Discriminative Semi-supervised Learning in Manifold Subspace for Face Recognition

Tue-Minh Dinh Vo, Hung Phuoc Truong, and Thai Hoang Le(✉)

Faculty of Information Technology,
University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
1112190@student.hcmus.edu.vn,
{tphung,lhthai}@fit.hcmus.edu.vn

Abstract. Linear Discriminant Analysis (LDA) is a commonly used method for dimensionality reduction, which preserves class separability. Despite its successes, it has limitations under some situations, including the small sample size problem. In practice, when the training data set is small, the covariance matrix of each class may not be accurately estimated. Moreover, LDA doesn't handle unlabeled data. In this paper, we propose a semi-supervised method called Discriminative Semi-supervised Learning in Manifold subspace (DSLM), which aims at overcoming all these limitations. The proposed method is designed to explore the discriminative information of labeled data and to preserve the intrinsic geometric structure of the data. We empirically compare our method with several related methods on face databases. Results are obtained from the experiments showing the effectiveness of our proposed method .

Keywords: Face recognition · Manifold · Semi-supervised

1 Introduction

In many areas of artificial intelligence, information retrieval, and data mining, one is often confronted with intrinsically low-dimensional data lying in a very high-dimensional space. This leads one to consider methods of dimensionality reduction that allow one to represent the data in a lower dimensional space. Two of the most popular techniques for this purpose are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

PCA is an unsupervised and an eigenvector method designed to model linear variation in high-dimensional data. PCA is guaranteed to discover the dimensionality of the subspace and produces a compact representation when the data is embedded in a linear subspace.

LDA is a supervised method. LDA searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. LDA encodes discriminating information in a linear separable space using bases are not necessarily orthogonal. When label information available, *e.g.* for classification task, LDA can achieve significant better performance than PCA. However, recent work [4] shows that when the training dataset is small, PCA can outperform LDA. The reason is covariance matrix of each class in

P.C. Vinh and V. Alagar (Eds.): ICCASA 2015, LNICST 165, pp. 243–253, 2016.
DOI: 10.1007/978-3-319-29236-6_24

LDA may not be accurately estimated. There are a lot of approaches that try to improve the performance of PCA and LDA, which are [1–3].

Recently, a number of research efforts have shown that the face images possibly reside on a nonlinear manifold [6, 10, 11, 16–18, 20–22]. Both PCA and LDA fail to discover the underlying structure when the face images lie on a manifold since they effectively see only the Euclidean structure. There has been some interest in the problem of developing low dimensional representations through kernel based techniques for face recognition [14,15]. These methods can discover the nonlinear structure of the face images. However, they are computationally expensive, and none of them explicitly considers the structure of the manifold on which the face images possibly reside. In the meantime, some nonlinear techniques have been proposed to discover the nonlinear structure of manifold, *e.g.* ISOMAP [13], LLE [6], Laplacian Eigenmap [12]. However, these nonlinear manifold learning techniques might not be suitable for face recognition since they do not generally provide a functional mapping between the high and low dimensional spaces that are valid both on and off the training data. There are a lot of approaches that try to address this issue by explicitly requiring an embedding function either linear or in reproducing kernel Hilbert space when minimizing the objective function [16–18]. One of the major limitations of these methods is that they fail to characterize the manifold structure of data when there are insufficient training samples. To solve this problem, many techniques have been proposed [19, 20] which have significantly improved the face recognition performance. However, these recognition algorithms struggle in achieving a reliable performance under more practical environments, where facial appearances are of large variations in illumination, expression, pose. An approach based on deep neural network has been proposed [5] to learn a nonlinear embedding from a high-dimensional data space to a low-dimensional space. However, this technique is computationally expensive and hard to determine the parameters.

In reality, we usually have small part of input data labeled, along with a large number of unlabeled data. Thus, semi-supervised learning has attracted an increasing amount of attention. Two well-known algorithms are extension of Support Vector Machine [21] and graph-based learning [10, 22]. Despite of their performance, it is unclear to determine the *good* graph.

In this paper, we propose a new semi-supervised dimensionality reduction algorithm, called Semi-supervised Learning in Manifold subspace (DSLM). Our proposed algorithm aims to find a projection which captures not only the discriminant structure inferred from the labeled data but also the intrinsic geometrical structure inferred from the whole training data. Specifically, the training data is used to build a graph incorporating neighborhood information in which each data point is represented as a linear combination of the neighboring data points. The graph provides a discrete approximation to the local geometry of the data manifold. In this way, DSLM can optimally preserves the manifold structure.

The rest of this paper is organized as follows: The Semi-supervised Learning in Manifold subspace (DSLM) algorithm is described in Sect. 2. A variety of experimental results are presented in Sect. 3. Section 4 discusses the effectiveness of our proposed algorithm. Finally, we provide some concluding remarks and suggestions for future work in Sect. 5.

2 Semi-supervised Learning in Manifold Subspace (DSLM)

2.1 The Objective Function

Suppose we have a set of n sample $\boldsymbol{X} = \{\boldsymbol{x_i}\}_{i=1}^{n}, \boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{D}}$ belonging to c classes. The basic idea of Linear Discriminant Analysis (LDA) is to seek directions on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. The objective function of LDA is as follow:

$$\boldsymbol{a}_{opt} = \arg\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T S_b \boldsymbol{a}}{\boldsymbol{a}^T S_w \boldsymbol{a}}. \tag{1}$$

where $\boldsymbol{S_w}$ is called the within-class scatter matrix and $\boldsymbol{S_b}$ is called the between-class scatter matrix. Define the total scatter matrix $\boldsymbol{S_t} = \boldsymbol{S_W} + \boldsymbol{S_b}$:

$$\boldsymbol{S_t} = \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^T. \tag{2}$$

where $\boldsymbol{\mu}$ is the total sample mean vector, n_k is the number of samples in the k-th class, $\boldsymbol{\mu^{(k)}}$ is the average vector of the k-th class, $\boldsymbol{x_i^{(k)}}$ is the i-th sample of the k-th class. Then the object function of LDA in Eq. (1) is equivalent to

$$\boldsymbol{a}_{opt} = \arg\max_{\boldsymbol{a}} \frac{\boldsymbol{a}^T \boldsymbol{S_b} \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{S_t} \boldsymbol{a}}. \tag{3}$$

We denote the matrix $\boldsymbol{X} = [\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(c)}]$ and the matrix $\boldsymbol{W_{LDA}}$ as

$$\boldsymbol{W_{LDA}} = \begin{bmatrix} \boldsymbol{W}^{(1)} & 0 & \cdots & 0 \\ 0 & \boldsymbol{W}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{W}^{(c)} \end{bmatrix}. \tag{4}$$

where $\boldsymbol{W^{(k)}}$ is a $n_k \times n_k$ matrix with all elements equal to $\frac{1}{n_k}$ and $\boldsymbol{X^{(k)}}$ is the data matrix of k-th class. The object function of LDA in Eq. (3) can be rewritten as [10]:

$$\boldsymbol{a}_{opt} = \arg\max_{a} \frac{\boldsymbol{a}^T \boldsymbol{X} \boldsymbol{W_{LDA}} \boldsymbol{X}^T \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{X} \boldsymbol{X}^T \boldsymbol{a}}. \tag{5}$$

When there is only one sample, LDA may be an ill-posed problem. When there is a small training set, overfitting may occur. The technique to solve those problem is regularization by introducing additional information. The optimization problem of regularized version of LDA can be written as follows [9]:

$$\max_{a} \frac{\boldsymbol{a}^T \boldsymbol{S}_b \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{S}_t \boldsymbol{a} + \alpha J(a)}. \tag{6}$$

where $J(a)$ controls the learning complexity of the hypothesis family, and the coefficient α controls balance between the model complexity and the empirical loss. The regularizer term $J(a)$ provides us the flexibility to incorporate our prior knowledge on some particular applications. The key to semi-supervised learning algorithm is the prior assumption of consistency. For classification, it means nearby points are likely to have the same label [7]. For dimensionally reduction, it means nearby points will have similar low-dimensional representations. Motivated by this intuition, we take advantage of the geometric properties of manifold patches. Specifically, if the data points lying on the same patch are likely to have the same label, which can be seen as Fig. 1.

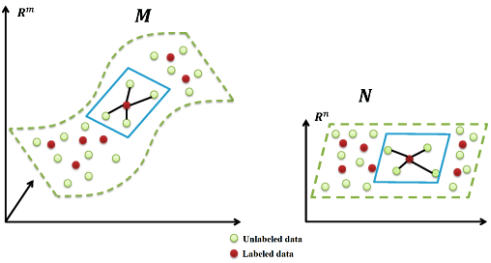

Fig. 1. Data points lie on same patch

Suppose $\boldsymbol{X}$ is from a smooth underlying manifold of dimensionality $d \ll D$. Each data points can be reconstructed from its neighbors with appropriate weights and these weights should be the same in low-dimensional space. Let $\boldsymbol{y_1}, \ldots, \boldsymbol{y_n} \in \mathbb{R}^{\boldsymbol{d}}$ be the corresponded mapped data. We have the cost function of a *good* map [6] under appropriate constraints as:

$$\Phi(\boldsymbol{y}) = \sum_i \left(\boldsymbol{y_i} - \sum_j \boldsymbol{W_{ij}} \boldsymbol{y_j} \right)^2. \tag{8}$$

which adds up the squared distances between all the data points and their reconstructions. $\boldsymbol{W_i}$ reveals the layout of the point around $\boldsymbol{x_i}$. Suppose the transformation is linear, that is, $\boldsymbol{y_i} = f(\boldsymbol{x_i}) = \boldsymbol{a}^T \boldsymbol{x_i}$. We define

$$\mathbf{z} = \boldsymbol{y} - \boldsymbol{W}\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{y}. \tag{9}$$

The cost function in Eq. (8) can be reduced to

$$\Phi(\boldsymbol{y}) = \sum_i \left(\boldsymbol{y_i} - \sum_j \boldsymbol{W_{ij}} \boldsymbol{y_j} \right)^2 = \sum_i (\mathbf{z_i})^2 = \boldsymbol{a}^T \boldsymbol{X} \boldsymbol{M} \boldsymbol{X}^T \boldsymbol{a}. \tag{10}$$

where $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W})$

Finally, we apply the approach of LDA and use the preserving local patches cost function as a regularizer term to make the objective function of DSLM:

$$\max_{a} \frac{\boldsymbol{a}^T \boldsymbol{S}_b \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{S}_t \boldsymbol{a} + \alpha J(a)} = \max_{a} \frac{\boldsymbol{a}^T \boldsymbol{S}_b \boldsymbol{a}}{\boldsymbol{a}^T (\boldsymbol{S}_t + \alpha \boldsymbol{X} \boldsymbol{M} \boldsymbol{X}^T) \boldsymbol{a}}. \tag{11}$$

Without loss of generality, we assume that the first n data points are labeled and ordered according to their labels. We use $\boldsymbol{X}_l = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l]$ to denote the labeled data matrix. We define the weight matrix $\boldsymbol{W} \in \mathbb{R}^{n \times n}$ as

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_{LDA} & 0 \\ 0 & 0 \end{bmatrix}, \tilde{\boldsymbol{I}} = \begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & 0 \end{bmatrix}.$$

where $\boldsymbol{W}_{LDA} \in \mathbb{R}^{l \times l}$ is defined in Eq. (4), I is an identity matrix of size $l \times l$.
We have

$$\boldsymbol{S}_b = \boldsymbol{X}_l \boldsymbol{W}_{LDA} \boldsymbol{X}_l^T = \boldsymbol{X} \boldsymbol{W} \boldsymbol{X}^T. \tag{12}$$

$$\boldsymbol{S}_t = \boldsymbol{X}_l \boldsymbol{X}_l^T = \boldsymbol{X} \tilde{\boldsymbol{I}} \boldsymbol{X}^T. \tag{13}$$

Then the objective function of DSLM in Eq. (11) can be rewritten as

$$\max_{a} \frac{\boldsymbol{a} \boldsymbol{X} \boldsymbol{W} \boldsymbol{X}^T \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{X} (\tilde{\boldsymbol{I}} + \alpha \boldsymbol{M}) \boldsymbol{X}^T \boldsymbol{a}}. \tag{14}$$

2.2 The Algorithm

Given data set $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$ includes labeled set $\boldsymbol{X}_l = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^l$ belonging to c classes and ordered according to their labels, and unlabeled set $\boldsymbol{X}_u = \{\boldsymbol{x}_i\}_{i=l+1}^n$. The k-th class have l_k samples, $\sum_{i=1}^{c} l_k = l$.

1. Construct the adjacency graph:
   In this step, we construct the adjacency graph G of all data set $\boldsymbol{X}$ by using the k-nearest neighbors method.
2. Compute the weights:
   In this step, we compute the weights on the edges of G. Let $\boldsymbol{W}$ be the weight matrix with W_{ij} having the weight of the edge from node i to node j, and 0 if there is no such edge. We define $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W})$ where $\boldsymbol{I}$ is the identity matrix of size $n \times n$.
   Please see [6] for details about how to compute $\boldsymbol{W}$.

3. Construct the graph for labeled data:
   In this step, we construct the weight matrix $\tilde{\boldsymbol{W}} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$ for labeled data

$$\tilde{\boldsymbol{W}} = \begin{bmatrix} \boldsymbol{W_l} & 0 \\ 0 & 0 \end{bmatrix}, \tilde{\boldsymbol{I}} = \begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & 0 \end{bmatrix}.$$

   where $\boldsymbol{W_l} \in \mathbb{R}^{\boldsymbol{l}\times\boldsymbol{l}}$ is defined in Eq. (9), $\boldsymbol{I}$ is an identity matrix of size $l \times l$.
4. Computing the projections:
   In this step, we compute the linear projections by solving the following generalized eigenvector problem

$$\boldsymbol{X}\tilde{W}\boldsymbol{X}^T\boldsymbol{a} = \lambda\boldsymbol{X}\left(\tilde{\boldsymbol{I}} + \alpha\boldsymbol{M}\right)\boldsymbol{X}^T\boldsymbol{a} \tag{16}$$

   It is easy to check that $\tilde{\boldsymbol{W}}$ is of rank c and we will have c eigenvectors with respect to non-zero eigenvalue [8]. Let $\boldsymbol{A} = [\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{c-1}]$ be the solution of Eq. (16), ordered according to their eigenvalues, $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{c-1} > 0$. A is a $n \times c$ matrix. The mapping subspace is as follows

$$\mathbf{x} \rightarrow \mathbf{z} = \boldsymbol{A}^T\mathbf{x}$$

3 Experimental Results

In this section, we investigate the use of our proposed approach for face recognition. We compare our DSLM algorithm with several representative dimension reduction algorithms, which include PCA, LDA, SDA [10]. PCA and LDA are the two most widely used subspace learning techniques for face recognition. SDA is the algorithm with high accuracy on semi-supervised face recognition [10].

3.1 Dataset Descriptions

The YALE face database contains 165 grayscale images of size 320 × 243 of 15 people (11 samples for person). The images demonstrate variations in lighting condition (left-light, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised and wink), and with/without glasses.

The ORL face database contains 400 gray images of size 92 × 112 of 40 people (10 samples for person). The images were captured at different times and have different variation including expressions (open or closed eyes, smiling or non-smiling) and face details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of face up to 20 degrees.

3.2 Data Preparation and Experimental Settings

In all the experiments, preprocessing to locate the faces was applying. 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 image for matching. The size of each cropped image in all the experiment is 32×32 pixels, with 256 gray levels per pixel. Thus, each image can be represented by 1024-dimesional vector in image space. No further preprocessing is done. 10 images of a person in YALE and 10 images of a person in ORL are displayed in Fig. 2.

(a) YALE face database (b) ORL face database

Fig. 2. Samples from YALE face database and ORL face database with different facial expression and details.

We use the semi-supervised setting for our experiments. That is, the available training set during the training phase contains both labeled and unlabeled examples, and the testing set is not available during the training phase. In this paper, we apply nearest-neighbor classifier for its simplicity. For each person in dataset, n images are randomly selected as the training set. Among these n images, l images are randomly selected and labeled which leaves other $n - l$ images unlabeled. We average the result over 25 random split. The recognition performance is measured by the accuracy:

$$Acc = \frac{Number\ of\ correctly\ classificated\ test\ samples}{Number\ of\ test\ samples} \times 100\,\%$$

3.3 Face Recognition with Different Dimensions

In this experiment, we fix $\alpha = 0.1$ for two methods SDA and DSLM. The number of nearest neighbors k is between 2 and 4, the recognition is carried out then. In general, the accuracy rates varies with the dimension of the face subspace. Figure 3 shows the plots of accuracy rates versus dimensionality reduction for the PCA, LDA, SDA and DSLM. The best result obtained in the optimal subspace and the corresponding dimensionality for each method are shown in Table 1. Note that the upper bound of dimensionality of SDA and DSLM is c where c is the number of classes. When there is a single labeled training image per class, LDA cannot be applied since the within-class

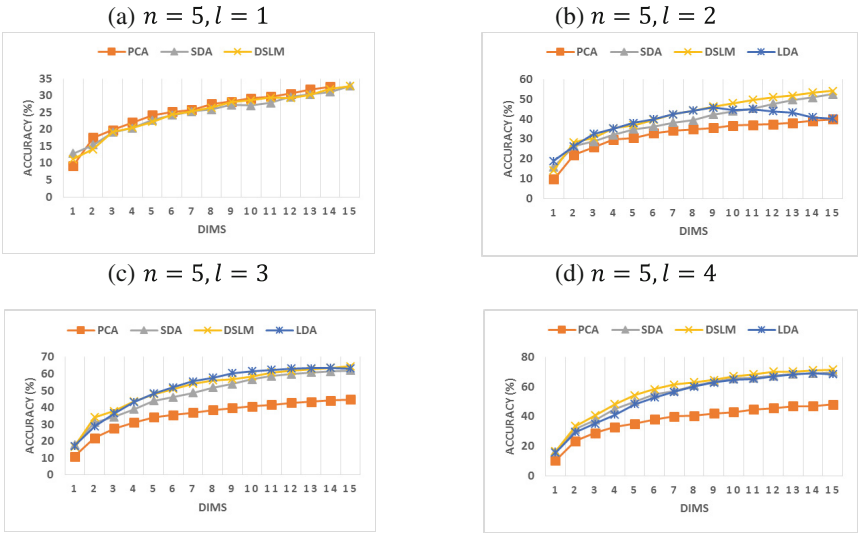

Fig. 3. Accuracy rates vs. dimensionality reduction on the YALE face database.

Table 1. Performance comparisons on the YALE face database

Method		PCA	LDA	SDA	DSLM
$n = 5$	$l = 1$	32.6 (14)	–	**32.8 (15)**	**32.8 (15)**
	$l = 2$	43.5 (29)	45.8 (9)5454	52.5 (15)	**54.2 (15)**
	$l = 3$	50.4 (44)	63.6 (14)	62.1 (15)	**64.6 (15)**
	$l = 4$	54.4 (59)	69.1 (14)	69.7 (15)	**71.6 (15)**

scatter matrix is the zero matrix. As can be seen, our DSLM algorithm performed the best for all the cases. Moreover, the optimal dimensionality obtained by DSLM, SDA and LDA is much lower than that obtained by PCA.

3.4 Face Recognition with Different k-Nearest Neighbors

The most important parameter in all of the manifold approaches which make use of the manifold structure is k-nearest neighbors. We test and compare two methods SDA and DSLM with different values of k. In this experiment, we use the ORL face database and fix $n = 7, l = 3$, $\alpha = 0.1$; k is chosen between 2 and 6; the recognition is carried out then. Figure 4 shows the plots of accuracy rates versus number nearest of neighbor. Table 2 shows the performance comparison of those. As can be seen, our DSLM algorithm performed better result. Moreover, the accuracy of our DSLM algorithm is stable with varying value of parameter k. It is shown that our DSLM algorithm is stability with varying size of patches on manifold.

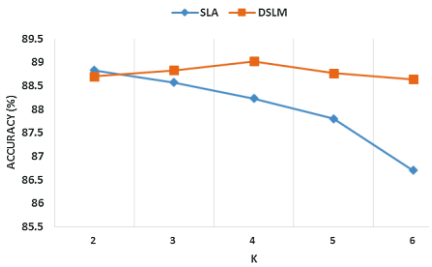

Fig. 4. Accuracy rates vs. *k*-nearest neighbor

Table 2. Performance comparison on ORL face database

k	SDA	DSLM
2	**88.83**	88.70
3	88.57	88.83
4	88.23	**89.02**
5	87.80	88.77
6	86.70	88.64

4 Discussion

It is worthwhile to high light several aspects of the proposed approach here:

1. Our proposed algorithm DSLM shares some similar properties with Semi-supervised Discriminant Analysis [10] algorithm. Both of them aim to find the optimal projection of the discriminative power of the labeled data and of the locality preserving power of manifold. However, their graphs which discover manifold structure are totally different. Thus, their objective functions are different.
2. Some manifold learning algorithms like ISOMAP, LLE, Laplacian eigenmaps are defined only on the training data points and it is unclear how to evaluate the map for new test points. DSLM can find the optimal linear projection. Thus, this makes it fast and suitable for practical applications, e.g. face recognition.
3. DSLM can be performed and product significant results in small datasets which cannot be achieved by LDA, which can be seen as experimental results.

5 Conclusion

In this paper, we proposed a new linear dimensionality reduction algorithm called Discriminative Semi-supervised Learning in Manifold subspace. By using a graph which characterizes the locality structure of manifold data and taking advance of discriminative power of LDA method, our algorithm can make use of both labeled data and unlabeled data points to find optimal projection. Experimental results on face recognition have demonstrated the effectiveness of our algorithm.

For future works, we are interested in applying the proposed method to other graphs which characterize better the geometric properties of the dataset. On the other hand, the algorithm should be investigated in supervised mode.

Acknowledgements. This research is supported by research funding from Science Research funding (T-2015.21) and Honors Program, University of Science, Vietnam National University - Ho Chi Minh City.

References

1. Truong, H., Le, T.: Fusion of bidirectional image matrices and 2D-LDA: an efficient approach for face recognition. In: Proceedings of Third Symposium on Information and Communication Technology, pp. 142–148. ACM, New York (2012)
2. Le, T., Truong, H., Do, H., Vo, D.: On approaching 2D-FPCA technique to improve image representation in frequency domain. In: Proceedings of Fourth Symposium on Information and Communication Technology, pp. 172–180. ACM, New York (2013)
3. Le, D.K., Truong, H., Le, T.: Facial expression recognition using statistical subspace. In: IEEE International Conference on Image Processing, pp. 5981–5985. IEEE (2014)
4. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. **23** (2), 228–233 (2001)
5. Huang, Y., Wang, W., Wang, L., Tan, T.: A general nonlinear embedding framework based on deep neural network. In: ICPR, pp. 732–737 (2014)
6. Roweis, S., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**, 2323–2326 (2000)
7. Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: Advances in Neural Information Processing Systems, vol. 16 (2003)
8. Golub, G.H., Loan, C.F.V.: Matrix computations. Johns Hopkins University Press, Maryland (1996)
9. Friedman, J.H.: Regularized discriminant analysis. J. Am. Stat. Assoc. **84**(405), 165–175 (1989)
10. Cai, D., He, X., Han, J.: Semi-supervised discriminant analysis. In: International Conference on Pattern Recognition, pp. 1–7 (2007)
11. Chang, Y., Hu, C., Turk, M.: Manifold of facial expression. In: Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures. France (2003)
12. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Advances in Neural Information Processing System, vol. 15. Canada (2001)
13. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**, 2319–2323 (2000)
14. Liu, Q., Huang, R., Lu, H., Ma, S.: Face recognition using kernel based fisher discriminant analysis. In: 5th FG. Washington, D.C. (2002)
15. Roweis, S., Saul, L., Hinton, G.: Global coordination of local linear models. In: Advances in Neural Information Processing System, vol. 14 (2001)
16. Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., Lin, S.: Graph embedding and extension: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
17. He, X., Niyogi, P.: Locality preserving projections. In: Advances in Neural Information Processing Systems, vol. 16. MIT Press, Cambridge (2003)
18. Chen, X., Wei, J., Li, J., Zhang, X.: Integrating local and global manifold structures for unsupervised dimensionality reduction. In: IJCNN. China (2014)
19. Lu, J., Tan, Y.P.: Cost-sensitive subspace learning for face recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition, pp. 2661–2666 (2010)
20. Lu, J., Tan, Y.P., Wang, G.: Discriminative multi-manifold analysis for face recognition from a single training sample per person. In: ICCV, pp. 1943–1950 (2011)

21. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: a geometric framework for learning from examples. J. Mach. Learn. Res. **7**, 2399–2434 (2006)
22. Goldberg, A.B., Li, M., Zhu, X.: Online manifold regularization: a new learning setting and empirical study. In: Proceedings of 19th European Conference on Machine Learning, pp. 393–407. Belgium (2008)