIST	0.556	0.569	0.272	0.484	0.600	0.663
CHFD	0.654	0.520	0.109	0.640	0.548	0.815

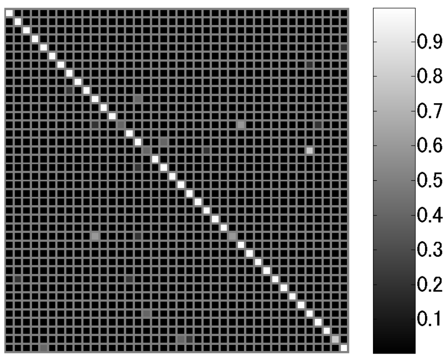

Fig. 7 Confusion matrix.

Fig. 8 The 36th person's images for training.

Fig. 9 The 17th person's images for testing.

the above methods because LDA encountered the singularity problem in our preliminary experiments. We determined all parameters in each method by using leave-one-out cross-validation. We used the ORL face database [7], the Indian face database [9] (IFD), the UMIST face database [10] and Caltech Human face (Front) dataset [11] (CHFD). Table 2 summarizes classification accuracy for each method on each dataset. The photo-retouch software achieved the highest accuracy on the IFD dataset and the proposed method achieved the highest accuracy on the other datasets.

5. Conclusion

In this paper, we proposed a non-iterative symmetric 2DLDA for supervised dimensionality reduction of matrices such as two-dimensional images. Experimental results show that the proposed method achieves comparable recognition rate with the conventional symmetric 2DLDA and is computationally efficient in the training stage of face recognition. We also compared the performance of the proposed method with that of no dimension reduction method (baseline), PCA+LDA, 2DLDA [4] and image shrinkage by a photo-retouch software on four benchmark datasets.

Acknowledgements

This work was partially supported by the Ministry of Education, Culture, Sports, Science and Technology under the Grant-in-Aid for Scientific Research (20700165).

References

[1] K. Fukunaga, Introduction to statistical pattern recognition, 2nd ed., Academic Press Professional, 1990.

[2] J. Yang, D. Zhang, X. Yong, and J. Yang, "Two-dimensional discriminant transform for face recognition," Pattern Recognit., vol.38, no.7, pp.1125–1129, 2005.

[3] J. Yang, J. Yang, A.F. Frangi, and D. Zhang, "Uncorrelated projection discriminant analysis and its application to face image feature extraction," Int. J. Pattern Recognition and Artificial Intelligence, vol.17, no.8, pp.1325–1347, 2003.

[4] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," Proc. NIPS, pp.1569–1576, 2005.

[5] K. Inoue and K. Urahama, "Non-iterative two-dimensional linear discriminant analysis," Proc. ICPR, pp.540–543, 2006.

[6] D. Luo, C. Ding, and H. Huang, "Symmetric two dimensional linear discriminant analysis (2DLDA)," Proc. CVPR, pp.2820–2827, 2009.

[7] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," Proc. 2nd IEEE Workshop on Applications of Computer Vision, 1994.

[8] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, 2nd ed., Wiley-Interscience, 2000.

[9] V. Jain and A. Mukherjee, The Indian Face Database, 2002. http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/

[10] D.B. Graham and N.M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences, vol.163, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman-Soulie and T.S. Huang, eds., pp.446–456, 1998.

[11] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," Proc. CVPR, vol.2, pp.264–271, 2003.