ORIGINAL ARTICLE

From NLDA to LDA/GSVD: a modified NLDA algorithm

Jun Yin · Zhong Jin

Abstract Linear discriminant analysis (LDA) often encounters small sample size (SSS) problem for high-dimensional data. Null space linear discriminant analysis (NLDA) and linear discriminant analysis based on generalized singular value decomposition (LDA/GSVD) are two popular methods that can solve SSS problem of LDA. In this paper, we present the relation between NLDA and LDA/GSVD under a condition and at the same time propose a modified NLDA (MNLDA) algorithm which has the same discriminating power as LDA/GSVD and is more efficient. In addition, we compare the discriminating capability of NLDA and MNLDA and present our interpretation about this. Experimental results on ORL, FERET, Yale face databases, and the PolyU FKP database support our viewpoints.

Keywords Linear discriminant analysis · Small sample size · Null space linear discriminant analysis · Linear discriminant analysis based on generalized singular value decomposition

1 Introduction

In pattern recognition, the data often lie in a high-dimensional space. In order to analyze the data, we need to reduce the dimensionality and find the most important features. Feature extraction is an essential technique to deal with this problem, and it has attracted much attention in computer vision and pattern recognition because of its great importance. In the past decades, numerous feature extraction methods have been proposed [1–12]. Linear discriminant analysis (LDA) [13] is one of the most popular feature extraction and dimensionality reduction methods, which has been widely applied in high-dimensional pattern recognition problems [14–16]. It seeks to find the optimal projection vector by maximizing between-class scatter and minimizing within-class scatter simultaneously. The criterion of LDA is maximizing

$$J_F(W) = \frac{J_b(W)}{J_w(W)} = \frac{\text{trace}(W^T S_b W)}{\text{trace}(W^T S_w W)}, \tag{1}$$

where S_b is the between-class scatter matrix, S_w is the within-class scatter matrix, and W is the projection matrix composed by optimal projection vectors. Suppose that we have M samples of c classes, S_b, S_w and the global scatter matrix S_t can be defined by

$$S_b = \frac{1}{M} \sum_{i=1}^{c} l_i (m_i - m)(m_i - m)^T, \tag{2}$$

$$S_w = \frac{1}{M} \sum_{i=1}^{c} \sum_{j=1}^{l_i} (x_{ij} - m_i)(x_{ij} - m_i)^T \tag{3}$$

and

$$S_t = \frac{1}{M} \sum_{i=1}^{c} \sum_{j=1}^{l_i} (x_{ij} - m)(x_{ij} - m)^T, \tag{4}$$

where x_{ij} is the training sample j in class i, m is the global centroid, m_i ($i = 1, 2, …, c$) is centroid of class i, and l_i ($i = 1, 2, …, c$) is the number of training samples in class i.

From formula (3), it can be easily known that the rank of S_w is not larger than the maximum of the dimensionality of

J. Yin (✉) · Z. Jin
School of Computer Science and Technology,
Nanjing University of Science and Technology,
Nanjing 210094, China
e-mail: yinjun8429@163.com

samples and the number of training samples. If the dimensionality of samples is too large, S_w will be singular. At this time, classical LDA cannot be performed successfully. This is so-called small sample size (SSS) problem, namely the number of training samples is relatively small compared to the dimensionality of samples. In order to solve this problem, several approaches have been proposed in the past few years. Among them, regularized LDA (RLDA) [17], PCA (principal component analysis) plus LDA (FDA) [14], null space LDA (NLDA) [18, 19], direct LDA (DLDA) [20], complete LDA (CLDA) [21, 22], inverse Fisher discriminant analysis (IFDA) [23], and LDA based on generalized singular value decomposition (LDA/GSVD) [24, 25] are the most well-known methods. RLDA makes S_w nonsingular by adding a diagonal matrix αI ($\alpha > 0$) to S_w. In FDA, PCA [26] is performed first for dimensionality reduction to make S_w nonsingular before performing LDA. NLDA assumes that the null space of S_w contains the most useful discriminative information and seeks projection vectors in the null space of S_w. DLDA searches discriminative information in the range space of S_b and solves the Fisher criterion (1) by diagonalizing S_b before diagonalizing S_w. CLDA extracts features separately from the null space of S_w and the range space of S_w. In IFDA, a restriction is added in PCA procedure and introduces inverse Fisher criterion. LDA/GSVD uses generalized singular value decomposition (GSVD) to solve singular problem in classical LDA.

We analyze NLDA with LDA/GSVD in theory and show their relationship under a condition which always holds for high-dimensional data. It can be found from the analysis that there is only a little difference between NLDA and LDA/GSVD. If we make a modification for NLDA, it will become equivalent to LDA/GSVD. As we know, the disadvantage of LDA/GSVD is its high computational time and NLDA is much more efficient. Therefore, we propose a modified NLDA (MNLDA) algorithm in this paper. If the given condition is satisfied, MNLDA has the same discriminating power as LDA/GSVD but is more efficient. In addition, we compare the discriminating power of NLDA and MNLDA. Sometimes NLDA outperforms MNLDA and MNLDA performs better sometimes. This mainly depends on the characteristic of the data. We give our interpretation about this according to the experiments.

The rest of the paper is organized as follows: Sect. 2 describes NLDA and LDA/GSVD. We analyze the relation between NLDA and LDA/GSVD in Sect. 3. Section 4 proposes MNLDA algorithm and presents the comparison of MNLDA, LDA/GSVD, and NLDA. Section 5 describes experiments on ORL, FERET, Yale face databases, and the Hong Kong polytechnic university finger-knuckle-print (PolyU FKP) database. Our conclusions are provided in Sect. 6 at last.

2 NLDA and LDA/GSVD

2.1 NLDA

Chen et al. [18] proved that the null space of S_w contains the most discriminative information and proposed NLDA algorithm which searches projection vectors in the null space of S_w. The criterion of NLDA is defined by

$$W = \arg \max_{W^T S_w W = 0} \text{trace}(W^T S_b W). \tag{5}$$

In order to improve the efficiency of the algorithm, the authors used a pixel grouping method to reduce the dimensionality of samples before performing NLDA in [18]. Based on the observation that the null space of S_t does not contain discriminative information, Huang et al. [27] removed the null space of S_t first and then perform NLDA.

NLDA algorithm is summarized as follows:

1. Calculate S_t by (4). Work out S_t's Eigen vectors p_1, p_2, ..., p_d corresponding to nonzero Eigen values.
2. Calculate S_b and S_w by (2) (3). Let $P = (p_1, p_2, ..., p_d)$. We get the new scatter matrices $\tilde{S}_b = P^T S_b P$ and $\tilde{S}_w = P^T S_w P$ after the removal of the null space of S_t.
3. Work out $\tilde{S}_w$'s Eigen vectors $q_1, q_2, ..., q_s$ corresponding to zero Eigen values and project the data into the null space of $\tilde{S}_w$. Now the between-class scatter matrix becomes $\hat{S}_b = Q^T \tilde{S}_b Q$, where $Q = (q_1, q_2, ..., q_s)$.
4. Work out $\hat{S}_b$'s Eigen vectors $r_1, r_2, ..., r_f$ corresponding to nonzero Eigen values. Let $R = (r_1, r_2, ..., r_f)$. We have the projection matrix of NLDA $G_n = PQR$.
5. For a sample x, the new NLDA feature is $y = G_n^T x$.

2.2 LDA/GSVD

Howland et al. [24] applied GSVD due to Paige et al. [28] to solve the SSS problem of LDA. For the n-dimensional data, we define

$$H_b = \left[\frac{\sqrt{l_1}}{\sqrt{M}}(m_1 - m), \frac{\sqrt{l_2}}{\sqrt{M}}(m_2 - m), ..., \frac{\sqrt{l_c}}{\sqrt{M}}(m_c - m) \right] \tag{6}$$

and

$$H_w = \frac{1}{\sqrt{M}}[X_1 - m_1 e_1, X_2 - m_2 e_2, ..., X_c - m_c e_c], \tag{7}$$

where $X_i = [x_{i1}, x_{i2}, ..., x_{il_i}] \in R^{n \times l_i}$ is the training sample matrix of class i and $e_i = [1, 1, ..., 1] \in R^{1 \times l_i}$. Then we have $S_b = H_b H_b^T$ and $S_w = H_w H_w^T$. Through applying GSVD to H_b^T and H_w^T, we obtain

$$U_b^T H_b^T V = \left[\overbrace{\Gamma_b}^{d}, \overbrace{0}^{n-d} \right] \tag{8}$$

and

$$U_w^T H_w^T V = \left[\overbrace{\Gamma_w}^{d}, \overbrace{0}^{n-d} \right], \tag{9}$$

where U_b and U_w are orthogonal, V is nonsingular, $\Gamma_b^T \Gamma_b$ and $\Gamma_w^T \Gamma_w$ are diagonal matrices which satisfy $\Gamma_b^T \Gamma_b + \Gamma_w^T \Gamma_w = I$ and d is the rank of $\begin{pmatrix} H_b^T \\ H_w^T \end{pmatrix}$. It follows from (8) (9) that

$$(U_b^T H_b^T V)^T U_b^T H_b^T V = V^T H_b H_b^T V = V^T S_b V$$
$$= [\Gamma_b, 0]^T [\Gamma_b, 0] = \begin{pmatrix} \Gamma_b^T \Gamma_b & 0 \\ 0 & 0 \end{pmatrix} \tag{10}$$

and

$$(U_w^T H_w^T V)^T U_w^T H_w^T V = V^T H_w H_w^T V = V^T S_w V$$
$$= [\Gamma_w, 0]^T [\Gamma_w, 0] = \begin{pmatrix} \Gamma_w^T \Gamma_w & 0 \\ 0 & 0 \end{pmatrix}. \tag{11}$$

Because $\Gamma_b^T \Gamma_b$ and $\Gamma_w^T \Gamma_w$ are diagonal and $\Gamma_b^T \Gamma_b + \Gamma_w^T \Gamma_w = I$, we have

$$V^T S_b V = \begin{pmatrix} I_j & & & \\ & \Lambda_k & & \\ & & 0_{d-k-j} & \\ & & & 0_{n-d} \end{pmatrix} \tag{12}$$

and

$$V^T S_w V = \begin{pmatrix} 0_j & & & \\ & \Sigma_k & & \\ & & I_{d-k-j} & \\ & & & 0_{n-d} \end{pmatrix}, \tag{13}$$

where $\Lambda_k \succ 0$ and $\Sigma_k \succ 0$ are diagonal, $\Lambda_k + \Sigma_k = I_k$ and the subscripts in I, 0, Λ, and Σ represent the size of the square matrices.

From (12) (13), it can be seen that V diagonalizes S_b and S_w simultaneously. Then the first $j + k$ columns of V which form the range space of S_b are selected as projection vectors. According to $\text{rank}(S_b) = j + k \leq c - 1$, the projection matrix of LDA/GSVD is often constructed by the first $c - 1$ columns of V.

The algorithm of LDA/GSVD proposed in [25] is summarized as follows:

1. Compute H_b and H_w by (6) (7).
2. Compute the singular value decomposition (SVD) of $H = \begin{pmatrix} H_b^T \\ H_w^T \end{pmatrix} \in R^{(c+M) \times n}$, which is $X^T H Y = \begin{pmatrix} Z & 0 \\ 0 & 0 \end{pmatrix}$.
3. Let $d = \text{rank}(H)$ and compute the SVD of $X(1{:}c, 1{:}d)$, which is $U_b^T X(1{:}c, 1{:}d) A = \Gamma_b$.
4. Compute $V = Y \begin{pmatrix} Z^{-1} A & 0 \\ 0 & I \end{pmatrix}$ and let $G_g = V(:, 1{:}c-1)$
5. For a sample x, the new LDA/GSVD feature is $y = G_g^T x$.

3 From NLDA to LDA/GSVD: their relation under a condition

NLDA and LDA/GSVD solves SSS problem of LDA by different ways. They seem to have no connection. In this section, through analysis we will present the relation between NLDA and LDA/GSVD under a condition. The condition usually holds for high-dimensional data.

3.1 A condition

Assume that the columns of training sample matrix $X = [X_1, X_2, \ldots, X_c]$ are linearly independent, where $X_i = [x_{i1}, x_{i2}, \ldots, x_{il_i}] \in R^{n \times l_i}$ is the training sample matrix of class i. As present in [19], under this assumption the following condition always holds:

$$\text{rank}(S_t) = M - 1, \text{rank}(S_b) = c - 1, \text{rank}(S_w) = M - c. \tag{14}$$

Because independence is often holds when the data have high dimensionality, condition (14) holds for high-dimensional data in many cases. While condition (14) is satisfied, we have

$$\text{rank}(S_b) + \text{rank}(S_w) = \text{rank}(S_t). \tag{15}$$

In the following analysis, we suppose that condition (14) is always satisfied.

3.2 Relation between NLDA and LDA/GSVD

From (12) (13), we know

$$\text{rank}(S_b) = \text{rank}(V^T S_b V) = j + k, \tag{16}$$

$$\text{rank}(S_w) = \text{rank}(V^T S_w V) = d - k - j + k = d - j \tag{17}$$

and

$$V^T S_t V = V^T S_b V + V^T S_w V$$

$$= \begin{pmatrix} I_j & & & \\ & \Lambda_k + \Sigma_k & & \\ & & I_{d-j-k} & \\ & & & 0_{n-d} \end{pmatrix} \quad (18)$$

$$= \begin{pmatrix} I_d & \\ & 0_{n-d} \end{pmatrix}.$$

It follows from (18) that

$$\mathrm{rank}(S_t) = \mathrm{rank}(V^T S_t V) = d. \quad (19)$$

If condition (14) is satisfied, from (15) (16) (17) we get

$$d = \mathrm{rank}(S_t) = \mathrm{rank}(S_b) + \mathrm{rank}(S_w)$$
$$= j + k + d - j = d + k. \quad (20)$$

It follows from (20) that $k = 0$. Then (12) (13) become

$$V^T S_b V = \begin{pmatrix} I_j & & \\ & 0_{d-j} & \\ & & 0_{n-d} \end{pmatrix} \quad (21)$$

and

$$V^T S_w V = \begin{pmatrix} 0_j & & \\ & I_{d-j} & \\ & & 0_{n-d} \end{pmatrix}. \quad (22)$$

As stated in Sect. 2.2, the first $j = \mathrm{rank}(S_b) = c - 1$ columns of V form the projection matrix of LDA/GSVD G_g. Hence,

$$G_g^T S_b G_g = I_j = I_{c-1} \quad (23)$$

and

$$G_g^T S_w G_g = 0_j = 0_{c-1}. \quad (24)$$

From Sect. 2.1, the projection matrix of NLDA $G_n = PQR$ satisfies

$$G_n^T S_b G_n = R^T Q^T P^T S_b PQR = R^T Q^T \tilde{S}_b QR = R^T \hat{S}_b R = \Lambda_f \quad (25)$$

and

$$G_n^T S_w G_n = R^T Q^T P^T S_w PQR = R^T Q^T \tilde{S}_w QR = R^T 0_s R = 0_f. \quad (26)$$

And we have $f = \mathrm{rank}(S_b) = c - 1$ under condition (14). So

$$G_n^T S_b G_n = \Lambda_{c-1} \quad (27)$$

and

$$G_n^T S_w G_n = 0_{c-1}. \quad (28)$$

From (23) (24) (27) (28), we can conclude that G_g and G_n are both derived from intersection of the null space of

S_w and the range space of S_b, and the only difference of two algorithms is the diagonal matrix Λ_{c-1} and I_{c-1}.

4 Modified NLDA

In this section, we develop a modified NLDA (MNLDA) algorithm whose performance is just equivalent to LDA/GSVD and much more efficient than LDA/GSVD. Further, the discriminating power and efficiency of MNLDA, LDA/GSVD, and NLDA are compared.

4.1 The algorithm

From Sect. 3.2, we know that the difference of NLDA and LDA/GSVD is the scale difference of Λ_{c-1} and I_{c-1} when condition (14) holds. If the scale of NLDA is modified, it will be equivalent to LDA/GSVD. Now we present the MNLDA algorithm as follows:

1. Calculate S_t by (4). Work out S_t's Eigen vectors p_1, p_2, …, p_d corresponding to nonzero Eigen values.
2. Calculate S_b and S_w by (2) (3). Let $P = (p_1, p_2, …, p_d)$ and $\Lambda = P^T S_t P$. We obtain the new scatter matrices $\tilde{S}_b = \Lambda^{-\frac{1}{2}} P^T S_b P \Lambda^{-\frac{1}{2}}$ and $\tilde{S}_w = \Lambda^{-\frac{1}{2}} P^T S_w P \Lambda^{-\frac{1}{2}}$ after the removal of the null space of S_t and the normalization of Λ.
3. Work out $\tilde{S}_w$'s Eigen vectors $q_1, q_2, …, q_s$ corresponding to 0 Eigen values. The projection matrix of MNLDA is $G_m = P\Lambda^{-\frac{1}{2}} Q$, where $Q = (q_1, q_2, …, q_s)$.
4. For a sample x, the new MNLDA feature is $y = G_m^T x$.

From the MNLDA algorithm, we have

$$G_m^T S_w G_m = Q^T \Lambda^{-\frac{1}{2}} P^T S_w P \Lambda^{-\frac{1}{2}} Q = Q^T \tilde{S}_w Q = 0_s \quad (29)$$

and

$$G_m^T S_t G_m = Q^T \Lambda^{-\frac{1}{2}} P^T S_t P \Lambda^{-\frac{1}{2}} Q = Q^T I_d Q = I_s. \quad (30)$$

It follows that

$$G_m^T S_b G_m = G_m^T S_t G_m - G_m^T S_w G_m = I_s - 0_s = I_s. \quad (31)$$

Since $\mathrm{rank}(S_t) = M - 1$ and $\mathrm{rank}(S_w) = M - c$ under condition (14), we have $s = (M - 1) - (M - c) = c - 1$. Hence

$$G_m^T S_b G_m = I_{c-1} \quad (32)$$

and

$$G_m^T S_w G_m = 0_{c-1}. \quad (33)$$

From (23) (24) (32) (33), we can make a conclusion that MNLDA is equivalent to LDA/GSVD.

4.2 Comparison of MNLDA, LDA/GSVD, and NLDA

The SVD process in step 2 of LDA/GSVD algorithm is time-consuming especially for high-dimensional data, whereas the Eigen value decomposition (EVD) process in MNLDA and NLDA algorithm is very quick. Hence, NLDA and MNLDA are more efficient than LDA/GSVD.

From the above discussion, it can be known that MNLDA and LDA/GSVD have the same discriminating power. Now we compare the discriminating capability of MNLDA and NLDA. In the literature [19], the authors indicated that step 4 of NLDA algorithm can be removed and proposed the improved NLDA (INLDA) algorithm, which is equivalent to NLDA when condition (14) is satisfied. The only difference between INLDA and MNLDA is the normalization process after the diagonalization of S_t. From the equivalence of NLDA and INLDA, we know that it is only the normalization process that leads to the different discriminating power of MNLDA and NLDA. In the experiments, it can be seen that this process increases the discriminating capability sometimes and reduces it sometimes. It mainly depends on the characteristics of the data.

5 Experiments

In this section, we will verify our viewpoints through face and finger-knuckle-print recognition experiments. NLDA, LDA/GSVD, and MNLDA are used for feature extraction separately. The dimensionality of reduced feature spaces is kept as $c - 1$. Then the nearest neighbor classifier with Euclidean metric is employed for classification.

5.1 Face recognition experiment

5.1.1 Face data sets

ORL [29], FERET [30–32], and Yale [33] face databases are applied in our experiments.

5.1.1.1 ORL face database ORL face database contains 400 face images of 40 individuals, each providing 10 different images. For some individuals, the images were taken at different times. The face images are centralized. The facial expression (open or closed eyes, smiling or nonsmiling) and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some titling and rotation of the face of up to 20°. Moreover, there is also some variation in the scale of up to 10%. The images are all grayscale and normalized to a resolution 92×112 pixels. Figure 1 shows ten images of one person.

5.1.1.2 FERET face database The FERET face database was sponsored by the US Department of Defense through the DARPA Program. It has become a standard database for testing and evaluating face recognition algorithms. We performed algorithms on a subset of the FERET database. The subset is composed of 1,400 images of 200 individuals, and each individual has seven images. It involves variations in face expression, pose, and illumination. In the experiment, the facial portion of the original image was cropped based on the location of eyes and mouth. Then we resized the cropped images to 80×80 pixels. Seven sample images of one individual are show in Fig. 2.

5.1.1.3 Yale face database The Yale face database was constructed at the Yale Center for Computational Vision and Control. It contains 165 gray-scale images of 15 individuals, and each person has 11 different images under various facial expressions (normal, happy, sad, sleepy, surprised, and wink), lighting conditions (left-light, center-light, right-light), and with/without glasses. In the experiment, each image was manually cropped and resized to 100×80 pixels. Figure 3 shows sample images of one person.

5.1.2 Experimental results

On ORL database, three images/people are randomly selected for training and the remaining images are used for testing. Features are extracted separately by NLDA, LDA/GSVD, and MNLDA. Then the nearest neighbor classifier with Euclidean distance is applied to classify the new features. Experiment is repeated for 10 times. Figure 4 shows the recognition rates versus the 10 different training sets. Table 1 lists the average recognition rates and standard deviations across 10 runs. Table 2 lists the computational time of each method.

Figure 4 shows that MNLDA has the same recognition rates as LDA/GSVD consistently and that NLDA

Fig. 1 Sample images of one person on ORL database

Fig. 2 Sample images of one person on FERET database

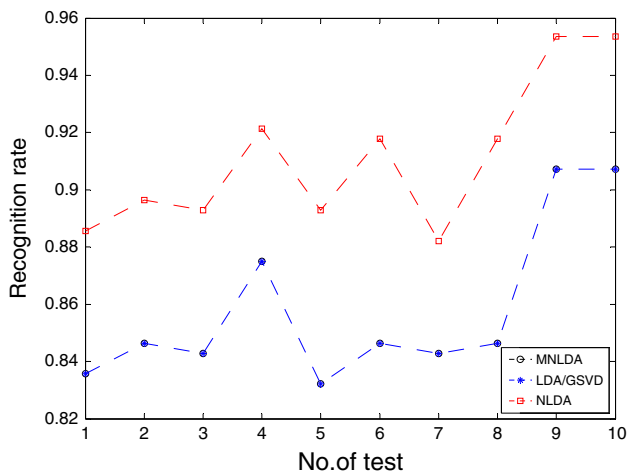**Fig. 3** Sample images of one person on YALE database

Fig. 4 The recognition rates of MNLDA, LDA/GSVD, and NLDA versus different training sets on ORL database when three images/people are randomly used for training

Table 1 The average recognition rates (%) and the corresponding standard deviation of MNLDA, LDA/GSVD, and NLDA on ORL database across 10 runs when 3 images/individual are randomly used for training

Methods	Average recognition rate	SD
MNLDA	85.8	2.8
LDA/GSVD	85.8	2.8
NLDA	91.1	2.6

Table 2 The computational time (s) of MNLDA, LDA/GSVD, and NLDA on ORL database across 10 runs when three images/individual are randomly used for training (CPU: pentium 1.8 GHz RAM: 1 Gb)

No. of test	MNLDA	LDA/GSVD	NLDA
1	1.6929	343.1512	1.3603
2	1.8350	321.4875	1.5972
3	2.3750	256.4447	1.5534
4	2.0999	253.8448	2.1527
5	2.1166	247.6787	1.5277
6	2.0889	278.3178	1.4554
7	2.0248	246.8955	1.4778
8	2.2749	273.0403	1.4944
9	2.0333	335.0893	1.4777
10	2.1582	251.4714	1.5194
Average	2.0700	280.7421	1.5616

outperforms MNLDA and LDA/GSVD irrespective of the variation of training sets. From Table 1, we can see that the average recognition rates of MNLDA and LDA/GSVD are only 86.1% while NLDA reaches 90.5%. From Table 2, it can be seen that NLDA and MNLDA are more efficient than LDA/GSVD. The average computational time of LDA/GSVD is about two hundred times as much as that of NLDA.

On FERET database, we select the first α (α varies from 2 to 5) images/people for training and the rest images for
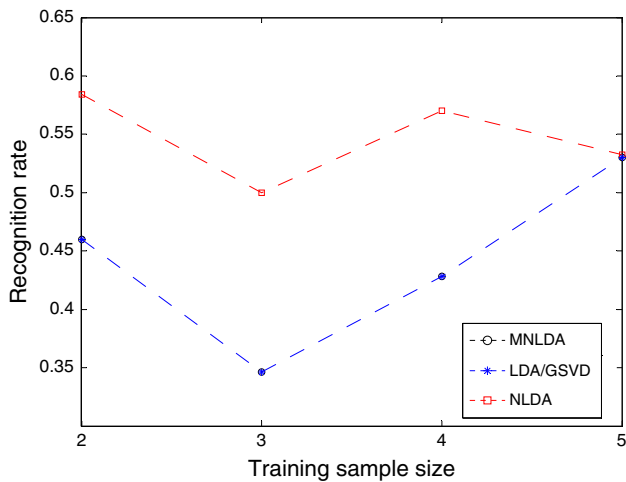
Fig. 5 The recognition rates of MNLDA, LDA/GSVD, and NLDA versus the training sample size on FERET database
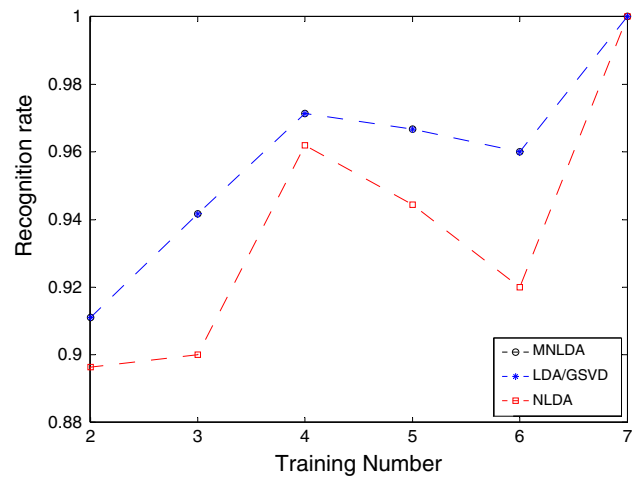

Fig. 6 The recognition rates of MNLDA, LDA/GSVD, and NLDA versus the training sample size on Yale database

Table 3 The computational time (s) of MNLDA, LDA/GSVD, and NLDA on FERET database when the first α (α varies from 2 to 5) images are used for training (CPU: pentium 1.8 GHz RAM: 1 Gb)

Training number	MNLDA	LDA/GSVD	NLDA
2	19.6648	139.3869	19.2463
3	44.0312	191.0451	42.2673
4	81.0988	241.3586	76.8765
5	132.0471	321.7521	124.2529

Table 4 The computational time (s) of MNLDA, LDA/GSVD, and NLDA on Yale database when the first β (β varies from 2 to 7) images are used for training (CPU: pentium 1.8 GHz RAM: 1 Gb)

Training number	MNLDA	LDA/GSVD	NLDA
2	0.2530	37.2991	0.2016
3	0.3952	48.4797	0.3335
4	0.6056	60.0472	0.5126
5	0.8108	71.4465	0.7250
6	1.0673	82.9715	0.9851
7	1.3731	94.3994	1.2975

testing. After feature extraction by NLDA, LDA/GSVD, and MNLDA, the nearest neighbor classifier with Euclidean distance is employed for classification. Figure 5 illustrates the recognition rates versus training sample size. The computational time of each method is shown in Table 3.

From Fig. 5, we can also see that MNLDA and LDA/GSVD have the identical recognition rates and NLDA has a higher recognition rates than them irrespective of the training sample size. From Table 3, it can be seen that MNLDA and NLDA are more efficient than LDA/GSVD obviously.

On Yale database, the first β (β varies from 2 to 7) images/people are selected to form the training set and the remaining $11 - \beta$ images are used for testing. NLDA, LDA/GSVD, and MNLDA are used for feature extraction separately. After feature extraction, the nearest neighbor classifier with Euclidean distance is used for classification. Figure 6 shows the recognition rates versus training sample size. The computational time of each method is listed in Table 4.

From Table 4, we can see that MNLDA and NLDA have higher efficiency than LDA/GSVD. Figure 6 shows that the recognition rates of MNLDA are the same as LDA/GSVD irrespective of the variation in training sample size.

These two points are consistent with the experimental results on ORL and FERET databases. However, from Fig. 6 we can also find an inconsistent point that NLDA outperforms MNLDA and LDA/GSVD on Yale database.

What is the reason of the results that NLDA outperforms MNLDA on ORL and FERET databases and MNLDA outperforms NLDA on Yale database? We suspect it can be explained as follows: From Sect. 4, we know that it is only the normalization process after the diagonalization of S_t that leads the different discriminating power of MNLDA and NLDA. In paper [34], the authors point out that the principal Eigen vectors of S_t contain two components: intrinsic difference ($\tilde{I}$) that discriminates different face identity and transformation difference ($\tilde{T}$) arising from all kinds of transformations such as expression and illumination. On ORL and FERET databases, $\tilde{I}$ may be contained on the principal Eigen vectors corresponding to large Eigen values while $\tilde{T}$ corresponding to small Eigen values. The normalization process after the diagonalization of S_t will enlarge the effect of $\tilde{T}$ and reduce the effect of $\tilde{I}$ which is important for recognition. So MNLDA involving normalization
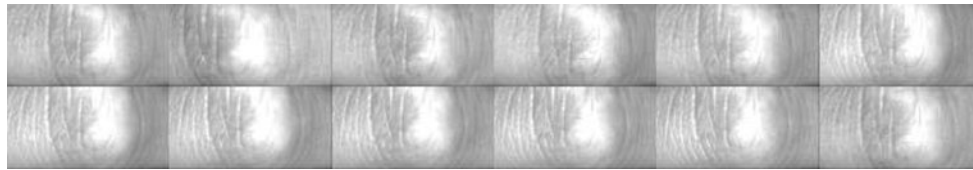
Fig. 7 Sample images of one *right* index finger

Table 5 The recognition rates (%) and computational time (s) of MNLDA and LDA/GSVD on the PolyU FKP database when the first six FKP images are used for training (CPU: pentium 1.8 GHz RAM: 1 Gb)

	MNLDA	LDA/GSVD
Recognition rate	55.2	63.0
Computational time	55.9234	149.0926

process performs worse. Inversely, on Yale database, $\tilde{I}$ corresponds to small Eigen values while $\tilde{T}$ large. So the normalization process will enlarge the effect of $\tilde{I}$ and reduce the effect of $\tilde{T}$. Hence MNLDA performs better.

5.2 Finger-knuckle-print recognition experiment on the PolyU FKP database

Finger-knuckle-print (FKP) images on the PolyU FKP database were collected from 165 volunteers, including 125 men and 40 women. Among them, 143 subjects were 20–30 years old and the others were 30–50 years old. The samples were collected in two separate sessions. In each session, the subject was asked to provide 6 images for each of the left index finger, the left middle finger, the right index finger, and the right middle finger. Therefore, 48 images from 4 fingers were collected from each subject. In total, the database contains 7,920 images from 660 different fingers. The average time interval between the first and the second sessions was about 25 days. The maximum and minimum intervals were 96 and 14 days, respectively. The images were processed by ROI extraction algorithm described in [35]. In the experiment, we select 1200 FKP images of the right index finger of 100 subjects and these images were resized to 55 × 110 pixels. Figure 7 shows 12 sample images of one right index finger.

The first six images collected in the first session are used for training and the rest six collected in the second session for testing. Firstly, LDA/GSVD and MNLDA are performed for feature extraction. Secondly, the nearest neighbor classifier is performed for classification. Table 5 lists the recognition rates and the computational time of two methods. Table 5 shows that MNLDA is much more efficient than LDA/GSVD. However, from Table 5, it can be seen that LDA/GSVD and MNLDA obtain different recognition rates. This is not consistent with the results received from face

recognition experiments. Why this happens? From the analysis above, we know LDA/GSVD and MNLDA have the same discriminating capability when condition (14) is satisfied. The number of training samples is 600 in this experiment. We find that $\text{rank}(S_t) = 593 \neq 600 - 1$, namely condition (14) is not satisfied. Therefore, the equivalence of LDA/GSVD and MNLDA does not hold.

6 Conclusion

In this paper, we analyze NLDA and LDA/GSVD algorithms and show their relationship under a condition which is often satisfied for high-dimensional data. Based on the relationship, MNLDA algorithm which is equivalent to LDA/GSVD is proposed. When the given condition is satisfied, MNLDA has the same discriminating power as LDA/GSVD and is more efficient. We also compare the efficiency and discriminating power of NLDA and LDA/GSVD. Like MNLDA, NLDA is more efficient than LDA/GSVD. Face recognition experiments demonstrate the equivalence of LDA/GSVD and MNLDA under the given condition and MNLDA is more efficient. FKP recognition experiment shows that these two algorithms have different discriminating power when the given condition does not hold. Moreover, the different discriminating power of NLDA and MNLDA mainly depends on the characteristic of the data. We give our interpretation about this through face recognition experiments.

References

1. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323
2. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326
3. Belkin M, Niyogi P (2003) Laplacian Eigen maps for dimensionality reduction and data representation. Neural Comput 15(6):1373–1396
4. He X, Yan S, Hu Y, Niyogi P, Zhang HJ (2005) Face recognition using Laplacian faces. IEEE Transact Pattern Anal Mach Intell 27(3):328–340

5. Fu Y, Huang TS (2005) Locally linear embedded Eigen space analysis. University Illinois Urbana-Champaign, Urbana, IL. Technical Report IFP-TR, ECE

6. Yang J, Zhang D, Yang JY, Niu B (2007) Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. IEEE Transact Pattern Anal Mach Intell 29(4):650–664

7. Yan S, Xu D, Zhang B, Zhang HJ, Yang Q, Lin S (2007) Graph embedding, extensions: a general framework for dimensionality reduction. IEEE Transact Pattern Anal Mach Intell 29(1):40–51

8. Chen HT, Chang HW, Liu TL (2005) Local discriminant embedding and its variants, In: Proceedings of CVPR 2005

9. Fu Y, Liu M, Huang TS (2007) Conformal embedding analysis with local graph modeling on the unit hyper sphere, In: Proceedings of CVPR 2007

10. Cai D, He X, Zhou K, Han J, Bao H (2007) Locality sensitive discriminant analysis, In: Proceedings of IJCAI

11. Yang WK, Sun CY, Zhang L (2011) A multi-manifold discriminant analysis method for image feature extraction. Pattern Recogn 44(8):1649–1657

12. Yang WK, Wang JG, Ren MW, Yang JY, Zhang L, Liu GH (2009) Feature extraction based on Laplacian bidirectional maximum margin criterion. Pattern Recogn 42(11):2327–2334

13. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley, NY

14. Belhumeur P, Hespanha J, Kriegman D (1997) Eigen faces versus fisher faces: recognition using class specific linear projection. IEEE Transact Pattern Anal Mach Intell 19(7):711–720

15. Zuo W, Zhang D, Yang J, Wang K (2006) BDPCA plus LDA: a novel fast feature extraction technique for face recognition. IEEE Transact Syst, Man, Cybern-Part B: Cybern 36(4):946–953

16. Jing XY, Zhang D, Tang YY (2004) An improved LDA approach. IEEE Transact Syst, Man, Cybern-Part B: Cybern 34(5):1942–1951

17. Friedman JH (1989) Regularized discriminant analysis. J Am Stat Assoc 84(405):165–175

18. Chen L-F, Liao H-YM, Ko M-T, Lin J-C, Yu G-J (2000) A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recogn 33(10):1713–1726

19. Ye J, Xiong T (2006) Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. J Mach Learn Res 7:1183–1204

20. Yu H, Yang J (2001) A direct LDA algorithm for high dimensional data with application to face recognition. Pattern Recogn 34(10):2067–2070

21. Yang J, Yang JY (2003) Why can LDA be performed in PCA transformed space? Pattern Recogn 36(2):563–566

22. Yang J, Alejandro FF, Yang J, Zhang D, Jin Z (2005) KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition. IEEE Transact Pattern Anal Mach Intell 27(2):230–244

23. Zhuang XS, Dai DQ (2005) Inverse fisher discriminate criteria for small sample size problem and its application to face recognition. Pattern Recogn 38(11):2192–2194

24. Howland P, Jeon M, Park H (2003) Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition. SIAM J Matrix Anal Appl 25(1):165–179

25. Howland P, Park H (2004) Generalizing discriminant analysis using the generalized singular value decomposition. IEEE Transact Pattern Anal Mach Intell 26(8):995–1006

26. Turk MA, Pentland AP (1991) Face recognition using Eigen faces. In: Proceedings of CVPR 1991

27. Huang R, Liu Q, Lu H, Ma S (2002) Solving the small sample size problem of LDA. In: Proceedings of ICPR 2002

28. Paige CC, Saunders MA (1981) Towards a generalized singular value decomposition. SIAM J Num Anal 18(3):398–405

29. http://www.cvc.yale.edu/projects/yalefaces/yalefaces.html

30. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face recognition algorithms. IEEE Transact Pattern Anal Mach Intell 22(10):1090–1104

31. Phillips PJ (2006) The facial recognition technology (FERET) database. http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

32. Phillips PJ, Wechsler H, Huang J, Rauss P (1998) The FERET database and evaluation procedure for face recognition algorithms. Image Vis Comput 16(5):295–306

33. http://www.cam-orl.co.uk

34. Wang X, Tang X (2004) A unified framework for subspace face recognition. IEEE Transact Pattern Anal Mach Intell 26(9):1222–1228

35. Zhang L, Zhang L, Zhang D, Zhu HL (2010) Online finger-knuckle-print verification for personal authentications. Pattern Recogn 43(7):2560–2571