Semi-supervised Tensor Graph-optimized Linear Discriminant Analysis for Two-dimensional Face Recognition

Shuhua Xu
Department of Maths of University of Shaoxing, Shaoxing, China
xushuhua_2012@163.com

***Abstract*—A new semi-supervised dimensionality reduction method called Semi-supervised Tensor Graph-optimized Linear Discriminant Analysis (STGLDA) is proposed for two-dimensional face recognition. Unlike recent proposed Tensor Locally Linear Discriminative Analysis (TLLDA), STGLDA offers some advantages over TLLDA. 1) STGLDA is originated from Graph-based Fisher Analysis (GbFA) while TLLDA is originated from the Local Fisher Discriminant Analysis (LFDA). In contrast to LFDA, GbFA encodes the richer discriminating information and there is no assumption that data obeys Gaussian distribution. 2) With the linear weighted way of fusing Principal Component Analysis (PCA) and GbFA, STGLDA preserves global scatter structure information and Graph-based discriminant information capturing local structure feature. TLLDA only pays attentions to preserve local structure information and discriminant information while ignores preserving global structure information. Experimental results on real face databases demonstrate that STGLDA is highly competitive with TLLDA and other tensor dimensionality reduction algorithm.**

***Index Terms*—face recognition, semi-supervised dimensionality reduction, linear discriminant analysis, tensor representation, graph optimization, information infusion, trade-off parameter**

I. INTRODUCTION

Face recognition is the computer technology applied in biometric identification. In practical applications of face recognition, high-dimensional face data not only is difficult to deal with but also contains redundant feature information. Therefore dimensionality reduction is the common pre-processing step in applications of data mining[1]. The past twenty years have witnessed rapid development on dimensionality reduction algorithms. Among them, Principal Component Analysis (PCA) [2] and Linear Discriminant Analysis (LDA) [3] are two most popular dimensionality reduction algorithms. LDA is a supervised method for feature extraction and dimensionality reduction and has been widely used in many applications such as face recognition [3-7]. Aiming at the problem of singular caused by the small sample size problems, some extensions of LDA have been developed [8-14], However, these extensions of LDA and LDA must take into account of the premise that data approximately obeys a Gaussian distribution that cannot always be satisfied in practicable applications. For solving the problem, Cui et al proposed a Graph-based Fisher Analysis (GbFA) [15]. In the term of the theory of spectrum graph [16-18], GbFA redefines the intrinsic graph based on the same-class samples and the penalty graph based on the not-same-class samples, making the original neighbor same-class samples much closer in the output space while pushing apart the original neighbor not-same-class samples in the output space. Hence GbFA encodes the rich discriminating information and enhances the classification. Moreover, there is no assumption that data obeys Gaussian distribution for GbFA.

With development of the tensor algebra, typical tensor dimensionality reduction algorithms [19, 20] are proposed. Usually tensor dimensionality reduction algorithms represent $n_1 \times ... \times n_m$ data as a point in tensor space $R^1 \times ... \times R^m$, which guarantee that projected low-dimensional data preserves spatial relations of high-dimensional data. [20] extended traditional LDA into second-order tensor space. [21] proposed Tensor Locality Sensitive Discriminant Analysis (TLSDA) to preserve the key structure of data by using the labeled samples and TLSDA has high performance as well as low time complexity. [22] proposed Tensor Locally Linear Discriminative Analysis (TLLDA) algorithm for image presentation, which is originated from the Local Fisher Discriminant Analysis (LFDA).

Semi-supervised dimensionality reduction exploits supervised information in labeled samples and fuses unsupervised information in unlabeled samples [23, 24]. Motivated above analysis, a Semi-supervised Tensor Graph-optimized Linear Discriminant Analysis (STGLDA) for face recognition is proposed, which is originated from PCA and GbFA. The algorithm firstly regards two-dimensional face images as a second-order matrix in the tensor space $R^{n_1} \otimes R^{n_2}$; then fuses PCA and GbFA with linear weighted way; finally gets two project matrixes U and V to achieve the projecting from high-dimensional data to low-dimensional data. Following the characteristic of tensor dimensionality reduction, the algorithm preserves global scatter structure and enhances discriminant information. Experimental results on YaleB and AR show that the proposed algorithm is efficient.

The rest of the paper is organized as follows: Section II reviews PCA and GbFA. STGLDA is introduced in

doi:10.4304/jsw.8.8.2047-2051

Section Ⅲ. In Section Ⅳ, we compare proposed STGLDA with TPCA, TLPP, OTNPE and TLLDA. The experimental results and analyses are presented. Finally, we provide some concluding remarks and future work in Section Ⅴ.

II. RELATED WORKS

A. Principal Component Analysis (PCA)

PCA is an unsupervised dimensionality reduction method for preserving the global data structure. Destination of PCA is to find iteratively the maximum-variance direction of the data points. The objective function of PCA is defined as:

$$T = \arg\max_{T}\left[tr\left(T^{\mathrm{T}}ST\left(T^{\mathrm{T}}T\right)^{-1}\right)\right] \tag{1}$$

where

$$\begin{aligned} S &= \frac{1}{n}\sum_{i=1}^{n}x_i x_i^{\mathrm{T}} - \frac{1}{n}\sum_{i,j=1}^{n}x_i x_j^{\mathrm{T}} \\ &= \frac{1}{2n}\sum_{i=1}^{n}\left(x_i - x_j\right)\left(x_i - x_j\right)^{\mathrm{T}} \end{aligned} \tag{2}$$

B. Graph-based Fisher Analysis (GbFA)

According to the graph theory, GbFA defines the intrinsic matrix W and the penalty matrix W'.

$$G_c = \sum_{i=1}^{n}\sum_{j=1}^{n}\left\|T^{\mathrm{T}}x_i - T^{\mathrm{T}}x_j\right\|^2 W_{ij} = 2tr\left\{T^{\mathrm{T}}XLXT\right\} \tag{3}$$

$$G_p = \sum_{i=1}^{n}\sum_{j=1}^{n}\left\|T^{\mathrm{T}}x_i - T^{\mathrm{T}}x_j\right\|^2 W'_{ij} = 2tr\left\{T^{\mathrm{T}}XL'XT\right\} \tag{4}$$

Where

$$W_{ij} = \begin{cases} \exp\left\{-\left\|-x_i - x_j\right\|^2/t\right\} & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases} \tag{5}$$

$$W'_{ij} = \begin{cases} \exp\left\{-\left\|-x_i - x_j\right\|^2/t\right\} & \text{if } y_i \neq y_j \\ 0 & \text{if } y_i = y_j \end{cases} \tag{6}$$

The objective function of GbFA is defined as follows:

$$\begin{aligned} T &= \arg\max_{T}\left(G_p\left(G_c\right)^{-1}\right) \\ &= \arg\max_{T}\left[tr\left(T^{\mathrm{T}}XL'XT\left(T^{\mathrm{T}}XLXT\right)^{-1}\right)\right] \end{aligned} \tag{7}$$

III. SEMI-SUPERVISED TENSOR GRAPH-OPTIMIZED LINEAR DISCRIMINANT ANALYSIS (STGLDA)

A. Objective function

According to these analyses, on the basis of PCA and GbFA, the paper proposed a Semi-supervised Tensor Graph-optimized Linear Discriminant Analysis (STGLDA) for face recognition.

Two-dimensional matrix face images are naturally represented by second-order tensors in the case of image sequences[25]. A two-dimensional matrix image is an element of the tensor space $R^{n_1} \otimes R^{n_2}$. Given training samples $X = \{x_i \mid x_i \in R^d\}_{i=1}^{n}$ that contain C classes, $n_i (1 \le i \le C)$ denotes the number of samples of the i-th class. According to Eq.(2) and Eq.(7), STGLDA aims to find two projecting U and V to get $Y = U^T XV = \{y_1, \ldots, y_n\}$, fusing PCA and GbFA with the linear weighted way under constrained conditions $U^{\mathrm{T}}U = I$ and $V^{\mathrm{T}}V = I$. The trace radio of objective function on STGLDA is defined as follows:

$$\left(U, V\right) = \arg\max_{U \in R^{n_1 \times l_1}, V \in R^{n_2 \times l_2}}\left(\hat{L}_b\left(\hat{L}_w\right)^{-1}\right) \tag{8}$$

where

$$\hat{L}_b = \beta\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left\|y_i - y_j\right\|^2 W'_{ij}\right) + \left(1-\beta\right)\frac{1}{2n}\sum_{i=1}\left\|\overline{y_i} - \overline{y}\right\|^2 \tag{9}$$

$$\hat{L}_w = \beta\frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{C_i}\left\|y_{ij} - \overline{y_i}\right\|^2 W_{ij} + \left(1-\beta\right)I \tag{10}$$

where β denotes the weighted trade-off parameter.

B. Theory Analysis

Since $\|A\|^2 = tr\left(AA^T\right)$, we can obtain

$$\begin{aligned} \hat{L}_b &= \beta\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left\|y_i - y_j\right\|^2 W'_{ij}\right) + \left(1-\beta\right)\frac{1}{2n}\sum_{i=1}\left\|\overline{y_i} - \overline{y}\right\|^2 \\ &= \beta\sum_{i=1}^{n}\sum_{j=1}^{n}\left(tr\left(\left(y_i - y_j\right)\left(y_i - y_j\right)^{\mathrm{T}}\right)W'_{ij}\right) \\ &\quad + \left(1-\beta\right)\frac{1}{2n}\sum_{i=1}tr\left(\left(\overline{y_i} - \overline{y}\right)\left(\overline{y_i} - \overline{y}\right)^{\mathrm{T}}\right) \\ &= \beta\sum_{i=1}^{n}\sum_{j=1}^{n}\left(tr\left(\left(U^{\mathrm{T}}x_iV - U^{\mathrm{T}}x_jV\right)\left(U^{\mathrm{T}}x_iV - U^{\mathrm{T}}x_jV\right)^{\mathrm{T}}\right)W'_{ij}\right) \\ &\quad + \left(1-\beta\right)\frac{1}{2n}\sum_{i=1}tr\left(\left(U^{\mathrm{T}}\overline{x_i}V - U^{\mathrm{T}}\overline{x}V\right)\left(U^{\mathrm{T}}\overline{x_i}V - U^{\mathrm{T}}\overline{x}V\right)^{\mathrm{T}}\right) \\ &= \beta\sum_{i=1}^{n}\sum_{j=1}^{n}tr\left(U^{\mathrm{T}}\left(x_i - x_j\right)VV^{\mathrm{T}}\left(x_i - x_j\right)^{\mathrm{T}}UW'_{ij}\right) \\ &\quad + \left(1-\beta\right)\frac{1}{2n}\sum_{i=1}tr\left(U^{\mathrm{T}}\left(\overline{x_i} - \overline{x}\right)VV^{\mathrm{T}}\left(\overline{x_i} - \overline{x}\right)^{\mathrm{T}}U\right) \\ &= tr\left[\beta\sum_{i=1}^{n}\sum_{j=1}^{n}\left(U^{\mathrm{T}}\left(x_i - x_j\right)VV^{\mathrm{T}}\left(x_i - x_j\right)^{\mathrm{T}}UW'_{ij}\right)\right] \\ &\quad + tr\left[\left(1-\beta\right)\frac{1}{2n}\sum_{i=1}U^{\mathrm{T}}\left(\overline{x_i} - \overline{x}\right)VV^{\mathrm{T}}\left(\overline{x_i} - \overline{x}\right)^{\mathrm{T}}U\right] \\ &= tr\left[\begin{array}{l} \beta\sum_{i=1}^{n}\sum_{j=1}^{n}\left(U^{\mathrm{T}}\left(x_i - x_j\right)\left(x_i - x_j\right)^{\mathrm{T}}W'_{ij}U\right) \\ +\left(1-\beta\right)\frac{1}{2n}\sum_{i=1}U^{\mathrm{T}}\left(\overline{x_i} - \overline{x}\right)\left(\overline{x_i} - \overline{x}\right)^{\mathrm{T}}U \end{array}\right] \\ &= tr\left[\begin{array}{l} \beta U^{\mathrm{T}}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left(x_i - x_j\right)\left(x_i - x_j\right)^{\mathrm{T}}W'_{ij}\right)\right)U \\ +\left(1-\beta\right)U^{\mathrm{T}}\left(\frac{1}{2n}\sum_{i=1}\left(\overline{x_i} - \overline{x}\right)\left(\overline{x_i} - \overline{x}\right)^{\mathrm{T}}\right)U \end{array}\right] \\ &= tr\left[U^{\mathrm{T}}\left(\beta S_b + \left(1-\beta\right)S\right)U\right] \end{aligned} \tag{11}$$

where $S_b = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left(x_i - x_j\right)\left(x_i - x_j\right)^{\mathrm{T}}W'_{ij}\right)$ and $S = \frac{1}{2n}\sum_{i=1}\left(\overline{x_i} - \overline{x}\right)\left(\overline{x_i} - \overline{x}\right)^{\mathrm{T}}$.

we can obtain

$$
\begin{aligned}
\hat{L}_w &= \beta \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\left\|y_{ij}-\overline{y_i}\right\|^2 W_{ij} + (1-\beta) I \\
&= \beta \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y_i}\right)\left(y_{ij}-\overline{y_i}\right)^T W_{ij} + (1-\beta) I \\
&= \beta \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\left(U^{\mathrm{T}} y_{ij} V - U^{\mathrm{T}}\overline{y_i} V\right)\left(U^{\mathrm{T}} y_{ij} V - U^{\mathrm{T}}\overline{y_i} V\right)^T W_{ij} + (1-\beta) I \\
&= \beta \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i} U^{\mathrm{T}}\left(x_{ij}-\overline{x_i}\right) V V^{\mathrm{T}}\left(x_{ij}-\overline{x_i}\right)^T U W_{ij} + (1-\beta) I \\
&= \beta \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i} U^{\mathrm{T}}\left(x_{ij}-\overline{x_i}\right)\left(x_{ij}-\overline{x_i}\right)^T U W_{ij} + (1-\beta) U^{\mathrm{T}} I U \\
&= U^{\mathrm{T}}\left[\beta \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\left(x_{ij}-\overline{x_i}\right)\left(x_{ij}-\overline{x_i}\right)^T W_{ij} + (1-\beta) I\right] U \\
&= U^{\mathrm{T}}\left[\beta S_W + (1-\beta) I\right] U
\end{aligned}
\tag{12}
$$

where $S_w = \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\left(x_{ij}-\overline{x_i}\right)\left(x_{ij}-\overline{x_i}\right)^T W_{ij}$.

Similarly, we have $\|A\|^2 = tr\left(A^T A\right)$ and also obtain

$$
\begin{aligned}
\hat{L}_b &= \beta \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left\|y_i - y_j\right\|^2 W'_{ij}\right) + (1-\beta)\frac{1}{2n}\sum_{i=1}^{n}\left\|\overline{y_i}-\overline{y}\right\|^2 \\
&= tr\left[V^{\mathrm{T}}\left(\beta S_b + (1-\beta) S\right) V\right]
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
\hat{L}_w &= \beta \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{C_i}\left\|y_{ij}-\overline{y_i}\right\|^2 W_{ij} + (1-\beta) I \\
&= V^{\mathrm{T}}\left[\beta S_W + (1-\beta) I\right] V
\end{aligned}
\tag{14}
$$

To ensure U and V converged, we iterate the procedure for several times until error conditions are satisfied.

C. Algorithm Steps

Input: Training samples $X = \{x_i \mid x_i \in R^{n_1 \times n_2}, 1 \le i \le n\}$ of the tensor space $R^{n_1} \otimes R^{n_2}$, error ε.

Output: two project matrixes $U(n_1 \times l_1)$ and $V(n_2 \times l_2)$.

Steps:

(1)Initial setting: $U_{(1)} = I(n_1, l_1)$, $V_{(1)} = I(n_2, l_2)$, $t = 1$, $U_{(1)}$ and $V_{(1)}$ denote respectively unit matrix, $X^v = \{x_i^v \mid x_i^v \in R^{n_1 \times l_2}, 1 \le i \le n\}$, $X^u = \{x_i^u \mid x_i^u \in R^{n_2 \times l_1}, 1 \le i \le n\}$, elements of X^v and X^u are zero.

(2)Set iteration circle variation t=1.

(3)Calculate $x_i^v = x_i V (1 \le i \le n)$.

(4) According to Eq. (5) and Eq.(6), calculate respectively W_{ij} and W'_{ij} in $X^v = \{x_1^v, \ldots, x_n^v\}$.

(5) According to Eq. (11) and Eq.(12), calculate respectively $\hat{L}_b$ and $\hat{L}_w$.

(6)Transform Eq.(8) into the generalized eigenvector problem to calculate $U_{(t)}$ and $U_{(t)}$ is normalized.

(7)Calculate $x_i^u = x_i^T U (1 \le i \le n)$.

(8) According to Eq. (5) and Eq.(6), calculate respectively W_{ij} and W'_{ij} in $X^u = \{x_1^u, \ldots, x_n^u\}$.

(9) According to Eq. (11) and Eq.(12), calculate respectively $\hat{L}_b$ and $\hat{L}_w$.

(10)Transform Eq.(8) into the generalized eigenvector problem to calculate $V_{(t)}$ and $V_{(t)}$ is normalized.

(11)If $\left\|U_{(t)} - U_{(t-1)}\right\| \prec \varepsilon$ and $\left\|V_{(t)} - V_{(t-1)}\right\| \prec \varepsilon$, then jump into step (12), else t =t +1 and jump into step (3).

(12)Obtain projecting matrix $U = U_{(t)}$ and $V = V_{(t)}$.

Ⅳ. EXPERIMENT

A. Experimental datasets

In the experiment, YaleB and AR face datasets are selected. They are described as followed:

(1) YaleB contains 2414 front-view face images of 38 individuals. For each individual, about 64 pictures were taken under various laboratory-controlled lighting conditions. In our experiments, we resize images to 32×32 pixels.

(2)AR consists of over 4000 face images of 126 individuals. For each individual, 26 pictures were taken in two sessions that separated by two weeks and each section contains 13 images, which include front view of faces with different expressions, illuminations and occlusions. In our experiment, we resize theses face images of AR to 30 ×30 pixels.

A group of face images on YaleB and AR are shown in Fig.1-Fig.2.

Figure 1. A group of face images on YaleB.

Figure 2. A group of face images on AR.

B. Experimental settings

In order to evaluate the performance of STGLDA, TPCA, TLPP, OTNPE and TLLDA are selected for making comparison. Parameter settings in various algorithms are shown in Table1.

TABLE I.
PARAMETER SETTINGS OF ALGORITHMS

Algorithms	Parameter settings
TPCA	no
STGLDA	$\kappa = 7$
TLLDA	$\kappa = 7$
OTNPE	$\kappa = 7$
TLPP	$\kappa = 7$

We repeated 40 times and the average of recognition accuracy is gotten as experimental results.

C. Experimental results

In experiments, the simplest nearest neighbor classification algorithm is adopted. Besides, we select randomly T images from each group face for training samples and remains for testing. Reduced dimensions(d

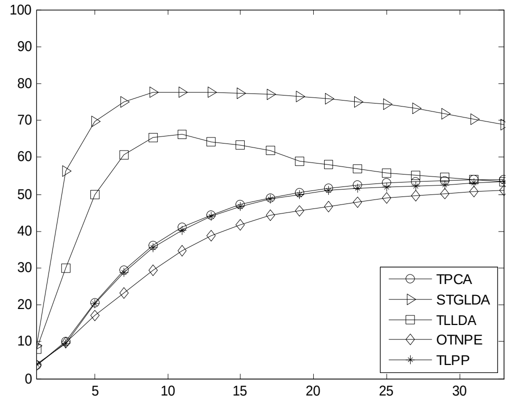

Figure 3. Recognition Accuracy VS. Reduced Dimensions(d×d) with L =8 on YaleB.

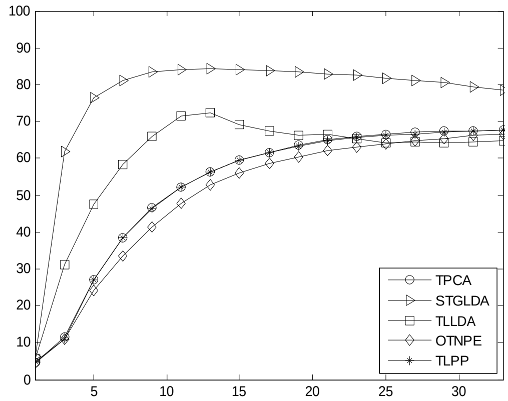

Figure 4. Recognition Accuracy (100%) VS. Reduced Dimensions(d×d) with L =16 on YaleB.

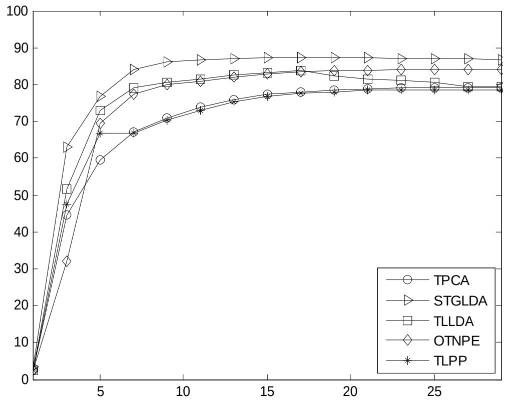

Figure 5. Recognition Accuracy VS. Reduced Dimensions(d×d) with L =7 on AR.

× d) are increased with the increment of two and corresponding classification accuracies are calculated. All experiments are repeated 40 times and average recognition accuracy is gotten as the experimental result.

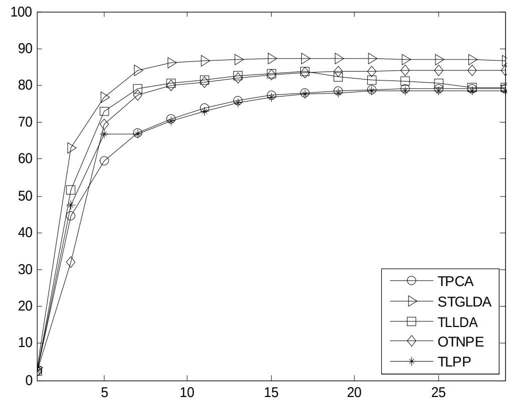

Figure 6. Recognition Accuracy VS. Reduced Dimensions(d×d) with L =14 on AR.

Experimental results on YaleB and AR are shown in Fig.3-Fig.6.

D. Experimental analysis

(1) The performance of TLLDA is superior to TPCA, TLPP and TLLDA owing to containing power discriminant information from TLDA and local structure information from TLPP. However, the recognition accuracy of STGLDA is great higher than in TLLDA on AR and YaleB with extern disturbs. The reason is that the radio of between-class discriminant information and within-class discriminant information defined in STGLDA has power robustness.

(2) TPCA performs much worse than STGLDA, which demonstrates that the way of the linear weighted way using Eq.(9) and Eq.(10) is efficient.

(3)Although TLPP and OTNPE have the good ability for capture local structure information, STGLDA has more performance of dimensionality reduction classification, which is explained that the intrinsic graph based on the same-class samples and the penalty graph based on the not-same-class samples defined in GbFA is inherited by STGLDA, encoding more discriminating information.

(4) STGLDA is superior to TLLDA. The reason is that TLLDA is originated from Local Fisher Discriminant Analysis (LFDA). In contrast with LFDA, GbFA has richer discriminant information. Moreover, TLLDA also ignores the global structure information of unlabeled samples.

V CONCLUSION

In the paper, a kind of dimensionality reduction algorithm named Semi-supervised Tensor Graph-optimized Linear Discriminant Analysis (STGLDA) for face recognition is proposed, which is based on Principal Component Analysis (PCA) and GbFA. The algorithm regards two-dimensional face images as a second-order matrix and fuses PCA and GbFA with linear weighted way, following the characteristic of tensor dimensionality reduction and preserving the global scatter structure and the enhanced discriminant information. Experimental results on YaleB and AR demonstrate the effectiveness of our algorithm.

However, each approach has its own advantages and disadvantages. PCA fails to capture global geometric characteristics. Moreover, the proposed algorithm is for two-dimensional face images. How to extend it to more order tensor to deal with data with more than three-dimension face images is also the future work.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their constructive comments and suggestions. The research is supported by NSF of China(grant No. 71171148) and NSF of Zhejiang province(grant No. LQ12F02007, Y201122544).

REFERENCES

[1] J.W. Long, X.J. Shen, and H.P. Chen, "A robust thresholding algorithm framework based on reconstruction and dimensionality reduction of the three dimensional histogram," *Journal of Computers,* vol.8, pp. 645-652, March 2013.

[2] I.T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 2002.

[3] S.J. Kim, A. Magnani and S.P. Boyd, "Robust fisher discriminant analysis," Advances in Neural Information Processing Systems, 2006: 659–666.

[4] W.S. Zheng, J. H. Lai, and S.S. Li, "1D-LDA vs. 2D-LDA: when is vector-based linear discriminant analysis better than matrix-based," *Pattern Recognition*, vol.41, pp. 2156–2172, July 2008.

[5] L.F. Chen, H.Y Mark Liao, M.T. Ko, J.C. Lin, and G.J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition,* vol. 33, pp. 1713–1726, October 2000.

[6] J. Yang, and J.Y. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recognition*, vol.36, pp. 563–566, February 2003.

[7] X.X. Zhang, and Y.D. Jia, "A linear discriminant analysis framework based on random subspace for face recognition," *Pattern Recognition*, vol.40, pp. 2585–2591, September 2007.

[8] S.W. Ji, and J.P. Ye, "Generalized linear discriminant analysis: a unified frame-work and efficient model selection," *IEEE Transactions on Neural Networks*, vol.19, pp. 1768–1782, October 2008.

[9] J.W. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letters*, vol.26, pp. 181–191, February 2005.

[10] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *The Annals of Statistics*, vol.23, pp. 73–102, January 1995.

[11] P. Howland, and H. Park, "Generalizing discriminant analysis using the generalized singular value decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 995–1006, August 2004.

[12] J.P. Ye, and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, pp. 929–941, June 2005.

[13] J.P. Ye, and T. Xiong, "Computational and theoretical analysis of null space and orthogonal linear discriminant analysis," *Journal of Machine Learning Research*, vol.7, pp. 1183–1204, December 2006.

[14] B. Yang, and Y.Y. Bu, "A comparative study on vector-based and matrix-based linear discriminant analysis," *Journal of Computers*, vol.6, pp. 818-824, April 2011.

[15] Y. Cui, and L.Y. Fan, "A novel supervised dimensionality reduction algorithm: Graph-based Fisher analysis," *Pattern Recognition*, vol.45, pp. 1471–1481, April 2012.

[16] M.F. Wang, "The new algorithm for finding the paths based on coding graph," *Journal of Software*, vol.7, pp.1-8, January 2012.

[17] E. Kerkouche, A. Chaoui, E.B. Bourennane, and O. Labbani, "On the use of graph transformation in the modeling and verification of dynamic behavior in UML models," *Journal of Software*, vol.5, pp. 1279-1291, November 2010.

[18] J. Chen, Y. Zhou, Y.J. Gao, B. Wang, L.B. Luo, and W.Y. Liu, "Effective improved graph transduction," *Journal of Software*, vol.6, pp. 1353-1360, July 2011.

[19] X.F. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," *Advances in Neural Information Processing Systems*, 2006.

[20] G. Dai, and D.Y. Yeung, "Tensor embedding methods," *Proceedings of the National Conference on Artificial Intelligence*, 2006: 330–335.

[21] Y.T. Wei, H. Li, and L.Q. Li, "Tensor locality sensitive discriminant analysis and its complexity," *International Journal of Wavelets Multiresolution and Information* Processing, vol.7, pp. 865–880, November 2009.

[22] Z. Zhang, and W.S. Chow, "Tensor locally linear discriminative analysis," *IEEE Signal Processing Letters*, vol.18, pp. 643–646, November 2011.

[23] Y.Q. Song, F.P. Nie, C.S. Zhang, and S.M. Xiang, "A unified framework for semi-supervised dimensionality reduction," *Pattern Recognition*, vol.41, pp.2789–2799, September 2008.

[24] S. Masashi, I. Tsuyoshi, N. Shinichi, and S. Jun, "Semi-Supervised local fisher discriminant analysis for dimensionality reduction," *Machine Learning*, vol.78, pp. 35–61, January 2010.

[25] Y.D. Zhang, S.H. Wang, and L.N. Wu, and Y.K. Huo, "Multi-channel Diffusion Tensor Image Registration via adaptive chaotic PSO," *Journal of Computers*, vol.6, pp. 825-829, April 2011.

Shuhua XU was born in Jiangxi, China, in 1977. She received her M.S. degree from Guizhou University, Guizhou, China, in 2004. Now she is an instructor of the School of Mathematics Science and physics Science. Her research interests include optimization theory and application, machine learning theory and pattern recognition.