Pattern Recognition 43 (2010) 3359–3363

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Two-dimensional supervised local similarity and diversity projection

Quan-Xue Gao*, Hui Xu, Yi-Ying Li, De-Yan Xie

State Key Laboratory of Integrated Services Networks, Xi Dian University, Xi'an 710071, P.R. China

A R T I C L E I N F O

Article history:
Received 7 July 2009
Received in revised form
9 April 2010
Accepted 13 May 2010

Keywords:
Manifold learning
Feature extraction
Diversity
2DLPP
Face recognition

A B S T R A C T

This paper presents a novel manifold learning method, namely two-dimensional supervised local similarity and diversity projection (2DSLSDP), for feature extraction. The proposed method defines two weighted adjacency graphs, namely similarity graph and diversity graph. The affinity matrix of similarity graph is determined by the spatial relationship between vertices of this graph, while affinity matrix of diversity graph is determined by the diversity information of vertices of its graph. Using these two graphs, the proposed method constructs local similarity scatter and diversity scatter, respectively. A concise feature extraction criterion is then raised via minimizing the ratio of the local similarity scatter to local diversity scatter. Thus, 2DSLSDP can well preserve not only the adjacency similarity structure, but also the diversity of data points, which is important for the classification. Experiments on the AR and UMIST databases show the effectiveness of the proposed method.

1. Introduction

Feature extraction is one of the fundamental problems in many areas, such as multivariable analysis, pattern recognition, computer vision, and machine learning. The aim of it is to seek the essential and meaningful features embedded in high-dimensional original data space, which can be useful for subsequent analysis. So far, many approaches have been developed for feature extraction. These methods can be roughly classified into two categories: unsupervised and supervised approaches.

Unsupervised feature extraction methods seek to find features which are useful and convenient for pattern representation. It is performed by seeking for the most representative low-dimensional and compact subspace for original high-dimensional data. One of the most widely used techniques in this category is principal component analysis (PCA). Supervised feature extraction methods, on the other hand, seek features which have good discriminant ability for classification. It is performed by seeking the most discriminative subspace in which the data from different classes can be clearly separated. One of the most representative techniques in this category is linear discriminant analysis (LDA). However, both PCA and LDA mainly rely on the assumption that samples satisfy the Gaussian distribution or reside on linear manifolds; thus, they suffer performance degradation in case of non-Gaussian distribution or nonlinear manifolds.

Recently, many studies have indicated that natural images [1–3], especially face images, possibly reside on or are close to low-dimensional nonlinear sub-manifold embedded in the original high-dimensional data space. Under this circumstance, PCA and LDA will fail to discover the essential and meaningful structure underlying the data points, thus leading to unsatisfactory performance. A large volume of manifold learning methods have been proposed to address this problem [1–10], the most representative methods are some unsupervised methods like laplacian eigenmap (LE) [3], locality preserving projections (LPP) [4], neighborhood preserving embedding (NPE) [5], and isometric projection [6], and some supervised methods such as supervised LPP (SLPP) [4], local discriminating projection (LDP) [8], and marginal fisher analysis (MFA) [10]. While all these methods have to rearrange the images from matrixes to vectors and are thus highly possible to lose useful structure information among image pixels, their two-dimensional counterparts directly apply to the two-dimensional images. Typical examples of such two-dimensional manifold learning methods include two-dimensional locality preserving projection (2DLPP) [7], two-dimensional discriminant locality preserving projections (2DDLPP) [9], and two-dimensional marginal fisher analysis (2DMFA). These methods have proven to be very effective for pattern classification and recognition applications.

2DLPP, which is one of the most representative two manifold learning techniques, in essence seeks to preserve the local similarity structure among data points, and achieves this goal via minimizing the ratio of the weighted local scatter to weighted global scatter [7], i.e. keeping the close data points as close as possible after projection. As an extreme case, it would prefer to compact all data points (possibly from the same class) to one single point. Obviously, this will make the obtained projection matrix overfit to the training data. Another drawback of 2DLPP is

* Corresponding author.
E-mail address: qxgao@xidian.edu.cn (Q.-X. Gao).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.05.017

that it cannot well preserve the diversity information of data points due to the adjacency similarity graph. However, the diversity of data points is important for pattern classification from the perspective of statistic.

What's more, the above-mentioned over-fitting problem of 2DLPP can be also thought of as a problem of impairment of the diversity of data. More specifically, the intra-class variations in the prototypes of a class in the feature space can also contribute to the robust recognition of the class. Here, we refer to such intra-class variations as the diversity of the data. Inspired by the manifold learning, a novel manifold learning approach, called two-dimensional supervised local similarity and diversity projection (2DSLSDP), is proposed to address the shortcoming of 2DLPP. 2DSLSDP defines two adjacency graphs: similarity graph and diversity graph. The affinity matrix of adjacency similarity graph is determined by the adjacency spatial relationship of vertices and measures the adjacency similarity structure of data; while affinity matrix of diversity graph embodies the diversity information of vertices of diversity graph. Two weighted scatters, namely similarity scatter and diversity scatter, are calculated from the two graphs, respectively, and a concise feature extraction criterion is then built by minimizing the ratio of the similarity scatter to diversity scatter. Thus, 2DSLSDP can well preserve both adjacency similarity inherent structure and diversity of data points. Experiments on the AR and UMIST databases show the effectiveness of the proposed method and its superiority over other two-dimensional feature extraction methods, including supervised and discriminant methods, considered in the experiments.

The rest of this paper is organized as follows: in Section 2, we review 2DLPP and discuss its shortcoming in detail. In Section 3, we present a novel manifold learning method, namely 2DSLSDP. Section 4 provides the summary of 2DSLSDP. The performance of 2DSLSDP is evaluated on the AR and UMIST databases in Section 5. The paper is concluded in Section 6 with discussion on future work.

2. Review of 2DLPP

Given N training images $A_i \in \mathbb{R}^{m\times n}(i=1,2,\ldots,N)$, where A_i denotes the ith training image. The spatial relationship between images can be modeled by a weighted graph $G_L=\{X,H\}$, namely adjacency similarity structure graph, with a vertex set $X=[A_1\ A_2\ \cdots\ A_N]$ and an affinity matrix $H\in\mathbb{R}^{N\times N}$. Then the objective function of 2DLPP can be rewritten as follows [7,11,12]:

$$w^* = \arg\min \frac{w^T X(L\otimes I_n)X^T w}{w^T X(D\otimes I_n)X^T w} \tag{1}$$

where w denotes the projection map, $\otimes$ the Kronecker product, and T the transpose operator. $L=D-H$ denotes the Laplacian matrix. D is a diagonal matrix whose elements on diagonal are row sum of H, i.e.$D_{ii}=\Sigma_j H_{ij}$. If A_i is among k nearest neighbors of A_j or A_j is among k nearest neighbors of A_i, then $H_{ij}=\exp\left(-\|A_i-A_j\|_F^2/t\right)$, otherwise, $H_{ij}=0$. $\|\bullet\|_F$ is Frobenius norm.

2DLPP efficiently preserves the local adjacency structure of data points (here, image samples) via Eq. (1). In order for a deep insight into the criterion of 2DLPP, we can rewrite the term $X(D\otimes I_n)X^T$ in the denominator of Eq. (1) as

$$X(D\otimes I_n)X^T = \sum_{i=1}^{N} D_{ii}A_iA_i^T \tag{2}$$

From Eq. (2), it can be clearly seen that $X(D\otimes I_n)X^T$ is a weight global scatter matrix, D_{ii} denotes the weight coefficients corresponding to image A_i and is determined by the affinity matrix H.

According to the definition of H_{ij}, if A_i and A_j is very close to each other, then H_{ij} is very large. Thus, if samples among the k nearest neighbors of A_i are very close to A_i, then D_{ii} is large. It means that sample A_i has a big contribution in estimating the total scatter matrix in Eq. (2). Fig. 1 shows some example two-dimensional data points on a plane. Following the above criterion, it is easy to know that the red point has a large weight in calculating the total scatter matrix because data points among its $k=3$ nearest neighbors are very close to it. On the other hand, the green point will be assigned with a small weight because its $k=3$ nearest neighbors have a sparser distribution around it. From the similarity viewpoint, such weight assignment is reasonable.

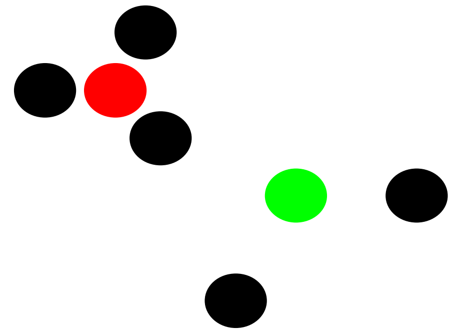

Fig. 1. An illustrative example of local adjacency-similarity and diversity. The red point carries more similarity information of its pattern, while the green point carries more diversity information of its pattern. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

However, from the perspective of statistic, if considering the diversity (i.e. the intra-class variations) of the data, such method becomes problematic. Still take the red and green points in Fig. 1 as an example. One can easily argue that the red point is more homogeneous to its neighbors than the green point and thus it conveys less distinctive information of its class. On the contrary, the green point would be a very important representative of its pattern. Keeping such representatives during the projection is expected to make the resulting classification system more robust and more generative. Therefore, data points like the green point in Fig. 1 should deserve large weights in calculating the total scatter matrix in Eq. (2). Unfortunately, referring to the definition of D_{ii}, it can be seen that the 2DLPP does not conform to this.

From the aforementioned discussion, we would like to raise the following remark on 2DLPP:

Remark: 2DLPP preserves the local adjacency similarity structure of data at the cost of losing some diversity information of the data which may be useful for classification.

Note that this remark is not limited to the 2DLPP method, although it is derived from the criterion of 2DLPP. However, this is not the topic of this paper. More thorough discussion on this will be reported in our future work. Another shortcoming of 2DLPP is that it as an unsupervised method does not use the class labels of images which can be very useful for discriminative feature learning and pattern classification. In the next section, we will propose a new feature extraction criterion, which considers both local similarity and diversity of data and explores the class labels given for the data. A novel feature extraction algorithm will be also presented to implement the proposed criterion.

3. The proposed method

3.1. Learning the diversity of data points

Given N training images $A_i \in \mathbb{R}^{m\times n}(i=1,2,\ldots,N)$. In order to capture the diversity of data in learning the projection, we encode it by weighted graph $G_d=\{X,B\}$, namely adjacency diversity graph to differ from the adjacency similarity graph mentioned in Section 2, with affinity matrix. The element B_{ij} in its affinity matrix B measures the diversity information conveyed by vertices A_i and A_j in graph G_d. It is defined as follows from the perspective of statistic:

$$B_{ij} = \begin{cases} \exp\left(-t_1/\|A_i - A_j\|_F^2\right), & \text{if } A_j \text{ is among } k_1 \text{ nearest neighbors of } A_i \\ 0, & \text{Otherwise} \end{cases} \tag{3}$$

Thus, for the purpose of preserving the diversity of the data during the projection, we impose the following objective:

$$\max tr\left\{ \sum_{ij} B_{ij}(Y_i - Y_j)(Y_i - Y_j)^T \right\} \tag{4}$$

where $Y_i \in R^{1\times n}$ is a low-dimensional representation of image A_i after projection, and $tr\{\cdot\}$ denotes the trace of matrix. $tr\left\{\sum_{ij} B_{ij}(Y_i - Y_j)(Y_i - Y_j)^T\right\}$is called the local diversity scatter.

Eq. (4) with our choice of symmetric weights B_{ij} defined by Eq. (3) incurs a heavy penalty if two points A_i and A_j, which are far apart in original image space are mapped very close to each other, i.e. if $\|Y_i - Y_j\|^2$ is very small. Therefore, maximizing it is an attempt to ensure that if diversity information between A_i and A_j is large, then the diversity information of Y_i and Y_j is also large. It is just the physical meaning of Eq. (4). From the perspective of statistic, Eq. (4) seeks to find the low-dimensional compact subspace for data points, which efficiently preserves the most diversity information of data points.

3.2. *Feature extraction criterion*

Our aim is to preserve both adjacency similarity structure and diversity structure of data points. From this point of view, a desirable projection should be the one that, at the same time, minimizes the weighted local similarity scatter, which is determined by adjacency similarity graph, and maximizes the weighted diversity scatter, which is determined by the diversity graph. As it happens, we can obtain just such a projection by the following criterion:

$$w^* = \arg\ \min \left(\frac{tr\left\{\sum_{ij} H_{ij}(Y_i - Y_j)(Y_i - Y_j)^T\right\}}{tr\left\{\sum_{ij} B_{ij}(Y_i - Y_j)(Y_i - Y_j)^T\right\}} \right) \tag{5}$$

where $tr\left(\sum_{ij} H_{ij}(Y_i - Y_j)(Y_i - Y_j)^T\right)$ denotes the local similarity scatter, $\{H_{ij}\}$ the affinity matrix of the adjacency similarity graph G_L. In order to better preserve the manifold structure of data points and explore the available class information of the data [8], the elements H_{ij} of affinity matrix are calculated in a supervised manner as follows:

$$H_{ij} = \begin{cases} \exp(-\|A_i - A_j\|_F^2/t)(1 + \exp(-\|A_i - A_j\|_F^2/t)), & \text{if } A_i \in \tau_i \text{ is among } k \text{ nearest neighbors of } A_j \in \tau_j \\ & \text{or } A_j \text{ is among } k \text{ nearest neighbors of } A_i \text{ and } \tau_i = \tau_j \\ \exp(-\|A_i - A_j\|_F^2/t)(1 + \exp(-\|A_i - A_j\|_F^2/t)), & \text{if } A_i \in \tau_i \text{ is among } k \text{ nearest neighbors of } A_j \in \tau_j \\ & \text{or } A_j \text{ is among } k \text{ nearest neighbors of } A_i \text{ and } \tau_i \neq \tau_j \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where τ_i denotes the class label of the ith image.

Substituting $Y_i = w^T A_i$ into the denominator of Eq. (5), then denominator can be rewritten as

$$\begin{aligned} &\frac{1}{2} tr\left(\sum_{ij} B_{ij}(Y_i - Y_j)(Y_i - Y_j)^T \right) \\ &= \frac{1}{2} w^T \left[\sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij}(A_i - A_j)(A_i - A_j)^T \right] w \\ &= w^T \left[\sum_{i=1}^{N} A_i (A_i)^T \sum_{j=1}^{N} B_{ij} - \sum_{i=1}^{N} A_i \sum_{j=1}^{N} B_{ij}(A_j)^T \right] w \\ &= w^T \left[\sum_{i=1}^{N} A_i (A_i)^T F_{ii} - \sum_{i=1}^{N} A_i \sum_{j=1}^{N} B_{ij}(A_j)^T \right] w \\ &= w^T \left[X(F \otimes I_n)X^T - X(B \otimes I_n)X^T \right] w = w^T X(L_d \otimes I_n)X^T w \end{aligned} \tag{7}$$

where $L_d = F - B$, B is an $N \times N$ symmetric matrix, F a diagonal matrix whose elements on diagonal are column sum of B, i.e. $F_{ii} = \Sigma_j B_{ij}$.

Likewise, substituting $Y_i = w^T A_i$ into the numerator of Eq. (5), we have

$$\frac{1}{2} tr\left(\sum_{ij} H_{ij}(Y_i - Y_j)(Y_i - Y_j)^T \right) = w^T X(L \otimes I_n)X^T w \tag{8}$$

Substituting Eq. (7) and Eq. (8) into Eq. (5), the feature extraction criterion becomes

$$w^* = \arg\min \left(\frac{w^T S_s w}{w^T S_d w} \right) \tag{9}$$

where $S_s = X(L \otimes I_n)X^T$ and $S_d = X(L_d \otimes I_n)X^T$ denote the weighted similarity and diversity scatter matrices constructed from the adjacency similarity and diversity graphs. They measure the similarity and diversity information of data, respectively. In Eq. (9), S_s is calculated by using the class label of images. So this criterion is a supervised feature extraction criterion.

The optimal projection w^* of Eq. (9) is the eigenvector of the generalized eigen-equation $S_s w = \lambda S_d w$, corresponding to the non-zero smallest eigenvalue λ. In practical applications, it usually needs $d (d \geq 2)$ projection directions. Denote $W \in \mathbb{R}^{m \times d}$ by the optimal projection matrix, then the columns of W are composed of the d orthonormal eigenvectors corresponding to the d non-zero smallest eigenvalues in terms of matrix theory.

After obtaining the optimal projection matrix W, the features can be obtained by projecting the images onto W. Denote Y_j and Y^* by the projected features of $A_j (j = 1, \ldots, N)$ and a probe image A^*, respectively, classification can then be realized by using the dissimilarity between Y^* and Y_j and nearest neighbor classifier. The dissimilarity between Y^* and Y_j can be defined as

$$d(Y^*, Y_j) = \sqrt{\sum_{i=1}^{n} \|y_i^* - y_i^j\|_2} \tag{10}$$

where y_i^* and y_i^j denote the ith column of Y^* and Y_j, respectively. $\|\bullet\|_2$ denotes the Euclidean distance. If $d(Y^*, Y_p) = \min_j d(Y^*, Y_j)$ and Y_p belongs to class τ_p, then A^* is assigned to the class τ_p, i.e. $A^* \in \tau_p$.

4. 2DSLSDP algorithm

In summary of the preceding description, the following provides the 2DSLSDP algorithm:

Step 1. *Constructing adjacency graphs*: let G_d and G_L denote two undirected graphs both over all data points. To construct G_L, if A_j is among the k nearest neighbors of A_i, then a directed edge is added from node i to j. To construct G_d, if A_j is among the k_1 nearest neighbors of A_i, then a directed edge is added from node i to j. Note that we also add a directed edge between node i and j if $\|A_i - A_j\|_F \le \varepsilon$. Where ε is sufficiently small, and $\varepsilon > 0$.

Step 2. *Computing the weights*: in this step, we compute the weights on the edge of two graphs G_d and G_L. Specify the affinity matrix H of G_L, the elements H_{ij} can be given by Eq. (6) and measures the local similarity of data points. For the affinity matrix B of G_d, the elements B_{ij} can be given by Eq. (3) and measures the diversity information of data points. Note that the elements B_{ij} can also be defined as $1 - H_{ij}$ if A_j is among the k nearest neighbors of A_i, or A_i is among the k nearest neighbors of A_j, otherwise, $B_{ij} = 0$.

Step 3. *Computing the projections*: in this step, we compute the linear projections w_i $(i = 1, 2, \ldots, d)$ which are the eigenvectors of the generalized Eigen-function, i.e. $X(L \otimes I_n)X^T w_i = \lambda_i X(L_d \otimes I_n)X^T w_i$, corresponding to the ith non-zero smallest eigenvalue λ_i, where $X = [A_1\ A_2\ \cdots\ A_N]$ denotes the data matrix, A_i is the ith image.

Step 4. *Extracting features*: in this step, low-dimensional representation $Y_i (i = 1, \ldots, N)$ and Y^* of the training image A_i and probe image A^* can be obtained by projecting A_i and A^* onto $W = [w_1\, w_2\ \cdots\ w_d]$, which is the optimal projection matrix.

Step 5. *Classifying probe images*: in this step, the classification of probe images can be realized by using the dissimilarity defined in Eq. (10) and the nearest neighbor classifier.

5. Experiments and analysis

The performance of the proposed algorithm 2DSLSDP has been evaluated on the AR and UMIST face databases. In the experiments, we compare 2DSLSDP with the most representative two-dimensional feature extraction methods, including unsupervised methods such as 2DPCA [13] and 2DLPP [7], and supervised methods, like 2DLDP (the two-dimensional version of LDP) [8], 2DLDA [14], 2DDLPP [9], and 2DMFA [10].

'The AR database (http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html) contains over 4000 color face images from 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of most persons were taken in two sessions, separated by two weeks. Each session contains 13 color images per person and 120 individuals (65 men and 55 women) participate in both sessions. In this database, the facial portion of each image was manually cropped and then normalized to the size of 50×40 [13]. The images from the first session with "neutral expression", "smile", "anger", "scream", "left light on", "right light on", and "both side light on" are selected for training images. The corresponding images from the second session are used for testing.

The UMIST database (http://images.ee.umist.ac.uk/danny/database.html) is a multi-view database, consisting of 575 images from 20 people and covering a wide range of poses from profile to frontal views. In this database, the first 19 images per class are used, among which the first 6 images per class are used for training and the remaining images for testing.

Table 1 lists the top recognition accuracy of these seven methods and the associated number of features on the AR and UMIST databases. The cures of recognition accuracies versus the number of projected vectors on the AR and UMIST databases are plotted in Figs. 2 and 3, respectively. From Table 1, Fig. 2, and Fig. 3, we can see the conclusions as follows:

First, comparing the two unsupervised methods, 2DLPP and 2DPCA, it is easy to find that 2DLPP outperforms 2DPCA in the recognition accuracy, especially on the UMIST database. The main reason may be that manifold learning method preserves the local intrinsic structure, which is robust to outliers and noise. Compared to the training and testing images in UMSIT database, it is easy to find that the pose of training image and testing images is different, while the variation in face pose may be viewed as outliers, which may cause the intrinsic structure of data points to embed in nonlinear manifold rather than linear manifold, so the global Euclidean structure preserved by 2DPCA does not efficiently reflect the essential structure of data. What's more, global

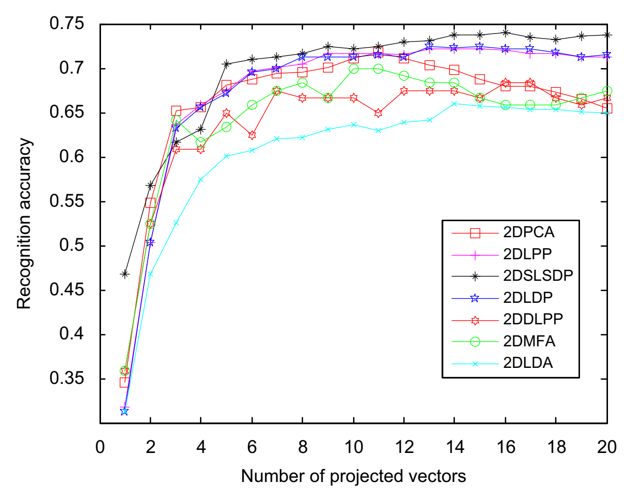

Fig. 2. The recognition accuracy versus the number of projected vectors of different methods on the AR database.

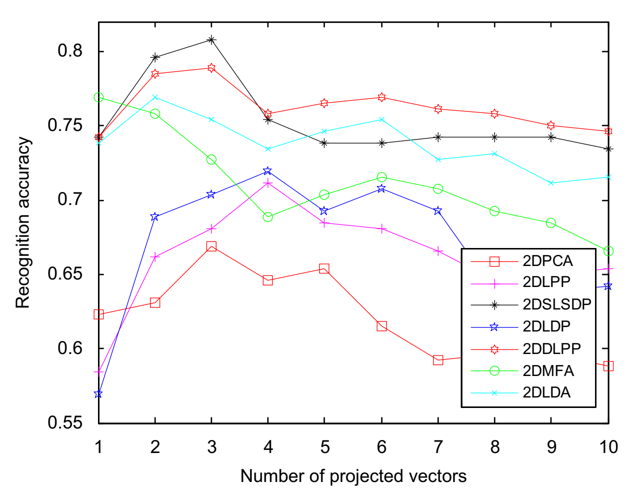

Fig. 3. The recognition accuracy versus the number of projected vectors of different methods on the UMIST database.

Table 1
Top recognition accuracy (%) of different methods on the AR and UMIST databases. The values in parentheses are the corresponding number of features.

Databases	2DPCA	2DLPP	2DLDA	2DDLPP	2DMFA	2DLDP	2DSLSDP
AR	71.79 (440)	72.14 (520)	65.95 (560)	68.33 (640)	70.00 (400)	72.50 (600)	74.05 (640)
UMIST	66.92 (276)	71.15 (368)	76.92 (184)	78.85 (276)	76.92 (92)	71.92 (368)	80.77 (276)

structure preserved by 2DPCA is not better to measure the information of data points under the nonlinear manifold.

Second, comparing the supervised methods 2DSLSDP and 2DLDP and the unsupervised methods 2DPCA and 2DLPP, it is easy to find that 2DSLSDP and 2DLDP outperform 2DPCA and 2DLPP at their best recognition accuracy. The reason may be that the samples among the *k* nearest neighbors of one image do not always belong to the same class in general. So the local similarity structure preserved by 2DLPP is not efficient enough for classification. The proposed method 2DSLSDP is obviously superior to the other methods, both supervise and unsupervised approaches. The reason may be that 2DSLSDP preserves not only the adjacency similarity between data points but also the diversity among data points. In contrast, none of the other methods, including supervised and unsupervised manifold learning approaches, can well deal with these two kinds of information simultaneously. For example, unsupervised method 2DPCA preserves the global Euclidean structure of data which is unsuitable to measure the inherent structure of nonlinear manifold and is poor at preserving the local similarity information of data, whereas manifold learning methods 2DLPP and 2DLDP can well preserve the local similarity of data, but they are highly possible to impair the diversity of data.

Third, the top recognition accuracy of 2DSLSDP is obviously superior to the discriminant methods 2DLDA, 2DDLPP, and 2DMFA when suitable parameters are adopted. The main reason may be that 2DDLPP and 2DMFA may overfit the training data; differently, 2DSLSDP can overcome this drawback due to the diversity graph it constructs. Moreover, 2DSLSDP is also superior to 2DLDA due to the deficiency of 2DLDA in handling nonlinear manifolds.

Fourth, the top recognition of 2DSLSDP outperforms all other methods when the number of projected vectors is larger than 5 on the AR database; 2DSLSDP is always better than unsupervised methods 2DPCA and 2DLPP and supervised method 2DLDP under the same number of the projected vectors on the UMIST database.

6. Conclusion

This paper presents a novel feature extraction method, called two-dimensional supervised local similarity and diversity projection (2DSLSDP). Different from 2DLPP and other supervised two-dimensional manifold learning methods, 2DSLSDP defines two weighted adjacency graphs, namely similarity graph and diversity graph, over the training data points. Similarity graph captures the adjacency intrinsic structure of data, while diversity graph embodies the diversity information of the data. Two weighted scatters are then calculated by using these two graphs, respectively. Experiments on the AR and UMIST face databases show the efficiency of the proposed method 2DSLSDP. What's more, the efficiency of 2DSLSDP indicates the importance of diversity information of data.

Note that most existing manifold learning algorithms, not just 2DLPP, have the potential of impairing the diversity of data, in our future work we will extend the idea presented in this paper to a more general graph-embedding framework.

Acknowledgement

We would like to thank all the anonymous reviewers for their constructive advices. We also thank Dr. Jian Yang for providing the cropped images of AR database. This project is supported by National Natural Science Foundation of China under Grants no. 60802075, the Key Laboratory of Integrated Services Networks of China under Grants no. ISN090403, and 111 Project of China (B08038).

References

[1] J.B Tenenbaum, V. de Silva, J.C. Langford., A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2320–2323.
[2] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[3] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373–1396.
[4] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacian faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.
[5] X. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding, Proceeding of the 10th IEEE International Conference on Computer Vision (2005).
[6] D. Cai, X. He, J. Han, Isometric projection, Association for the Advancement of Artificial Intelligence (2007).
[7] D. Hu, G. Feng, Z. Zhou, Two-dimensional locality preserving projections with its application to palmprint recognition, Pattern Recognition 40 (2007) 339–342.
[8] H. Zhao, S. Sun, Z. Jing, J. Yang, Local structure based on supervised feature extraction, Pattern Recognition 39 (2006) 1546–1550.
[9] R. Zhi, Q. Ruan, Facial expression recognition based on two-dimensional discriminant locality preserving projections, Neurocomputing 71 (2008) 1730–1734.
[10] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin., Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2007) 40–51.
[11] Z. Hu, Comment on two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognition 41 (4) (2008) 1426.
[12] D. Hu, G. Feng, Z. Zhou, Reply on comment on two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognition 41 (4) (2008) 1427.
[13] J. Yang, D. Zhang, A.F. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (1) (2004) 131–137.
[14] H. Xiong, M.N.S. Swamy, M.O. Ahmad, Two-dimensional FLD for face recognition, Pattern Recognition 38 (7) (2005) 1121–1124.

Quan-xue Gao received the BE degree in 1998 from the University of Xi'an Highway, Xi'an China, the Master of engineering degree in 2001 from the Gan Su University of Science and Technology, Lan Zhou, China, and the PhD degree from the Northwestern Polytechnical University, Xi' an, China, in 2005. In 2006–2207, he was an associate research assistant in the Hong Kong Polytechnic University.