Tensor Locally Linear Discriminative Analysis

Zhao Zhang, *Student Member, IEEE*, and W. S. Chow, *Senior Member, IEEE*

***Abstract*—This letter presents a Tensor Locally Linear Discriminative Analysis (TLLDA) method for image presentation. TLLDA is originated from the Local Fisher Discriminant Analysis (LFDA), but TLLDA offers some advantages over LFDA. 1) TLLDA can preserve the local discriminative information of image data as LFDA. 2) TLLDA represents images as matrices or 2-order tensors rather than vectors, so TLLDA keeps the spatial locality of pixels in the images. 3) TLLDA avoids the singularity that may be suffered by LFDA. 4) TLLDA is faster than LFDA. Simulations on two real databases are conducted to verify the validity of TLLDA. Results show that TLLDA is highly competitive with some widely used techniques.**

***Index Terms*—Dimensionality reduction, discriminant analysis, tensor representation, trace ratio optimization.**

I. Introduction

THE *Principal Component Analysis* (PCA) [1], *Fisher Linear Discriminant Analysis* (LDA) [1] and *Locality Preserving Projections* (LPP) [2] are the three most popular dimensionality reduction (DR) methods. Based on the ideas of LDA and LPP, an efficient discriminant technique called *Local Fisher Discriminant Analysis* (LFDA) [3] was proposed for multimodal DR. LFDA can preserve the local and discriminant information of data effectively. Note that these methods all perform in vector space, extracting the features from vector patterns directly. To learn a LDA, PCA, LPP or LFDA subspace, images firstly need to be converted to 1-D vectors in high-dimensional vector space. But images are intrinsically matrices or 2-order tensors. To solve this issue, multidimensional extensions of PCA, LDA and LPP, namely 2DPCA [11], 2DLDA [10] and tensorized LPP (TLPP) [7], are presented. These methods aim to compute the subspaces directly from the 2-D image matrices rather than 1-D vectors.

This letter represents images of size $n_1 \times n_2$ as matrices in tensor space $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$. We then propose a *Tensor Locally Linear Discriminative Analysis* (TLLDA) by tensorizing LFDA. Compared with LFDA, TLLDA gives natural representations of images. TLLDA is computationally efficient as the decomposed matrices are of size $n_1 \times n_1$ or $n_2 \times n_2$, which is smaller than that of size $n \times n (n = n_1 \times n_2)$ in LFDA. TLLDA avoids the singularity problem [6]. The trace ratio optimization [8], [9], [12] is used to solve TLLDA enabling TLLDA to deliver more specific solution.

The outline of this letter is described as follows. Section II reviews LFDA. Section III shows our method. In Section III, we conduct simulations. The conclusion is given in Section V

Manuscript received June 29, 2011; revised August 01, 2011; accepted August 07, 2011. Date of publication August 22, 2011; date of current version September 22, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yong Man Ro.

The authors are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: itzzhang@ee.cityu.edu.hk; eetchow@cityu.edu.hk).

Digital Object Identifier 10.1109/LSP.2011.2165538

II. Local Fisher Discriminant Analysis (LFDA)

Let $x_i \in \mathbb{R}^n \quad (i = 1, 2, \ldots, m)$ be vectors of n-dimensional data and $y_i (\in \{1, 2, \ldots, c\})$ be the corresponding class labels of samples, where c is the class number, LFDA finds a $n \times d$ projection matrix Ξ projecting x_i into low-dimensional representation $\tilde{x}_i = \Xi^{\mathrm{T}} x_i$, where notation ^{T} is the transpose of a matrix or vector. Let $\tilde{S}^{(lwc)}$ and $\tilde{S}^{(lbc)}$ be the local intra- and inter-class scatter of LFDA respectively, we have

$$\tilde{S}^{(lwc)} = \frac{1}{2} \sum_{i,j=1}^{m} \|x_i - x_j\|^2 \tilde{P}_{i,j}^{(wc)},$$
$$\tilde{S}^{(lbc)} = \frac{1}{2} \sum_{i,j=1}^{m} \|x_i - x_j\|^2 \tilde{P}_{i,j}^{(bc)} \tag{1}$$

where $\|\cdot\|$ is the Euclidean distance. Let m_l be the number of points in class l, then matrices $\tilde{P}^{(wc)}$ and $\tilde{P}^{(bc)}$ are defined as

$$\tilde{P}_{i,j}^{(wc)} = \begin{cases} W_{i,j}/m_l, & \text{if } y_i = y_j = l \\ 0, & \text{else if } y_i \neq y_j \end{cases}, \tag{2}$$

$$\tilde{P}_{i,j}^{(bc)} = \begin{cases} W_{i,j}/(1/m - 1/m_l), & \text{if } y_i = y_j = l \\ 1/m, & \text{else if } y_i \neq y_j \end{cases} \tag{3}$$

where W represents the similarity between x_i and x_j. W can be defined by heat kernel [4] or local scaling heuristic method [5]. $W_{i,j}$ is large if x_i and x_j are neighbors, and else small. The neighbors of x_i is determined by k-nearest neighbor search [4]. Then the objective function of LFDA is defined as

$$\underset{\Xi \in \mathbb{R}^{n \times d}}{\mathrm{Max}} \ \mathrm{Tr}\left(\Xi^{\mathrm{T}} \tilde{A}^{(lbc)} \Xi\right), \quad \text{s.t.} \Xi^{\mathrm{T}} \tilde{A}^{(lwc)} \Xi = I \tag{4}$$

where I is an identity matrix, $\mathrm{Tr}(\cdot)$ denotes the matrix trace of matrix. Then $\Xi = [\psi_1 | \psi_2 | \ldots | \psi_d]$ is solved from $\tilde{A}^{(lbc)} \psi_j = \lambda_j \left(\tilde{A}^{(lwc)} + \beta I\right) \psi_j$, where βI with $\beta = 0.01$ is added to avoid the singularity in this letter. Thus the LFDA projection axes are not orthogonal.

III. Tensor Locally Linear Discriminative Analysis

A. Matrix (or, Tensor) Interpretation of Images

Let $x \in \mathbb{R}^{n_1 \times n_2}$ be an image of size $n_1 \times n_2$, then a second order tensor of x is mathematically defined as $x = \sum_{i,j} (u_i^{\mathrm{T}} x v_j) u_i v_j^{\mathrm{T}}$. in tensor space $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$, where $(u_1, u_2, \ldots, u_{n_1})$ is a set of basis functions in $\mathbb{R}^{n_1}$ and $(v_1, v_2, \ldots, v_{n_2})$ is a set of basis functions in $\mathbb{R}^{n_2}$ [7], [12]. So

$\{u_i v_j^{\mathrm{T}}\}$ forms a basis of tensor space $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$. Let U be a subspace of R^{n_1} spanned by $\{u_i\}_{i=1}^{d_1}$ and V be a subspace of R^{n_2} spanned by $\{v_i\}_{i=1}^{d_2}$, then tensor product $U \otimes V$ is a subspace of $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$. By projecting each $x_i \in \mathbb{R}^{n_1 \times n_2}$ onto $U \otimes V$, we obtain tensors $\eta_i = U^{\mathrm{T}} x_i V \in \mathrm{R}^{d_1 \times d_2} (d_1 \leq n_1, d_2 \leq n_2)$.

B. The Objective Function

Let $\tilde{\Upsilon}^{(lbc)}$ and $\tilde{\Upsilon}^{(lwc)}$ denote the inter- and intra-class scatters of TLLDA, we have the following classwise expressions:

$$\tilde{\Upsilon}^{(lwc)} = \frac{1}{2} \sum_{i,j=1}^{m} \|\eta_i - \eta_j\|^2 \tilde{P}_{i,j}^{(wc)},$$
$$\tilde{\Upsilon}^{(lbc)} = \frac{1}{2} \sum_{i,j=1}^{m} \|\eta_i - \eta_j\|^2 \tilde{P}_{i,j}^{(bc)} \tag{5}$$

where $\eta_i = U^{\mathrm{T}} x_i V, \eta_j = U^{\mathrm{T}} x_j V$. The weight matrices $\tilde{P}^{(wc)}$ and $\tilde{P}^{(bc)}$ are respectively defined as

$$\tilde{P}_{i,j}^{(wc)} = \begin{cases} W_{i,j}/\sqrt{m}, & \text{if } y_i = y_j \\ 0, & \text{else if } y_i \neq y_j \end{cases}, \tag{6}$$
$$\tilde{P}_{i,j}^{(bc)} = \begin{cases} W_{i,j}/(1/\sqrt{2m} - 1/\sqrt{m}), & \text{if } y_i = y_j \\ 1/\sqrt{m}, & \text{else if } y_i \neq y_j \end{cases} \tag{7}$$

where W is similarly defined as LFDA. Then the following trace ratio (TR) problem [8], [9] is defined for TLLDA:

$$(U, V) = \underset{U \in \mathbb{R}^{n_1 \times d_1}, V \in \mathbb{R}^{n_2 \times d_2}}{\arg\max} \operatorname{Tr}\left(\tilde{\Upsilon}^{(lbc)}\right) \Big/ \operatorname{Tr}\left(\tilde{\Upsilon}^{(lwc)}\right) \tag{8}$$

from which the transforming axes U and V can be computed.

C. Computational Analysis

Let $D^{(bc)}$ be a diagonal matrix with entries being $D_{ii}^{(wc)} = \sum_j \tilde{P}_{i,j}^{(wc)}$. Since $\|A\|^2 = \operatorname{Tr}(AA^{\mathrm{T}})$, we can obtain

$$\tilde{\Upsilon}^{(lwc)} = \frac{1}{2} \sum_{i,j=1}^{m} \operatorname{Tr}\left((\eta_i - \eta_j)(\eta_i - \eta_j)^{\mathrm{T}}\right) \tilde{P}_{i,j}^{(wc)}$$
$$= \operatorname{Tr}\left(\sum_{i=1}^{m} D_{ii}^{(wc)} \eta_i \eta_i^{\mathrm{T}} - \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(wc)} \eta_i \eta_j^{\mathrm{T}}\right). \tag{9}$$

By substituting η_i and η_j into (9), we can obtain

$$\tilde{\Upsilon}_{+}^{(lwc)} = \operatorname{Tr}\left(U^{\mathrm{T}}\left(F_V^{(wc)} - H_V^{(wc)}\right) U\right) \tag{10}$$

where $F_V^{(wc)} = \sum_{i=1}^{m} D_{ii}^{(wc)} x_i V V^{\mathrm{T}} x_i^{\mathrm{T}}, H_V^{(wc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(wc)} x_i V V^{\mathrm{T}} x_j^{\mathrm{T}}$. Similarly, $\tilde{\Upsilon}^{(lbc)}$ is converted to

$$\tilde{\Upsilon}_{+}^{(lbc)} = \operatorname{Tr}\left(U^{\mathrm{T}}\left(F_V^{(bc)} - H_V^{(bc)}\right) U\right) \tag{11}$$

where $D_{ii}^{(wc)} = \sum_j \tilde{P}_{i,j}^{(wc)}, F_V^{(bc)} = \sum_{i=1}^{m} D_{ii}^{(bc)} x_i V V^{\mathrm{T}} x_i^{\mathrm{T}}, H_V^{(bc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(bc)} x_i V V^{\mathrm{T}} x_j^{\mathrm{T}}$. Also, we have $\|A\|^2 = \operatorname{tr}(A^{\mathrm{T}} A)$, so $\tilde{\Upsilon}^{(lwc)}$ and $\tilde{\Upsilon}^{(lbc)}$ become

$$\tilde{\Upsilon}_{-}^{(lwc)} = \operatorname{Tr}\left(V^{\mathrm{T}}\left(F_U^{(wc)} - H_U^{(wc)}\right) V\right), \tag{12}$$
$$\tilde{\Upsilon}_{-}^{(lbc)} = \operatorname{Tr}\left(V^{\mathrm{T}}\left(F_U^{(bc)} - H_U^{(bc)}\right) V\right) \tag{13}$$

where $F_U^{(wc)} = \sum_{i=1}^{m} D_{ii}^{(wc)} x_i^{\mathrm{T}} U U^{\mathrm{T}} x_i, H_U^{(wc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(wc)} x_i^{\mathrm{T}} U U^{\mathrm{T}} x_j, F_U^{(bc)} = \sum_{i=1}^{m} D_{ii}^{(bc)} x_i^{\mathrm{T}} U U^{\mathrm{T}} x_i, H_U^{(bc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(bc)} x_i^{\mathrm{T}} U U^{\mathrm{T}} x_j$. Then (8) becomes

$$\underset{U \in \mathbb{R}^{n_1 \times d_1}, V \in \mathbb{R}^{n_2 \times d_2}}{\operatorname{Max}} \frac{\operatorname{Tr}\left(U^{\mathrm{T}}\left(F_V^{(bc)} - H_V^{(bc)}\right) U\right)}{\operatorname{Tr}\left(U^{\mathrm{T}}\left(F_V^{(wc)} - H_V^{(wc)}\right) U\right)}, \tag{14}$$

$$\underset{U \in \mathbb{R}^{n_1 \times d_1}, V \in \mathbb{R}^{n_2 \times d_2}}{\operatorname{Max}} \frac{\operatorname{Tr}\left(V^{\mathrm{T}}\left(F_U^{(bc)} - H_U^{(bc)}\right) V\right)}{\operatorname{Tr}\left(V^{\mathrm{T}}\left(F_U^{(wc)} - H_U^{(wc)}\right) V\right)}. \tag{15}$$

Note that the above problems are all subjected to U and V, so they cannot be solved independently. This work adopts the similar computational method as [7] to compute U and V. By setting U to be an identity matrix and $\tilde{x}_i = x_i^{\mathrm{T}} U$, we obtain

$$\tilde{F}_U^{(wc)} = \sum_{i=1}^{m} D_{ii}^{(wc)} \tilde{x}_i \tilde{x}_i^{\mathrm{T}}, \quad \tilde{H}_U^{(wc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(wc)} \tilde{x}_i \tilde{x}_j^{\mathrm{T}}$$
$$\tilde{F}_U^{(bc)} = \sum_{i=1}^{m} D_{ii}^{(bc)} \tilde{x}_i \tilde{x}_i^{\mathrm{T}}, \quad \tilde{H}_U^{(bc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(bc)} \tilde{x}_i \tilde{x}_j^{\mathrm{T}}. \tag{16}$$

Thus the optimization problem in (15) can be simplified as $\operatorname{Max}_{V \in \mathbb{R}^{n_2 \times d_2}, V^{\mathrm{T}} V = I_{d_2 \times d_2}} \operatorname{Tr}(V^{\mathrm{T}}(\tilde{F}_U^{(bc)} - \tilde{H}_U^{(bc)})V)/\operatorname{Tr}(V^{\mathrm{T}}(\tilde{F}_U^{(wc)} - \tilde{H}_U^{(wc)})V)$ with respect to $V^{\mathrm{T}} V = I_{d_2 \times d_2}$, from which V can be computed. After V is obtained, we similarly set $\widehat{x}_i = x_i V$,then $F_V^{(wc)}, H_V^{(wc)}, F_V^{(bc)}$ and $H_V^{(bc)}$ can also be simplified as

$$\tilde{F}_V^{(wc)} = \sum_{i=1}^{m} D_{ii}^{(wc)} \widehat{x}_i \widehat{x}_i^{\mathrm{T}}, \quad \tilde{H}_V^{(wc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(wc)} \widehat{x}_i \widehat{x}_j^{\mathrm{T}}$$
$$\tilde{F}_V^{(bc)} = \sum_{i=1}^{m} D_{ii}^{(bc)} \widehat{x}_i \widehat{x}_i^{\mathrm{T}}, \quad \tilde{H}_V^{(bc)} = \sum_{i,j=1}^{m} \tilde{P}_{i,j}^{(bc)} \widehat{x}_i \widehat{x}_j^{\mathrm{T}}. \tag{17}$$

Similarly the optimization problem in (14) can be converted to $\operatorname{Max}_{U \in \mathbb{R}^{n_1 \times d_1}, U^{\mathrm{T}} U = I_{d_1 \times d_1}} \operatorname{Tr}(U^{\mathrm{T}}(\tilde{F}_V^{(bc)} - \tilde{F}_V^{(bc)})U)/\operatorname{Tr}(U^{\mathrm{T}}(\tilde{F}_V^{(wc)} - \tilde{H}_V^{(wc)})U)$ subject to $U^{\mathrm{T}} U = I_{d_1 \times d_1}$. Thus U can be similarly obtained.

D. Effective Solution Using Trace Ratio Optimization

This section shows how to compute U and V with the iterative ITR [8]. We consider the following TR problem:

$$\underset{\Theta^{\mathrm{T}} \Theta = I}{\operatorname{Max}} \operatorname{Tr}\left(\Theta^{\mathrm{T}} \tilde{L}^{(bc)} \Theta\right) \Big/ \operatorname{Tr}\left(\Theta^{\mathrm{T}} \tilde{L}^{(wc)} \Theta\right). \tag{18}$$

Then ITR tackles (18) by directly optimizing the objective $\operatorname{Tr}(\Theta^{\mathrm{T}} \tilde{L}^{(bc)} \Theta)/\operatorname{Tr}(\Theta^{\mathrm{T}} \tilde{L}^{(wc)} \Theta)$ by assuming that the column vectors of Θ are unitary and orthogonal together. Given TR value λ^v at step v, matrix Θ^v can be obtained from solving the following optimization problem:

$$\Theta^v = \underset{\Theta^{\mathrm{T}} \Theta = I}{\arg\max} \operatorname{Tr}\left(\Theta^{\mathrm{T}}\left(\tilde{L}^{(bc)} - \lambda^v \tilde{L}^{(wc)}\right) \Theta\right) \tag{19}$$

and renew λ^{v+1} as the TR value given by $\Theta^v : \lambda^{v+1} = \operatorname{Tr}(\Theta^{v\mathrm{T}} \tilde{L}^{(bc)} \Theta^v)/\operatorname{Tr}(\Theta^{v\mathrm{T}} \tilde{L}^{(wc)} \Theta^v)$ until converging to the

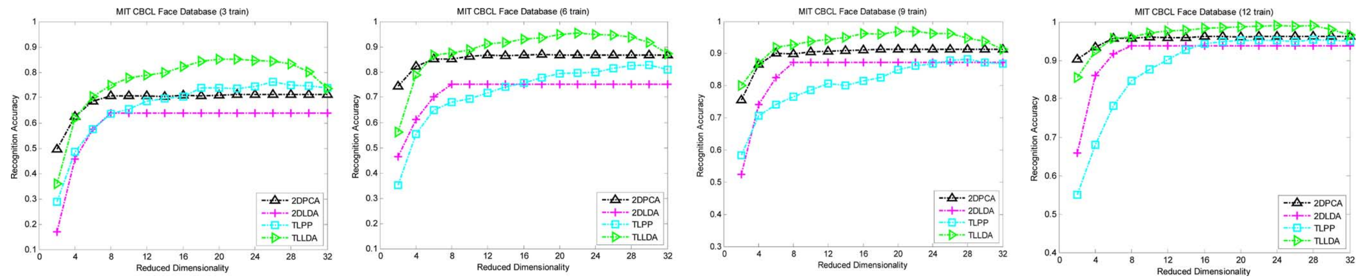

Fig. 1. Recognition accuracy versus reduced dimensionality on the MIT CBCL face database.

TABLE I
PERFORMANCE COMPARISON ON THE MIT CBCL FACE DATABASE

Result / Method	MIT CBCL (3 train)			MIT CBCL (6 train)			MIT CBCL (9 train)			MIT CBCL (12 train)		
	Mean	Best	Dim	Mean	Best	Dim	Mean	Best	Dim	Mean	Best	Dim
2DPCA	0.6892	0.7122	24	0.8546	0.8701	18	0.8971	0.9129	18	0.9554	0.9617	22
2DLDA	0.5935	0.6382	8	0.7205	0.7501	8	0.8393	0.8723	8	0.9134	0.9368	8
TLPP	0.6671	0.7607	26×26	0.7240	0.8271	30×30	0.8068	0.8828	28×28	0.8821	0.9527	28×28
TLLDA	0.7637	0.8525	20×20	0.8877	0.9534	22×22	0.9329	0.9679	20×20	0.9676	0.9899	28×28

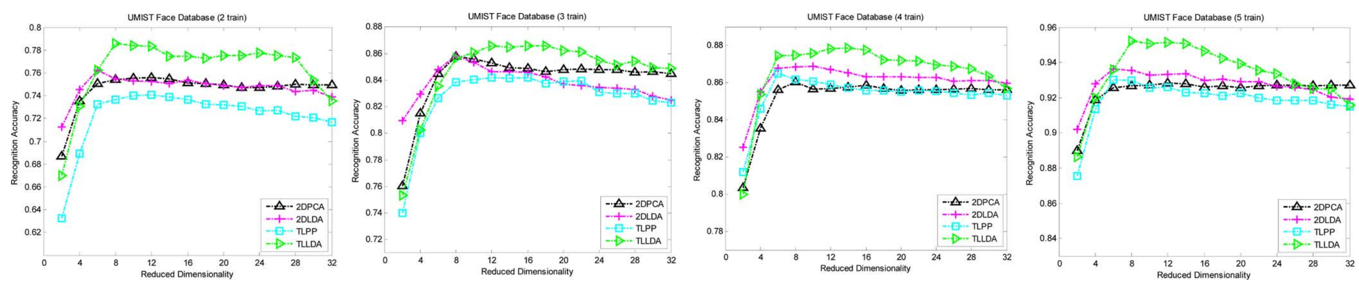

Fig. 2. Recognition accuracy versus reduced dimensionality on the UMIST face database.

global optimum [9]. Note that Θ^0 was initialized as an arbitrary orthogonal matrix in ITR so ITR maybe unstable due to the randomness. Most importantly, a bad initialization may greatly increase the number of iterations. This letter initializes $\lambda^0 = 0$ instead of initializing Θ^0 to be a columnly orthogonal matrix. To achieve V, we firstly fix U. Let $\tilde{L}_U^{(bc)} = \tilde{F}_U^{(bc)} - \tilde{H}_U^{(bc)}, \tilde{L}_U^{(wc)} = \tilde{F}_U^{(wc)} - \tilde{H}_U^{(wc)}, \tilde{L}_V^{(bc)} = \tilde{F}_V^{(bc)} - \tilde{H}_V^{(bc)}, \tilde{L}_V^{(wc)} = \tilde{F}_V^{(wc)} - \tilde{H}_V^{(wc)}$, by setting $\tilde{L}^{(bc)} = \tilde{L}_U^{(bc)}$ and $\tilde{L}^{(wc)} = \tilde{L}_U^{(wc)}$, then V can be obtained by ITR. After V is computed, U can be similarly achieved by setting $\tilde{L}^{(bc)} = \tilde{L}_V^{(bc)}$ and $\tilde{L}^{(wc)} = \tilde{L}_V^{(wc)}$. Note that the delivered projection matrix is orthogonal.

IV. SIMULATION RESULTS AND ANALYSIS

A. Simulation Setting and Data Preparation

This section tests 2DPCA, 2DLDA, TLPP, and TLLDA. The heat kernel $\exp(-\|x_i - x_j\|/t)$ is used to set the weights for TLPP, LFDA and TLLDA, where t is calculated as all pairwise distances among points. The number of neighbors, k, is set to 11. Two real databases are tested. The first one is the MIT CBCL face database (Available at http://cbcl.mit.edu/software-datasets/heisele/download/download.html). The second one is UMIST database (Available at http://www.sheffield.ac.uk/eee/research/iel/research/face). The face images are all re-sized into 32×32 pixels. This letter considers $n_1 = n_2, d_1 = d_2$. In solving TLPP and TLLDA, all the images are denoted as (32×32)-dimensional matrices. For recognition, the one-nearest-neighbor (1NN) learner with Euclidean metric is used. For each case, results are averaged over ten random splits of training and test samples.

B. Face Recognition on MIT CBCL Database

The synthetic sample set from the MIT CBCL database is tested. In this set, there are 324 images of one of ten individuals. In our study, 200 images per individual are selected for simulations. For each individual, $L(= 3, 6, 9, 12)$ images are selected for training the face subspaces and the rest are for testing. Fig. 1 plots the accuracy of each method vs. reduced dimensions. For each method, we only show its performance in the d- or $(d \times d)$-dimensional subspace, where $d = 1, 2, \ldots, 32$. Seeing from Fig. 1, the performance of each method varies with reduced dimensions. We also show the mean and best result of each method in Table I. The subspaces according to the best records are optimal subspaces. Observing from Fig. 1 and Table I, we conclude that TLLDA delivers the highest accuracies in most cases. The 2DPCA and TLPP perform comparatively to our method, whilst 2DLDA performs poorly on this dataset.

C. Face Recognition on UMIST Database

This study tests 2DPCA, 2DLDA, TLPP, and TLLDA using the UMIST database. This database consists of 564 images of 20 persons. For each individual, $L(= 2, 3, 4, 5)$ images are selected for training. Fig. 2 plots the results over reduced dimensions. For each method, we show its performance in the d- or $(d \times d)$-dimensional subspace, where $d = 1, 2, \ldots, 32$. The mean and best result by each method are list in Table II. We have the following

TABLE II
PERFORMANCE COMPARISON ON THE UMIST FACE DATABASE

Result / Method	UMIST Face (2 train)			UMIST Face (3 train)			UMIST Face (4 train)			UMIST Face (5 train)		
	Mean	Best	Dim	Mean	Best	Dim	Mean	Best	Dim	Mean	Best	Dim
2DPCA	0.7460	0.7558	12	0.8413	0.8579	8	0.8523	0.8605	8	0.9210	0.9283	12
2DLDA	0.7473	0.7623	6	0.8380	0.8582	8	0.8610	0.8689	10	0.9254	0.9329	6
TLPP	0.7221	0.7404	12×12	0.8266	0.8421	12×12	0.8534	0.8648	6×6	0.9175	0.9310	6×6
TLLDA	0.7627	0.7857	8×8	0.8471	0.8659	16×16	0.8660	0.8788	14×14	0.9368	0.9521	8×8
Method	Mean	Best	Time	Mean	Best	Time	Mean	Best	Time	Mean	Best	Time
LFDA	0.7370	0.7607	4.5204	0.8215	0.8478	4.6774	0.8492	0.8659	4.9276	0.9206	0.9295	5.1615
TLLDA	0.7627	0.7857	0.2677	0.8471	0.8659	0.3106	0.8660	0.8788	0.3927	0.9368	0.9521	0.4689

observations. 1) The performance of each method varies with reduced dimensions. As the number of dimensions increases, the result of 2DPCA is monotonically increasing, but the increasing dimensions cause the 2DLDA, TLPP and TLLDA results to deteriorate. 2) TLLDA outperforms other methods in most cases. TLPP is worse for each case. 2DPCA and 2DLDA are comparable. As training sample size increases, the accuracy of 2DLDA increases faster than 2DPCA and TLPP. We also compare TLLDA with LFDA. The same settings are tested. We compute d-dimensional vector subspace for LFDA. The results are given in the bottom of Table II. We see TLLDA outperforms LFDA. The running time is significantly reduced by TLLDA.

V. CONCLUDING REMARKS

This letter introduces a Tensor Locally Linear Discriminative Analysis technique TLLDA for image feature extraction. The key difference of TLLDA and LFDA is TLLDA uses the matrix representation of images, enabling TLLDA to extract the features from 2-D images directly. TLLDA has the analytic solution of projection matrix as LFDA. We examine TLLDA by simulations over two real databases. From all our tested cases, the overall performance of TLLDA is comparable or even outperforms some state-of-the-art methods. We also experimentally observe that TLLDA is more computationally efficient than LFDA.

REFERENCES

[1] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, 2001.
[2] X. F. He, S. C. Yan, Y. X. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, 2005.
[3] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, 2007.
[4] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.
[5] L. Zelnik-Manor and P. Perona, L. K. Saul, Y. Weiss, and L. Bottou, Eds., "Self-tuning spectral clustering," in *NIPS*, 2005, pp. 1601–1608.
[6] F. P. Nie, S. Xiang, and C. S. Zhang, "Neighborhood MinMax projections," in *Proc. IJCAI*, San Francisco, CA, 2007, pp. 993–998.
[7] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Advances in Neural Information Processing Systems*, Vancouver, BC, Canada, 2005.
[8] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, 2007, pp. 1–8.
[9] Y. Jia, F. Nie, and C. S. Zhang, "Trace ratio problem revisited," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 729–735, 2008.
[10] H. Kong, E. K. Teoh, J. G. Wang, and R. Venkateswarlu, "Two dimensional fisher discriminant analysis: Forget about small sample size problem," in *Proc. ICASSP*, 2005, pp. 761–764.
[11] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 26, no. 4, pp. 131–137, 2004.
[12] F. P. Nie, S. Xiang, Y. Q. Song, and C. S. Zhang, "Extracting the optimal dimensionality for local tensor discriminant analysis," *Patt. Recognit.*, vol. 42, no. 1, pp. 105–114, 2009.