

Symmetric Two Dimensional Linear Discriminant Analysis (2DLDA)

Dijun Luo, Chris Ding, Heng Huang
University of Texas at Arlington
701 S. Nedderman Drive Arlington, TX 76013
dijun.luo@gmail.com, {chqding, heng}@uta.edu

Abstract

Linear discriminant analysis (LDA) has been successfully applied into computer vision and pattern recognition for effective feature extraction. High-dimensional objects such as images are usually transform as 1D vectors before the LDA transformation. Recently, two-dimension LDA (2DLDA) methods have been proposed which reduced the dimensionality of images without transforming the matrices into vectors. However, the objective function for 2DLDA remains an unresolved problem. In this paper, we (1) propose a symmetric LDA formulation which resolves the ambiguity problem, and (2) propose an effective algorithm to solve the symmetric 2DLDA objective. Experiments on UMIST, CMU PIE, and YaleB images databases show that our approach outperforms the other 2DLDA methods in terms of both classification accuracy and objective function results.

1. Introduction

Fisher linear discriminant analysis (LDA) has been successfully applied to computer vision and pattern recognition applications in the past few years. As a subspace analysis approach to learn the low-dimensional structure of high-dimensional data, LDA seeks for a set of vectors that maximized Fisher Discriminant Criterion. It simultaneously minimized the within-class scatter while maximizing the between-class scatter in the projective feature vector space. Typical transform-based methods include Fisherfaces [1] and its variations [8]. Belhumeur *et al.* discussed the comparison between Fisherface and Eigenface in [1] and also showed that LDA was better than Principle Component Analysis (PCA).

Recently several two-dimensional LDA (2DLDA) methods have been presented recently. Liu *et al.* [6], Li and Yuan [5], and Xiong *et al.* [9] formulated the image based between-class scatter matrix and within-class scatter matrix calculation. Those methods didn't convert the images into vectors so that reduced the dimensionality of image matrices.

Song *et al.* [7] and Yang *et al.* [10] used the column by column correlations to reduce the number of columns. Yang *et al.* [12] improved and gave out an algorithm to reduce the number of columns first and reduce the number of rows later. This method is an order dependent algorithm. Ye *et al.* [14] introduce an order independent 2DLDA with an iterative solution algorithm. They consider the projection of the data onto a space which is the tensor product of two vector spaces. In Inoue and Urahama's paper [4], they pointed out Ye's iterative algorithm does not necessarily increase objective function monotonically. Because one more objective function exists for 2DLDA and it can decrease when we maximize another one. Inoue and Urahama [4] proposed a method for selecting the transformation matrices which give higher value of the generalized Fisher criterion than both Yang's and Ye's 2DLDAs. They also gave out a non-iterative algorithm in which the rows and columns matrices are treated independently.

However, a fundamental problem with 2DLDA remains unresolved. For 1D vectors, the between-class and within-class scatter matrices are uniquely defined. For 2D matrices such as images, the between-class and within-class scatter matrices are not unique. There is a large number of possible choices, see Eqs. (10-14). In other words, the 2DLDA objective functions used in all previous work are incomplete.

In this paper, we present a novel approach that resolves this ambiguity problem of 2DLDA. We begin with a key observation: if the 2D matrix objects are symmetric, then the between-class and within-class scatter matrices are uniquely defined. Motivated from this observation, we propose a new data representation which (1) enforces symmetry and (2) are equivalent to the standard transforms in 2DLDA. This symmetric data representation therefore defines a unique 2DLDA objective function, which is consistent generalization from 1DLDA (see Section 3).

In Section 4, we propose an effective algorithm to solve the new 2DLDA objective function. In Section 5, we perform experiments on three well-known face images databases to evaluate the performance of other 2DLDA methods and 2DSVD (representing the methods from unsu-

pervised learning family). Our experimental results demonstrate that our approach outperforms other 2DLDA methods.

2. Review of LDA and Existing 2DLDA

2.1. Classic LDA

The classical LDA transforms the original data into a much lower dimensional space where the classification task is much easier to perform. Let the original be a set of 1D vectors: $X = (x_1, \dots, x_n) \in R^{m \times n}$, with the knowledge that they are partitioned into k pattern classes $\Pi = \{\pi_1, \dots, \pi_k\}$, with π_i contains n_i data points in the i -th class. In LDA, the transformation to the lower dimensional (sub)-space is

$$y_i = G^T x_i, \quad (1)$$

where G is the transformation to the subspace. We often write $(y_1, \dots, y_n) = G^T(x_1, \dots, x_n)$ or $Y = G^T X$. The goal of LDA is to find the G such that different classes are more separated in the transformed space and thus more easily distinguished from each other.

Define the between-class scatter matrix S_b and within-class scatter matrix S_w

$$S_b(x) = \sum_{j=1}^k n_j (m_j - m)(m_j - m)^T, \quad (2)$$

$$S_w(x) = \sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - m_j)(x_i - m_j)^T, \quad (3)$$

where $m_j = \frac{1}{n_j} \sum_{x_i \in \pi_j} x_i$ is its class mean and $m = \frac{1}{n} \sum_{i=1}^n x_i$ is the global mean. In the transformed space, S_b and S_w are transformed into

$$S_b(Y) = G^T S_b(X) G,$$

$$S_w(Y) = G^T S_w(X) G.$$

The optimization criteria of LDA is that G is chosen such that different classes are more separated from other ($\max S_b(Y)$) and each class is more compact ($\min S_w(Y)$). This leads to the standard LDA optimization objective function:

$$\max_G J(G) = \text{tr} \frac{S_b(Y)}{S_w(Y)} = \text{tr} \frac{G^T S_b(X) G}{G^T S_w(X) G}. \quad (4)$$

Note that $\text{tr}(A/B) = \text{tr}(B^{-1}A) = \text{tr}(AB^{-1})$.

2.2. 2DLDA

In 2DLDA, we deal with a set of images $X = (X_1, \dots, X_n), X_i \in \mathbb{R}^{r \times c}$. With similar intuition of classical LDA, 2DLDA tries to seek a bilinear transformation

$$Y_i = L^T X_i R, \quad (5)$$

such that different classes are more separated. The key issue is how to select the subspace L and R based on between-class and within-class scatter matrices.

We immediately see there is a fundamental ambiguity problem: There are two ways to define the within-class scatter matrix S_w

$$S_w(XX^T) = \sum_{j=1}^k \sum_{x_i \in \pi_j} (X_i - M_j)(X_i - M_j)^T, \quad (6)$$

$$S_w(X^T X) = \sum_{j=1}^k \sum_{x_i \in \pi_j} (X_i - M_j)^T (X_i - M_j), \quad (7)$$

and there two ways to define the between-class scatter matrix S_b

$$S_b(XX^T) = \sum_{j=1}^k n_j (M_j - M)(M_j - M)^T, \quad (8)$$

$$S_b(X^T X) = \sum_{j=1}^k n_j (M_j - M)^T (M_j - M). \quad (9)$$

Therefore, in the transformed space, we have corresponding

$$S_b(Y Y^T), S_b(Y^T Y),$$

$$S_w(Y Y^T), S_w(Y^T Y).$$

In general, images are not symmetric: $X_i \neq X_i^T$. Thus

$$S_b(Y Y^T) \neq S_b(Y^T Y),$$

$$S_w(Y Y^T) \neq S_w(Y^T Y).$$

For this reason, the LDA objective function is ambiguous. We have a large number of choices:

$$J_1 = \text{tr} \frac{S_b(Y Y^T)}{S_w(Y Y^T)}, \quad (10)$$

$$J_2 = \text{tr} \frac{S_b(Y^T Y)}{S_w(Y^T Y)}, \quad (11)$$

$$J_3 = \text{tr} \left[\frac{S_b(Y Y^T)}{S_w(Y Y^T)} + \frac{S_b(Y^T Y)}{S_w(Y^T Y)} \right], \quad (12)$$

$$J_4 = \text{tr} \left[\frac{S_b(Y^T Y)}{S_w(Y^T Y)} \frac{S_b(Y^T Y)}{S_w(Y^T Y)} \right], \quad (13)$$

$$J_5 = \text{tr} \frac{S_b(Y Y^T) + S_b(Y^T Y)}{S_w(Y Y^T) + S_w(Y^T Y)}, \quad (14)$$

etc. Which is the right one? Why?

3. Symmetric 2DLDA

Our main contribution in this paper is to introduce a new data representation to resolve the ambiguity problem of the existing 2DLDA.

Our approach is motivated by a key observation: if the images is symmetric, i.e., $X_i = X_i^T$, then

$$S_w(XX^T) = S_w(X^T X),$$

$$S_b(XX^T) = S_b(X^T X).$$

Then the ambiguity problem in Eqs. (10-14) will be resolved. For this reason, we introduce a new data representation (bilinear transformation).

3.1. Symmetric Bilinear Transformation

Here we introduce a symmetric bilinear transformation:

$$\begin{pmatrix} Y_i & Y_i^T \\ X_i & X_i^T \end{pmatrix} = \Gamma^T \begin{pmatrix} X_i \\ X_i^T \end{pmatrix} \Gamma, \quad \Gamma = \begin{pmatrix} L \\ R \end{pmatrix}. \quad (15)$$

Using this symmetric transformation, the scatter matrices become unique and the ambiguity problem of existing 2DLDA is resolved.

We now show that the bilinear transformation of Eq.(15) is equivalent to the linear transformation of Eq.(5). We have the following equalities:

$$\Gamma^T \begin{pmatrix} X_i \\ X_i^T \end{pmatrix} \Gamma = \begin{pmatrix} L^T X_i R & R^T X_i^T L \end{pmatrix} \quad (16)$$

Thus $Y_i = L^T X_i R$ holds identically. We also have

$$\left\| \begin{pmatrix} X_i \\ X_i^T \end{pmatrix} - \Gamma \begin{pmatrix} Y_i \\ Y_i^T \end{pmatrix} \Gamma^T \right\|^2 = 2\|X_i - LY_i R^T\|^2 \quad (17)$$

Therefore, optimizations using (L, R) are equivalent to optimizations using Γ .

3.2. Symmetric 2DLDA

Using the symmetric bilinear transformation of Eq.(15) we have the following theorem:

Theorem 1: The unique LDA objective function for 2DLDA is

$$J_{2DLDA} = \text{tr} \frac{S_b(YY^T)}{S_w(YY^T)} = \text{tr} \frac{S_b(Y^T Y)}{S_w(Y^T Y)} \quad (18)$$

$$= \text{tr} \left(\frac{R^T S_b^L R}{R^T S_w^L R} + \frac{L^T S_b^R L}{L^T S_w^R L} \right) \quad (19)$$

where

$$S_w^R = \sum_{j=1}^k \sum_{X_i \in \pi_j} (X_i - M_j) R R^T (X_i - M_j)^T, \quad (20)$$

$$S_w^L = \sum_{j=1}^k \sum_{X_i \in \pi_j} (X_i - M_j)^T L L^T (X_i - M_j). \quad (21)$$

$$S_b^R = \sum_{j=1}^k n_j (M_j - M) R R^T (M_j - M)^T, \quad (22)$$

$$S_b^L = \sum_{j=1}^k n_j (M_j - M)^T L L^T (M_j - M). \quad (23)$$

The proof is rather lengthy and is given in Appendix A.

4. Related Work on 2DLDA

With the notation introduced in Theorem 1, it is now convenient to discuss earlier work relating to 2DLDA. Yang *et al.* [10] first propose to use

$$J'_2 = \text{tr} \frac{S_b(Y^T Y)}{S_w(Y^T Y)} = \text{tr} \frac{L^T S_b L}{L^T S_w L}.$$

This idea is the further development of 2DPCA [11]. Ye *et al.* [14] uses

$$J'_1 = \text{tr} \frac{S_b(YY^T)}{S_w(YY^T)} = \text{tr} \frac{R^T S_b^L R}{R^T S_w^L R}, \quad (24)$$

$$J'_2 = \text{tr} \frac{S_b(Y^T Y)}{S_w(Y^T Y)} = \text{tr} \frac{L^T S_b^R L}{L^T S_w^R L}. \quad (25)$$

This idea is the further development of GLRAM [13]. Inoue *et al.* [4] emphasized the inconsistency of the optimization approach of Ye *et al.* [14] and provide several alternatives. This issue is explained more carefully below.

4.1. Inconsistency of Independent Alternating Optimization Approach

In both Ye's [14] and Inoue *et al.*'s [4] approaches, they optimize the two objectives Eqs.(24,25)

$$\max_R J'_1(R, L), \quad \max_L J'_2(R, L),$$

in an independent and alternating way. They obtain R by maximizing J'_1 (ignoring J'_2) and then obtain L by maximizing J'_2 (ignoring J'_1). A fundamental drawback of this approach is the inconsistency of optimization. When maximizing J'_1 (without considering J'_2), J'_2 could be and is often decreased. Similarly, when maximizing J'_2 (without considering J'_1), J'_1 could be and is often decreased.

Inoue and Urahama [4] noticed and emphasized this inconsistency. They provided some quick fixes, but did not provide a satisfactory solution.

A satisfactory solution to this problem has two technical challenges. First, when maximizing J'_1 , we need to take into account of J'_2 . But in order to do so, we need to know how J'_1 and J'_2 should be combined. One choice is simple

additive combination: $J = J'_1 + J'_2$. Another choice should be

$$J = \text{tr} \frac{R^T S_b^L R + L^T S_b^R L}{R^T S_w^L R + L^T S_w^R L}, \quad (26)$$

There are many other choices. Clearly, this requires a rigorous definition of the objective function for 2DLDA. Our symmetric LDA is motivated by this challenge.

The second challenge is how to optimize the resulting objective function. The solution for $\max_R J'_1$ is simply given by the eigenvectors of $(S_w^L)^{-1} S_b^L$, as in standard LDA. For the objectives such as Eq.(26) or Eq.(19) the functional dependency on R, L via Eqs.(20,21,33,23) are rather complicated. The required algorithm could be far more complicated than the eigenvector solution in the independent-alternating approach. Our work resolves this challenge by developing the derivative based approach as described in the next section.

5. Computational Algorithms for Symmetric 2DLDA

The objective of Symmetric 2DLDA in Eq. (19) is no longer the trace of a single ratio two scatter matrices. As discussed in §4.1, this objective can not be solved directly by calculating eigenvectors (as in standard LDA). Fortunately we are able to develop an efficient algorithm to solve this problem by using gradient-ascend approach. This approach requires the derivatives of the objective. The derivatives of matrix functions be worked out using basic matrix algebra such as in the book [2]. We skip the rather lengthy derivation and present the results in the following Lemma:

Lemma 1. Let $P_L = L^T S_b^L L$, $Q_L = L^T S_w^R L$, $P_R = R^T S_b^R R$, and $Q_R = R^T S_w^L R$. The derivatives of the objective function J_{2DLDA} of Eq.(19) are the following. For $\frac{\partial J}{\partial R}$, we have

$$\frac{\partial}{\partial R} \text{tr} \frac{R^T S_b^L R}{R^T S_w^L R} = 2S_b^L R Q_R^{-1} - 2S_w^L R Q_R^{-1} P Q_R^{-1}, \quad (27)$$

and

$$\begin{aligned} & \frac{\partial}{\partial R} \text{tr} \frac{L^T S_b^R L}{L^T S_w^R L} \\ &= 2 \sum_{k=1}^K \sum_{A_i \in \pi_k} (A_i - M_k)^T L Q_L^{-1} L^T (A_i - M_k) R \\ & \quad - 2 \sum_{k=1}^K (M_k - M)^T L Q_L^{-1} P_L Q_L^{-1} L^T (M_k - M) R. \end{aligned} \quad (28)$$

For $\frac{\partial J}{\partial L}$, we have

$$\frac{\partial}{\partial L} \text{tr} \frac{L^T S_b^R L}{L^T S_w^R L} = 2S_b^R L Q_L^{-1} - 2S_w^R L Q_L^{-1} P_L Q_L^{-1}, \quad (29)$$

$$\begin{aligned} & \frac{\partial}{\partial L} \text{tr} \frac{R^T S_b^L R}{R^T S_w^L R} \\ &= 2 \sum_{k=1}^K \sum_{A_i \in \pi_k} (A_i - M_k) R Q_R^{-1} R^T (A_i - M_k)^T L \\ & \quad - 2 \sum_{k=1}^K (M_k - M) R Q_R^{-1} P_R Q_R^{-1} R^T (M_k - M)^T L. \end{aligned} \quad (30)$$

We present the results in four parts to make them more transparent.

Algorithm 1 Solving Symmetric 2DLDA using gradient.

Input

- A set of images $\{X_i\}_{i=1}^n$ and their class labels.
- Initial L_0, R_0
- Frequency c for orthogonalization

Initialization

- $L \leftarrow L_0, R \leftarrow R_0$,
- Calculate $M_k, k = 1, 2, \dots, K$ and M ,
- $t \leftarrow 0$

Do

Calculate $S_w^L, S_b^L, S_w^R, S_b^R$.
 $R \leftarrow R + \delta \frac{\partial J}{\partial R}$
 $L \leftarrow L + \delta \frac{\partial J}{\partial L}$
 $t \leftarrow t + 1$
 if $(t \bmod c) = 0$
 $R \leftarrow \text{eigenvectors of } (S_w^L)^{-1} S_b^L$
 $L \leftarrow \text{eigenvectors of } (S_w^R)^{-1} S_b^R$
 endif

Until stopping criteria is satisfied.

Output L, R

Using the explicit formulations of the gradient above, we develop Algorithm 1 in Table. The step size δ is set to

$$\delta = 0.02 \langle |R| \rangle / \langle |\partial J / \partial R| \rangle$$

where $\langle |A| \rangle = \sum_{i=1}^m \sum_{j=1}^n |A_{ij}| / nm$.

The parameter c is to control the frequency for $R^T S_b^L R, R^T S_w^L R, L^T S_b^R L$, and $L^T S_w^R L$ to be nearly diagonal. This implies that in the transformed space, S_b^L, S_w^L, S_b^R , and S_w^R are nearly diagonal, i.e., data are simultaneously uncorrelated w.r.t. $S_t^L = S_b^L + S_w^L$ and are uncorrelated w.r.t. $S_t^R = S_b^R + S_w^R$. In the experiments, we set $c = 3$.

6. Experimental Results

We compare performance of symmetric 2DLDA to the method of Ye *et al.* [13] on three standard face images datasets.

6.1. UMIST

The first benchmark UMIST faces is for multi-view face recognition, which is challenging in computer vision because the variations between the images of the same face in viewing direction are almost always larger than image variations in face identity. A robust face recognition system should be able to recognize the person even though the testing image and training images have quite different poses. This dataset contains 20 persons with 18 images for each. All these images of UMIST database are cropped and re-sized into 28×23 images.

In this dataset, we first visualize the discriminant capability of subspace of our approach. Since typically the distance of two two-dimension object X_i is evaluate in the L, R subspace: $Y_i = L^T X_i R$, we want to see the major components of subspace in two-dimension plot, which are shown Figure 1. In this figure, we compare PCA, 2DSVD, Ye's 2DLDA, and our method. We use 3 (shown in first two columns), 6 (shown in the third and fourth columns), and 9 (shown in the last two columns) images per person to train the data across all the persons and pick up three persons, 18 images for each, to plot in the selected subspace. For PCA, we plot first principle component (as x-axis) versus second principle (y-axis) in (a) columns and first principle component (x-axis) versus third principle component (y-axis) in (b) columns. For 2DSVD, Ye's 2DLDA, and our method, we use $l_1^T X_i r_1$ as x-axis and $l_1^T X_i r_2$ as y-axis to plot image X_i in (a) columns and $l_1^T X_i r_1$ as x-axis and $l_2^T X_i r_1$ as y-axis to plot in (b) columns, where l_1, l_2, r_1, r_2 are the first two columns of L and R respectively. In each plot, blue points denote training dataset which are used to seek the subspace, and red points are testing images. From the plots, we can see that unsupervised methods (PCA and 2DSVD) are not suitable to generate a discriminative subspace. Instead, supervised methods (Ye's 2DLDA and our method) are much better. We can also see that the more training images we use, the more discriminative the subspace is. The plots also indicate that our method enjoy much more discriminative capability than Ye's 2DLDA since our method directly solve the 2DLDA objective.

Instead of using cross validation, we systematically evaluate the classification accuracy and objective function value using a cross validation-like scheme which is much more challenging: The images are equally divided into 5 parts. Then pick one part as the training dataset and the rest 4 parts as testing dataset. Repeat it 5 times by picking up different parts as training dataset. This is scheme only uses 20% of the images to train and 80% to test. We set $k = 5, 8, 12$ and compare the classification with subspace of 2DSVD, Ye's 2DLDA and original space (direct Nearest Neighbor classification). We plot the average Nearest Neighbor classification accuracy and the corresponding 2DLDA objective defined in Eq. (19). For original space and 2DSVD, we

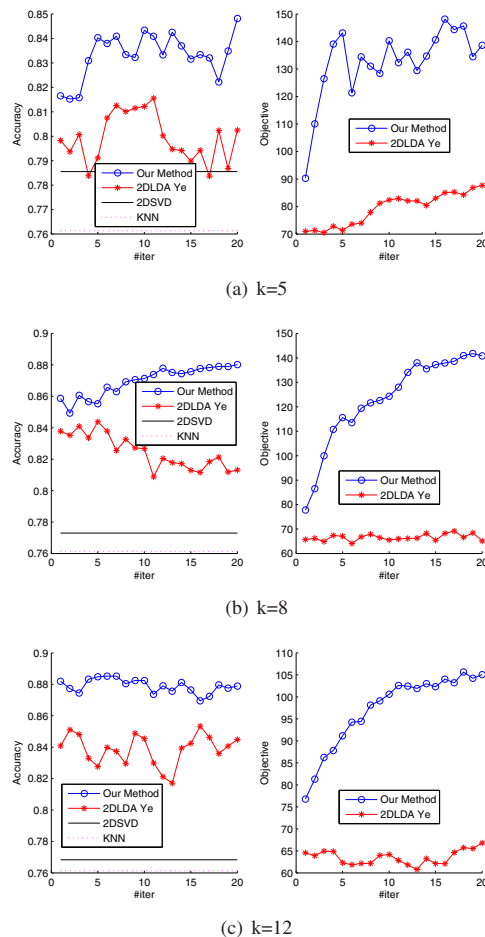


Figure 2. Classification accuracy and objective function value comparison on UMIST.

omit the objective function value. The figure shows that our method is significantly better in term of objective function since we are explicitly solving the objective function. Our method also generate better subspace (3%-5% better in classification accuracy), see Figure 2(a), 2(b), and 2(c). We also summarize the best results in classification and objective in Table 1.

6.2. CMU PIE Dataset

The second dataset is CMU PIE (Face Pose, Illumination, and Expression) face database which contains 68 subjects with 41,368 face images as a whole. Preprocessing to locate the faces was applied. Original images were normalized (in scale and orientation) such that the two eyes were aligned at the same position. Then, the facial areas were cropped into the final images for matching. The size of each cropped image is 64×64 pixels, with 256 grey levels per pixel. No further preprocessing is done. In our

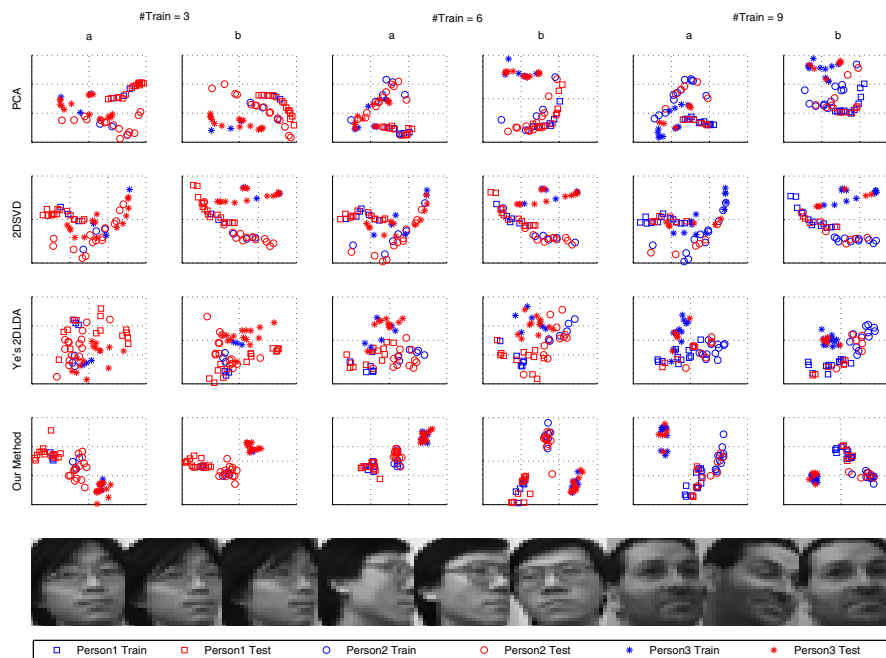


Figure 1. Visualization of discriminative capability of PCA, 2DSVD, Ye's 2DLDA and our method on UMIST dataset. Shown are 3 cases for training the subspace using 3, 6, and 9 images per person.

UMIST	k=5		k=8		k=12	
	Acc	Obj	Acc	Obj	Acc	Obj
KNN	0.764	-	0.764	-	0.764	-
Y2DLDA	0.815	87.6	0.764	69.1	0.853	66.8
S2DLDA	0.848	148.1	0.880	141.7	0.885	105.6
PIE	k=8		k=12		k=16	
	Acc	Obj	Acc	Obj	Acc	Obj
KNN	0.537	-	0.537	-	0.537	-
Y2DLDA	0.623	459.8	0.678	366.1	0.691	327.2
S2DLDA	0.651	514.4	0.678	411.7	0.696	339.7
YaleB	k=8		k=12		k=16	
	Acc	Obj	Acc	Obj	Acc	Obj
KNN	0.572	-	0.572	-	0.572	-
Y2DLDA	0.766	418.8	0.804	348.3	0.791	303.2
S2DLDA	0.768	446.0	0.799	393.2	0.796	336.9

Table 1. Summarization of best classification accuracy and 2DLDA objective function value in UMIST, PIE, and YaleB face datasets. Methods: KNN in original space, Ye's 2DLDA (Y2DLDA), and Symmetric 2DLDA (S2DLDA).

experiment, we randomly pick 10 different combinations of pose, face expression, and illumination condition. Finally we have $68 \times 10 = 680$ images.

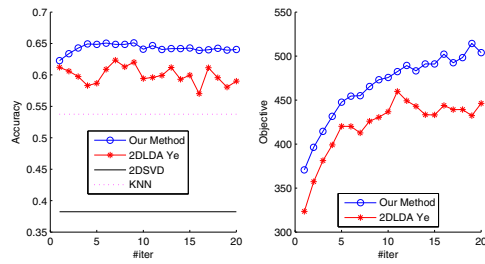
We use the same scheme discussed above in this dataset except we set $k = 8, 12, 16$ since the image size is larger than UMIST. The results can be found in Figure 3(a), 3(b),

and 3(c). In this dataset, we generate a reasonable better result in classification (about 5% better) when $k = 8$ and slightly better in the other cases. Overall, the classification accuracy is much more stable. Our method is also always outperforms Ye's method in 2DLDA objective (roughly 30% better). The summarization of the results are shown in Table 1.

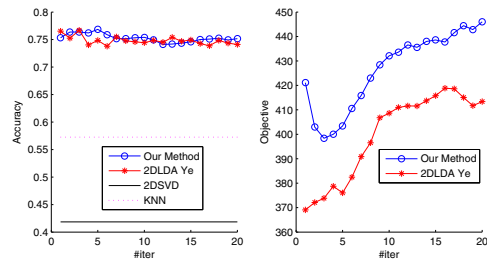
6.3. YaleB Dataset

The final face images benchmark used in our experiment is the combination of extended and original Yale database B [3]. These two databases contain single light source images of 38 subjects (10 subjects in original database and 28 subjects in extended one) under 576 viewing conditions (9 poses \times 64 illumination conditions). Thus, for each subject, we got 576 images under different lighting conditions. The facial areas were cropped into the final images for matching [3], including: 1) preprocessing to locate the faces was applied; 2) original images were normalized (in scale and orientation) such that the two eyes were aligned at the same position. The size of each cropped image in our experiments is 192×168 pixels, with 256 gray levels per pixel. We randomly pick up 20 images for each person and also subsample the images down to 48×42 .

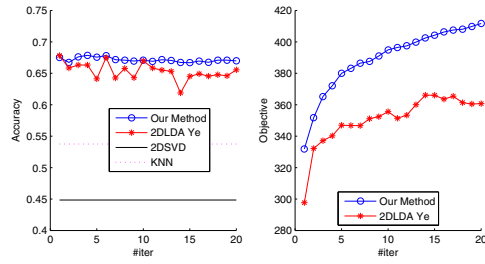
In this dataset, the solutions of our method is till stably better than Ye's 2DLDA in 2DLDA objective (see the right



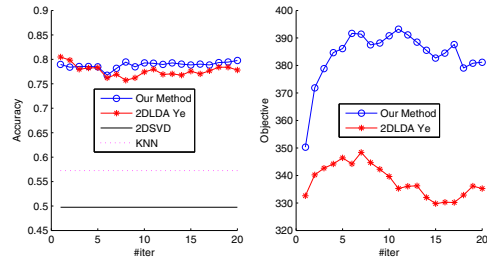
(a) k=8



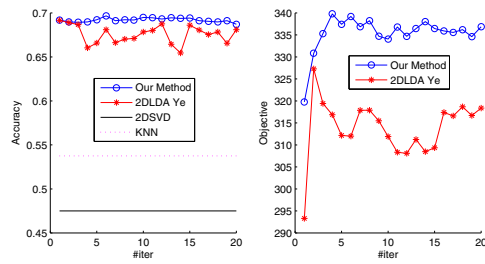
(a) k=8



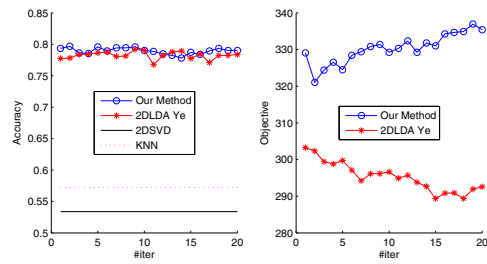
(b) k=12



(b) k=12



(c) k=16



(c) k=16

Figure 3. Classification accuracy and objective function value comparison on PIE.

Figure 4. Classification accuracy and objective function value comparison on YaleB.

part of Figure 4(a), 4(b), and 4(c)). And the classification accuracy is close to each other. In this YaleB dataset, the faces are taken in 64 different illumination conditions and we only pick up 20% of the data to train. Thus both Ye's 2DLDA and Symmetric 2DLDA are easy to overfit in the training set. However, Symmetric 2DLDA is still slightly better and more stable overall. We can also find the summarization of the results in Table 1.

7. Conclusion

This paper first points out the LDA ambiguity problem that was existing in the previous 2DLDA objective functions. After that, we present a novel symmetric data representation and resolve the ambiguity problem using the complete objective function for 2DLDA. Our 2DLDA (Symmetric 2DLDA) is the first one to define 2DLDA using ratio trace formulation with complete between-class scatter matrices and within-class scatter matrices. An effective

computational algorithm is also given out to solve the complete objective function of Symmetric 2DLDA. The experimental results show that our method achieves better face recognition accuracy than other 2DLDA methods. The new methodology improves the feature extraction and pattern classification in computer vision and machine learning related research.

2DLDA is a kind of bilinear discriminant analysis methods. Tensor based LDA can be generated using the approach presented in this paper. Our future work will exploit the multi-linear discriminant analysis problem.

Acknowledgement

This work is partially supported by NSF DMS-0844497, NSF CCF-0830780, and University of Texas Stars Award.

Appendix A

Proof of Theorem 1: We write the within-class scatter matrices $S_w(YY)$ as follows:

$$S_w(YY) = \sum_{j=1}^n \sum_{Y_i \in \pi_j} \begin{pmatrix} (Y_i - M_j^Y) & (Y_i - M_j^Y)^T \end{pmatrix} \cdot \begin{pmatrix} (Y_i - M_j^Y) & (Y_i - M_j^Y)^T \end{pmatrix} \quad (31)$$

Substituting Eq. (15) into Eq. (31), we get:

$$\begin{aligned} S_w(YY) &= \sum_{j=1}^n \sum_{X_i \in \pi_j} \begin{pmatrix} R & L \end{pmatrix}^T \begin{pmatrix} (X_i - M_j)^T & (X_i - M_j) \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix}^T \begin{pmatrix} (X_i - M_j)^T & (X_i - M_j) \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix} \\ &= \sum_{j=1}^n \sum_{X_i \in \pi_j} \begin{pmatrix} R & L \end{pmatrix}^T \text{diag} \begin{pmatrix} (X_i - M_j) R R^T (X_i - M_j)^T \\ (X_i - M_j)^T L L^T (X_i - M_j) \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix} \\ &= \begin{pmatrix} R & L \end{pmatrix}^T \begin{pmatrix} S_w^R & \\ & S_w^L \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix}. \end{aligned} \quad (32)$$

Similarly, we can write the between-class scatter matrices as:

$$S_b(YY) = \begin{pmatrix} R & L \end{pmatrix}^T \begin{pmatrix} S_b^R & \\ & S_b^L \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix}. \quad (33)$$

The standard LDA objective function is

$$\begin{aligned} J_{2DLDA}(L, R) &= \frac{\text{tr} S_b(YY)}{\text{tr} S_w(YY)} = \frac{\begin{pmatrix} R & L \end{pmatrix}^T \begin{pmatrix} S_b^R & \\ & S_b^L \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix}}{\begin{pmatrix} R & L \end{pmatrix}^T \begin{pmatrix} S_w^R & \\ & S_w^L \end{pmatrix} \begin{pmatrix} R & L \end{pmatrix}} = \text{tr} \begin{pmatrix} R^T S_b^L R & 0 \\ 0 & L^T S_b^R L \end{pmatrix} \\ &= \text{tr} \begin{pmatrix} (R^T S_w^L R)^{-1} & 0 \\ 0 & (L^T S_w^R L)^{-1} \end{pmatrix} \cdot \begin{pmatrix} R^T S_b^L R & 0 \\ 0 & L^T S_b^R L \end{pmatrix} = \text{tr} \frac{R^T S_b^L R}{R^T S_w^L R} + \text{tr} \frac{L^T S_b^R L}{L^T S_w^R L} \end{aligned}$$

References

- [1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.
- [2] K. Fukunaga. Introduction to statistical pattern recognition. *Academic Press Professional, 2nd edition*, 1990.
- [3] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.
- [4] K. Inoue and K. Urahama. Non-iterative two-dimensional linear discriminant analysis. *Proceedings of the 18th International Conference on Pattern Recognition*, 2:540–543, 2006.
- [5] M. Li and B. Yuan. 2d-lda: A novel statistical linear discriminant analysis for image matrix. *Pattern Recognition Letters*, 26(5):527–532, 2005.
- [6] K. Liu, Y. Cheng, and J. Yang. Algebraic feature extraction for image recognition based on an optimal discriminant criterion. *Pattern Recognition*, 26(6):903–911, 1993.
- [7] F. Song, S. Liu, and J. Yang. Orthogonalized fisher discriminant. *Pattern Recognition*, 38(2):311–313, 2005.
- [8] D. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):831–836, 1996.
- [9] H. Xiong, M. Swamy, and M. Ahmad. Two-dimensional fld for face recognition. *Pattern Recognition*, 38(7):1121–1124, 2005.
- [10] J. Yang, J. Yang, A. F. Frangi, and D. Zhang. Uncorrelated projection discriminant analysis and its application to face image feature extraction. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(8):1325–1347, 2003.
- [11] J. Yang, D. Zhang, A. F. Frangi, and J. Yang. Twodimensional pca: A new approach to appearancebased face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1), 2004.
- [12] J. Yang, D. Zhang, X. Yong, and J. Yang. Two-dimensional linear discriminant transform for face recognition. *Pattern Recognition*, 38(7):1125–1129, 2005.
- [13] J. Ye. Generalized low rank approximations of matrices. *International Conference on Machine Learning*, 2004.
- [14] J. Ye, R. Janardan, and Q. Li. Two-dimensional linear discriminant analysis. *Advances in Neural Information Processing Systems (NIPS 2004)*, 17:1569–1576, 2004.