Neurocomputing 74 (2011) 629–637

Contents lists available at ScienceDirect

Neurocomputing

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

Sparse two-dimensional local discriminant projections for feature extraction

Zhihui Lai *, Minghua Wan, Zhong Jin, Jian Yang

School of Computer Science, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, PR China

ARTICLE INFO

Article history:
Received 3 February 2010
Received in revised form
7 July 2010
Accepted 22 September 2010
Communicated by: G.-B. Huang
Available online 20 October 2010

Keywords:
Feature extraction
Sparse subspace
Elastic Net
Spectral analysis

ABSTRACT

Two-dimensional local graph embedding discriminant analysis (2DLGEDA) and two-dimensional discriminant locality preserving projections (2DDLPP) were recently proposed to directly extract features form 2D face matrices to improve the performance of two-dimensional locality preserving projections (2DLPP). But all of them require a high computational cost and the learned transform matrices lack intuitive and semantic interpretations. In this paper, we propose a novel method called sparse two-dimensional locality discriminant projections (S2DLDP), which is a sparse extension of graph-based image feature extraction method. S2DLDP combines the spectral analysis and L_1-norm regression using the Elastic Net to learn the sparse projections. Differing from the existing 2D methods such as 2DLPP, 2DDLP and 2DLGEDA, S2DLDP can learn the sparse 2D face profile subspaces (also called sparsefaces), which give an intuitive, semantic and interpretable feature subspace for face representation. We point out that using S2DLDP for face feature extraction is, in essence, to project the 2D face images on the semantic face profile subspaces, on which face recognition is also performed. Experiments on Yale, ORL and AR face databases show the efficiency and effectiveness of S2DLDP.

1. Introduction

Techniques for feature extraction in supervised or unsupervised learning tasks have attracted much attention in computer vision and pattern recognition [1]. Due to the wide applications of appearance-based image recognition, many methods have been developed over the past few decades. In some applications such as face recognition and palmprint recognition, 2D images are usually transformed into 1D vectors through column by column or row by row concatenation. The resulting 1D image vectors of faces or palmprints usually lead to a high dimensional image vector space. Therefore, it is necessary for us to perform dimensionality reduction on the high dimensional data and obtain a compact representation of the data. Two of the most fundamental linear dimensionality reduction methods are principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [3]. PCA is an unsupervised learning method which aims to preserve total variances by maximizing the trace of feature covariance matrix. As a supervised learning algorithm, LDA aims to preserve the discriminant information maximizing between-class scatter and minimizing the within-class scatter simultaneously. Since the dimensions of 1-D vectors of the patterns are usually very high, a large computational cost involved in a big dense matrix eigen-decomposition is unavoidable when PCA and LDA are used for dimensionality reduction. Moreover, an intrinsic limitation of traditional LDA is that it fails to work when the within-class scatter matrix becomes singular, which is known as the small sample size (SSS) problem. In order to enhance the performance for classification, a number of variant PCA-based and LDA-based algorithms, such as probabilistic-based PCA [4,5], probabilistic linear discriminant analysis [6], geometric mean and harmonic mean based subspsace learning methods [7–9], discriminative common vectors [10], etc., were proposed in recent years.

Comparing with traditional PCA and LDA, two dimensional principal component analysis (2DPCA) [11] extracts image features directly from 2D image matrices rather than 1D vectors so that the image matrices do not need to be transformed into vectors. An image covariance matrix is directly constructed from the original image matrices for feature extraction. The optimal projection axes are its orthogonal eigenvectors corresponding to its larger eigenvalues. Due to the smaller size of the image variance matrix than the classical covariance matrix, 2DPCA requires less time to extract image features and achieves a better recognition rate. However, 2DPCA is suitable for data representation but not for classification. Therefore, Li and Yuan [12] extended the idea of using image matrices for LDA and presented 2DLDA for 2D face recognition. As a result, the SSS problem in LDA is naturally solved since the 2D within-class scatter matrix is nonsingular. There are some other image-based methods such as [13,14] designed for avoiding singularity of the within-class scatter matrix in LDA. In addition, a representatively extension of image-based method is the bilinear 2DLDA proposed in [15] using iteration algorithm to obtain the

* Corresponding author.
E-mail address: lai_zhi_hui@163.com (Z. Lai).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.09.010

optimal projections. Motivated by the successes of the bilinear 2DLDA for face recognition, Tao et al. [16] developed a general tensor discriminant analysis for gait recognition.

However, PCA, LDA and their 2D versions fail to discover and preserve the local information. A number of linear dimensionality reduction techniques have been developed to address this problem. Recently, He et al. [17,18] proposed a linear method named locality preserving projections (LPP) to approximate the eigenfunctions of the Laplace Beltrami operator on the manifold. LPP preserves the local manifold structure modeled by a nearest neighbor graph of the patterns so that the learned subspace preserves local information. This idea was further extended and a general framework of graph embedding was proposed in [19] for dimensionality reduction. Based on the idea of local patch alignment, a systematic framework was developed in [20] for understanding the common properties and intrinsic difference in different manifold learning algorithms. However, those vertor-based methods usually encounter SSS problems and cannot be implemented directly because of the singularity of matrix.

Motivated by the 2DPCA and 2DLDA, which operate directly on 2D image matrices, two-dimensional locality preserving projection (2DLPP) [21–23] was proposed for face feature extraction. More recently, some variant versions of 2DLPP such as graph-based 2DLGEDA [25], two dimensional discriminant locality preserving projection (2DDLPP) [24] and its variation [26] were also proposed to improve the performance of the 2DLPP. However, there still exist some problems in 2DLPP-based methods, including 2DLGEDA and 2DDLPP. Firstly, the computational cost of 2DLPP-based methods is high because it involves dense matrix eigen-decomposition and operations on Kronecker products of matrices. Secondly, although 2D-based methods learn from face images, they can not learn the face-like semantic transformation matrices similar to Eigenfaces (PCA), Fisherfaces (LDA) and Laplacianfaces [18].

From the aspect of understanding the extracted features in a low-dimensional subspace, one of the common major disadvantages of all the linear methods mentioned above is that the learned projection axes are linear combinations of ***all*** the original features, hence the projection axes are often hard to give physical interpretations to the extracted features. In order to interpret the physical meaning hiding in the learned projection axes, there has been great interest in sparse subspace learning in recent years. Zou et al. [27] proposed sparse PCA (SPCA) using the Elastic Net [28] for L_1-penalized regression on regular principle components and least angle regression [29]. d'Aspremont et al. [30] relaxed the hard cardinality constraint and obtained a convex approximation using semi-definite programming. Moghaddam et al. [31,32] proposed a spectral bounds framework for sparse subspace learning. Particularly, they proposed both exact and greedy algorithms for sparse PCA and sparse LDA. There are also related works on SPCA such as [33,34] via regularized low rank matrix approximation. Recently, Cai et al [35,36] proposed a unified sparse subspace learning (USSL) with spectral regression for special weighted/affined matrices. In order to give a reasonable interpretation to the extracted features, the existing sparse feature extraction methods such as SPCA, SLDA and USSL have to directly operate on the high dimensional vectors. As a result, two disadvantages are encountered. First, the processes of learning sparse projections are rather time-consuming when the dimensions of the features are very high. Second, when the sparse subspace learning algorithms are used in some special applications such as face recognition and palmprint recognition, 2D images are usually reshaped into high dimensional vectors, and thus some useful structural information embedded in the original 2D images may be lost.

Motivated by the image-based 2D feature extraction methods and sparse subspace learning algorithms, in this paper, based on 2DLGEDA, we propose a novel sparse subspace learning framework called sparse two-dimensional locality discriminant projections (S2DLDP) for two-dimensional image feature extraction. S2DLDP can be viewed as a sparse extension of graph-based 2D image feature extraction methods. S2DLDP not only further improves the performance of 2DLGEDA, but also learns an intuitive and interpretable subspace. Unlike the sparse feature extraction methods that operate on the very high dimensional image vectors, S2DLDP directly works on 2D image matrices.

Our contributions to the feature extraction problems are as follows. Firstly, we propose a novel framework, which imposes the sparseness in the graph-based feature extraction methods. Secondly, we give the spectral analysis of Kronecher product of the local neighborhood graph matrices. Thirdly, the 3 vector-based (1D-based) regression methods (i.e. ridge regression, lasso regression and Elastic Net) are extended into the image-based (2D-based) regression methods. At last, we obtain a novel framework for sparse two-dimensional feature extraction based on L_1L_1 -norm regression using the spectral analysis and the extended Elastic Net. Thus, the proposed framework generalizes the manifold-learning-based non-sparse two-dimensional linear feature extraction methods to sparse cases. Moreover, we show the semantic 2D sparsefaces learned by S2DLDP. The sparsefaces conquer the drawback that the transform matrices learned by the 2D methods have not semantic face properties. Therefore, we walk a step forward in the direction on learning the semantic face basis from vector-based pattern to image-base pattern. Furthermore, we give the intuitive explanations to the learned subspaces on which the 2D raw face images are projected. It is pointed out that the essence of using S2DLDP for face feature extraction is to project the 2D face images on the face profile subspaces for representations and thus face recognition is also performed on the profile subspaces.

The rest of the paper is organized as follows. In Section 2, we briefly review the 2DLDEGA and 2DDLPP. S2DLDP algorithm is described in Section 3. In Section 4, experiments are carried out to evaluate our S2DLDP algorithm. Finally, the conclusions and future work are given in Section 5.

2. Review of 2DLGEDA and 2DDLPP

Two-dimensional local discriminant methods have been proposed as a supervised extension of 2DLPP that works directly on 2D images. Assume that $X_i \in R^{n_1 \times n_2} (i=1,2,\ldots,m)$ are the 2D image matrices of the training images, where $n_1 > n_2$ and m is the number of training samples. Suppose A is an $n_2 \times d$-dimensional matrix, where each column of A is a unitary column vector and $d \le n_2$. The purpose of 2D feature extraction techniques is to seek an optimal projective matrix $A=(\varphi_1,\varphi_2,\ldots,\varphi_d)$ and then map a 2D image from $n_1 \times n_2$-dimensional image space into an $n_1 \times d$-dimensional space by the following linear projection:

$$Y_i = X_i A (i=1,2,\ldots,m) \tag{1}$$

2.1. Two-dimensional local graph embedding dscriminant analysis (2DLGEDA)

The goal of 2DLGEDA [25] is to preserve the 2D image within-class compactness and maximize the between-class separability. 2D image within-class compactness is characterized from the intrinsic graph by the term

$$\begin{aligned} S_w &= \sum_{i=1}^{m}\sum_{j=1}^{m} \|Y_i - Y_j\|_F^2 W_{ij}^w = \sum_{i=1}^{m}\sum_{j=1}^{m} \|X_i\varphi - X_j\varphi\|_F^2 W_{ij}^w \\ &= 2\varphi^T X[(D^w - W^w)\otimes I_{n_1}]X^T\varphi = 2\varphi^T X(L^w \otimes I_{n_1})X^T\varphi \end{aligned} \tag{2}$$

$$W_{ij}^{w} = \begin{cases} 1, & X_i \in N_{k_w}^{+}(X_j) \text{ or } X_j \in N_{k_w}^{+}(X_i) \\ 0, & \text{otherwise} \end{cases}$$

where $X=[X_1^T, X_2^T, \ldots, X_m^T]^T$ is the 2D image training sample matrix of size $mn_1 \times n_2$, and $D$is a diagonal matrix whose entries are column or row sums of W, I_{n_1} is an identity matrix of order n_1, operator $\otimes$ is the Kronecher product of the matrices, $N_k^{+}(X_i)$ indicates the samples in the k_w nearest neighbors of X_i in the same class and φ denotes the projection vector, $L^w = D^w - W^w$.

Similarly, 2D image between-class separability is characterized from the between-class graph by the term

$$S_b = \sum_{i=1}^{m}\sum_{j=1}^{m} \|Y_i - Y_j\|_F^2 W_{ij}^b = \sum_{i=1}^{m}\sum_{j=1}^{m} \|X_i\varphi - X_j\varphi\|_F^2 W_{ij}^b$$
$$= 2\varphi^T X[(D^b - W^b)\otimes I_{n_1}]X^T\varphi = 2\varphi^T X(L^b \otimes I_{n_1})X^T\varphi \tag{3}$$

$$W_{ij}^{b} = \begin{cases} 1, & \text{if } (i,j) \in P_{k_b}(c_i) \text{ or } (i,j) \in P_{k_b}(c_j) \\ 0, & \text{otherwise} \end{cases}$$

where $X=[X_1^T, X_2^T, \ldots, X_m^T]^T$ is the 2D image training sample matrix of size $mn_1 \times n_2$, and $P_{k_p}(c_i)$ is a set of data pairs that are in the k_p nearest pairs among the set $\{(i,j) | i \in \pi_c, j \notin \pi_c\}$, where $\pi_c$denotes the index set of cth class and c varies from 1 to the number of classes. D^b is a diagonal matrix whose entries are column or row sums of W^b, and $L^b = D^b - W^b$.

Finally, the criterion of 2DLGEDA is formally similar to the Fisher criterion since they are both Reyleigh quotients and the optimal projections can be obtained from solving the generalized eigen-equation:

$$X^T(L^b \otimes I_{n_1})X\varphi = \lambda X^T(L^w \otimes I_{n_1})X\varphi \tag{4}$$

where λ is generalized eigenvalue corresponding to the eigenvector φ. Then, the optimal transformation matrix of 2DLGEDA is composed of the eigenvectors associated with the d largest eigenvalues.

2.2. Two-dimensional discriminant locality preserving projections

Two-dimensional discriminant locality preserving projections (2DDLPP) were proposed in [24] and applied to facial expression recognition. The objective function of the 2DDLPP is to maximize $\overline{S}_b$, the weighted sum of the distances between means of the samples of different classes, and at the same time, to minimize $\overline{S}_w$, the weighted sum of the distances within the same class. The discriminator of the 2DDLPP measuring the between-class separability can be represented as

$$\overline{S}_b = \sum_{i=1}^{m}\sum_{j=1}^{m} \|Y_i - Y_j\|_F^2 \overline{W}_{ij}^b = \sum_{i=1}^{m}\sum_{j=1}^{m} \|X_i\varphi - X_j\varphi\|_F^2 \overline{W}_{ij}^b$$
$$= 2\varphi^T X[(\overline{D}^b - \overline{W}^b)\otimes I_{n_1}]X^T\varphi = 2\varphi^T X(\overline{L}^b \otimes I_{n_1})X^T\varphi \tag{5}$$

$$\overline{W}_{ij}^{b} = \begin{cases} \left(\frac{1}{m_c} - \frac{1}{m}\right) \times \exp\left(-\|M_i - M_j\|^2/t\right), & \text{if } X_i \text{ and } X_j \text{ are in the different class} \\ \frac{1}{m}, & \text{otherwise} \end{cases}$$

where $\overline{D}^b$is a diagonal matrix whose entries are column or row sums of $\overline{W}^b$, and $\overline{L}^b = \overline{D}^b - \overline{W}^b$; m_c is the number of cth class, M_i is the sample mean value of the ith class and t is the Gaussian kernel parameter.

The discriminator of the 2DDLPP measuring the within-class compactness can be represented as

$$\overline{S}_w = \sum_{i=1}^{m}\sum_{j=1}^{m} \|Y_i - Y_j\|_F^2 \overline{W}_{ij}^w = \sum_{i=1}^{m}\sum_{j=1}^{m} \|X_i\varphi - X_j\varphi\|_F^2 \overline{W}_{ij}^w$$
$$= 2\varphi^T X[(\overline{D}^w - \overline{W}^w)\otimes I_{n_1}]X^T\varphi = 2\varphi^T X(\overline{L}^w \otimes I_{n_1})X^T\varphi \tag{6}$$

$$\overline{W}_{ij}^{w} = \begin{cases} \exp\left(-\|X_i - X_j\|^2/t\right), & \text{if } X_i \text{ and } X_j \text{ are in the same class} \\ 0, & \text{otherwise} \end{cases}$$

Similar to 2DLGEDA, the optimal projections of 2DDLPP can be obtained solving the generalized eigen-equation:

$$X(\overline{L}^b \otimes I_{n_1})X^T\varphi = \lambda X^T(\overline{L}^w \otimes I_{n_1})X\varphi \tag{7}$$

The optimal transformation matrix of 2DDLPP is composed of the eigenvectors associated with the first d largest eigenvalues of Eq. (7).

2.3. Discussions about the 2D discriminant feature extraction framework

It can be found that both 2DLGEDA and 2DDLPP are one of the graph-based representative methods of the general 2D dimensionality reduction framework: $X^T(G_a \otimes I)X\varphi = \lambda X^T(G_b \otimes I)X\varphi$, where G_a and G_b represent the local neighborhood graphs (or their graph Laplacian) defined in different ways. Many existing 2D-based methods such as those in [21–26] can also be described in this framework.

There are two disadvantages of these 2D-based supervised feature extraction frameworks. One is its high computational complexity. The main computational cost is due to the calculations of $X^T(G_a \otimes I_{n_1})X$ and $X^T(G_b \otimes I_{n_1})X$. The other is that the learned projection axes are linear combination of all the original features, hence it is often hard to give them a reasonable interpretation. In the face recognition case, although 2D-based methods learn from face images, they all cannot learn the face-like semantic transformation matrices (subspaces) similar to the vector-based Eigenfaces, Fisherfaces and Laplacianfaces. In the following section, we focus on the two disadvantages and develop the proposed S2DLDP framework from 2DLGEDA.

3. Sparse two-dimensional local discriminant projections (S2DLDP)

3.1. The model of S2DLDP

Our purpose is to develop an algorithm to extract the sparse features which can be interpreted intuitively or semantically from 2D face images. Our idea is to impose a sparseness constrained condition in Eq. (4). The model of S2DLDP is given as follows:

$$\begin{cases} X^T(L^b \otimes I_{n_1})X\varphi = \lambda X^T(L^w \otimes I_{n_1})X\varphi \\ \text{subject to} \quad \text{Card}(\varphi) \le K \end{cases} \tag{8}$$

where φ is the column vector corresponding to eigenvalue λ and *Card*(φ) denotes the number of non-zero elements of φ. The only difference between Eqs. (4) and (8) is that a sparse constraint is imposed in Eq. (8). On the one hand, directly solving the generalized eigen-function of Eq. (4) cannot obtain the sparse projections. On the other hand, the computations of $X^T(L^b \otimes I_{n_1})X$ and $X^T(L^w \otimes I_{n_1})X$ are expensive. Therefore, in order to avoid expensive computations and obtain the sparse solutions using L_1-norm regression, we develop two theorems in the following section.

3.2. The theorems

Theorem 1. *Assume that λ is an eigenvalue and y is the corresponding eigenvector of the generalized eigenvalue problem*

$$(L^b \otimes I_{n_1})y = \lambda(L^w \otimes I_{n_1})y \tag{9}$$

If there exist φ satisfying $y = X\varphi$, then λ and φ will be the eigenvalue and the corresponding eigenvector of Eq. (4).

Proof. We have

$$X^T(L^b \otimes I_{n_1})X\varphi = X^T(L^b \otimes I_{n_1})y = \lambda X^T(L^w \otimes I_{n_1})y = \lambda X^T(L^w \otimes I_{n_1})X\varphi.$$

Thus, λ and φ are the eigenvalue and the corresponding eigenvector of Eq. (4).■

From Theorem 1, instead of directly solving the generalized eigen-problem in Eq. (4), we can solve the eigen-problem in Eq. (9), and then find φ such that $X\varphi = y$. However, directly solving the eigen-problem in Eq. (9) is still computationally expensive due to the eigen-decomposition of a large size matrix. The following theorem can further help us reduce the cost of computing the eigenvectors of Eq. (9).

Theorem 2. *Assume that λ is the eigenvalue and z is the corresponding eigenvector of the generalized eigen-problem*

$$L^b z = \lambda L^w z \tag{10}$$

Let v be any non-zero unit vector in R^{n_1} and $y = z \otimes v$, then λ and y are the eigenvalue and the corresponding eigenvector of the generalized eigen-problem in Eq. (9).

Proof. We have

$$(L^b \otimes I_{n_1})y = (L^b \otimes I_{n_1})(z \otimes v) = (L^b z) \otimes (I_{n_1} v) = \lambda(L^w z) \otimes (I_{n_1} v)$$
$$= \lambda(L^w \otimes I_{n_1})(z \otimes v) = \lambda(L^w \otimes I_{n_1})y$$

Therefore, λ and y are the eigenvalue and corresponding eigenvector of the generalized eigen-problem in Eq. (9). ■

From Theorem 2, instead of solving the generalized eigen-problem in Eq. (9), we can solve the generalized eigen-problem in Eq. (10) to obtain the generalized eigenvectors of Eq. (9) and save the computational cost. As a result, we can obtain λ and z from Eq. (10) with lower computational cost because the sizes of L^b and L^w are smaller than $L^b \otimes I_{n_1}$ and $L^w \otimes I_{n_1}$ in Eqs. (2)–(4).

3.3. The sparse solutions of S2DLDP

For $y = X\varphi$ in Theorem 1, such φ might not exist. There exist three methods to obtain the approximating solutions: (i) ridge regression [37], (ii) lasso regression [37] and (iii) Elastic Net [28]. Since we only focus on the 2D image based feature extraction methods, we directly extend these three vector-based (1D-based) regression methods into image-based (2D-based) cases. For more details, please see the related references. The 2D extension of the 3 regression methods are stated as follows:

(i) ***2D extension of the ridge regression:*** We can use 2D extension of the ridge regression [37] to solve this problem

$$\varphi = \arg\min_{\varphi} \left(\sum_{i=1}^{m} \sum_{h=1}^{n_1} (X_i(h,:) \times \varphi - y_i)^2 + \alpha \sum_{j=1}^{n_2} \overline{\varphi}_j^2 \right) \tag{11}$$

where $\alpha \geq 0$ and $X_i(h,:)$ is the hth row of image matrix X_i, y_i is the ith element of y and $\overline{\varphi}_j$ is the jth element of φ. However, the ridge penalty does not provide a sparse solution.

(ii) ***2D extension of the lasso regression:*** With a L_1-norm penalty on φ, we have the 2D extension of the lasso regression [37]:

$$\varphi = \arg\min_{\varphi} \left(\sum_{i=1}^{m} \sum_{h=1}^{n_1} (X_i(h,:) \times \varphi - y_i)^2 + \beta \sum_{j=1}^{n_2} \left| \overline{\varphi}_j \right| \right), \tag{12}$$

where $\left|\overline{\varphi}_j\right|$ denotes the absolute value of $\overline{\varphi}_j$. Due to the nature of the L_1 penalty, some coefficients will be shrunk to zero if β is large enough. However, the lasso has several limitations as pointed out in [28], i.e. the number of selected features by the lasso is limited by the number of samples. Therefore, lasso regression is not the optimal method to obtain the sparse solutions.

(iii) ***2D extension of the Elastic Net:*** Recently, the Elastic Net was widely used in feature selection. The Elastic Net generalizes the lasso to overcome its limitations combining both the ridge and lasso penalty. In this paper, the Elastic Net is extended to 2D case and used to obtain the sparse solutions:

$$\varphi = \arg\min_{\varphi} \left(\sum_{i=1}^{m} \sum_{h=1}^{n_1} (X_i(h,:) \times \varphi - y_i)^2 + \alpha \sum_{j=1}^{n_2} \overline{\varphi}_j^2 + \beta \sum_{j=1}^{n_2} \left| \overline{\varphi}_j \right| \right). \tag{13}$$

According to the above two theorems and Eq. (13), spectral regression with L_1 penalty combining the spectral graph analysis provides an effective approach for sparse subspace learning problem. Thus, the optimal sparse projections of the proposed method can be given by regression with L_1 norm. The optimal sparse approximate solutions of Eq. (8) are called sparse two-dimensional local discriminant projections (S2DLDP).

Denote $y^l = z \otimes v^l$ $(l = 1,\ldots,d)$, where $l \leq d \leq n_2$ and z is the eigenvector associated with the first largest eigenvalue by solving the eigen-equation Eq. (10) and v^l's are mutually orthogonal unit vectors. Then, the optimal projections can be obtained from the following optimal problem:

$$\varphi_l = \arg\min_{\varphi} \left(\sum_{i=1}^{m} \sum_{h=1}^{n_1} (X_i(h,:) \times \varphi - y_i^l)^2 + \alpha \sum_{j=1}^{n_2} \overline{\varphi}_j^2 + \beta \sum_{j=1}^{n_2} \left| \overline{\varphi}_j \right| \right) \tag{14}$$

where y_i^l $(l = 1,\ldots,d)$ denotes the ith element of y^l. Thus, the optimal sparse projection matrix of S2DLDP is $A = (\varphi_1, \varphi_2, \ldots, \varphi_d)$.

Remark: Regression methods, such as ridge regression, lasso regression and Elastic Net, are usually performed on the vector-based (1D based) optimization problems. Eqs. (11–14) extend the vector-based regression problems to the image-based (2D based) regression problems. The essence of image-based regression is that each row (or column) of the images is viewed as a vector and then perform regression on these vectors as vector-based regression.

Note that there are at most n_1 orthogonal n_1-dimensional vectors v^l $(l = 1,\ldots,n_1)$ and we thus can obtain $n_1$eigenvectors $y^l = z \otimes v^l$ $(l = 1,\ldots,n_1)$ for each eigenvector of Eq. (10) for regression using Eq. (14). Therefore, for each eigenvector of Eq. (10) we obtain n_1 sparse projections, which construct a sparseface defined in Section 4.1.1 in this paper. As a result, S2DLDP can obtain at most $m \times n_1$ n_2-dimensional projection axes, which are far more than the ones learned by the other 2D methods (the 2DLPP-based methods can only learn n_2 $(n_2 < n_1)$ projection axes since the projection axes are obtained from generalized eigenfunctions of $n_2 \times n_2$ matrices similar to Eq. (4)). Note that only the eigenvector z corresponding to the largest eigenvalue of Eq. (11) is used for constructing y^l for regression in the S2DLDP algorithm since using too much sparse projections cannot significantly improve the recognition rate but increase the computational burden. If necessary, one can use all the $m \times n_1$ projection axes.

We summarize S2DLDP algorithm in Table 1. Once the optimal sparse projections are obtained, the samples can be projected to the low-dimensional sparse subspace for classification.

Table 1
S2DLDP algorithm.

Step 1. Construct the similarity matrices W^b, $W^w$and the corresponding Laplace matricesL^b, L^w.
Step 2. Compute the first eigenvector associated with the largest eigenvalue of the eigen-equation (10).
Step 3. Select d mutually orthogonal unit vectors $v_1, v_2, \ldots, v_d$.
Step 4. Compute $y^1, y^2, \ldots y^d$as $y^i = z \otimes v_i (i = 1,2,\ldots,d)$.
Step 5. Compute the d sparse solutions according to Eq. (14) to construct transformation matrix $A = (\varphi_1, \varphi_2, \ldots, \varphi_d)$.

4. Experiment and analysis

In order to evaluate the proposed S2DLDP algorithm, we compare it with 2DPCA, 2DLDA, 2DLPP and 2DLGEDA on Yale, ORL and AR face databases. The Yale face database was used to examine the performance of S2DLDP when both facial expressions and illumination and sample size are varied. The ORL face database was used to evaluate the performance of S2DLDP under conditions where the pose, face expression vary. The AR face database was employed to test the performance of S2DLDP under conditions where there are variations in time, facial expressions and large lighting conditions. The nearest neighborhood classifier with Euclidean distance is used in all the experiments.

4.1. Experiments on the Yale face database

The Yale face database (http://www.cvc.yale.edu/projects/yalefaces/yalefaces.html) contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 50×40 pixels. Fig. 1 shows sample images of one person in the Yale database.

4.1.1. Sparsefaces: an intuitive explanation

At first, we explore the learned projection matrix, i.e. the "faces" learned by 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed S2DLDP. The first six images per class were used to learn the projection matrix. The 2D "faces" learned by 2DPCA, 2DLDA, 2DLPP and 2DLGEDA are potted in Fig. 2(a). The faces learned by S2DLDP with $Card(\varphi) = 4:4:20$ are shown in Fig. 2(b). In order to clearly show the properties of the faces learned by S2DLDP, binary images (non-zero values are set to be 0 and the zero values are set to be 1) of Fig. 2(b) are shown in Fig. 2(c). As we can see from Fig. 2, none of the "faces" learned by 2DPCA, 2DLDA, 2DLPP and 2DLGEDA can show us semantic face-like images. The faces learned by the proposed S2DLDP are significantly different form the other methods. The non-zero elements in the matrices are automatically formed face profiles. Thus, S2DLDP learns a set of semantic face profiles (or contours). We call the faces learned by S2DLDP as sparsefaces or profile subspaces since each projection matrix is a sparse matrix and show us a face-profile-like image.

Fig. 1. Sample images of one person in the Yale database.

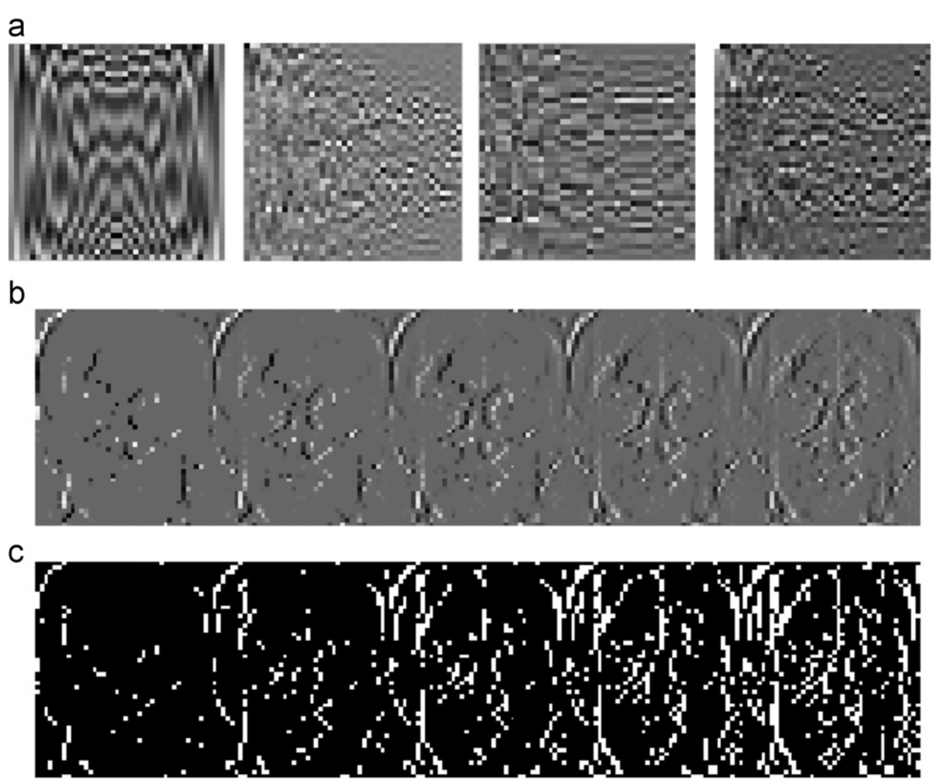


Fig. 2. The "faces" learned by 2D methods. (a) From left to right: 2DPCA face, 2DLDA face, 2DLPP face and 2DLGEDA face. (b) The sparsefaces learned by S2DLDP. (c) Black and white showing the sparsefaces corresponding to (b), the white points are non-zeros loadings and the black areas are zero elements.

Now, let us consider the feature extraction step. A 2D face image X_i is projected on the sparsefaces subspaces, i.e. $Y_i = X_i A = X[\varphi_1, \varphi_2, \ldots, \varphi_d]$. Form this equation we can easily find that only the non-zero elements in φ_i $(i=1,\ldots,d)$ are contributive to Y_i in low-dimensional subspace. Thus, only the non-zero elements in the projective matrix A are contributive for low-dimensional 2D face image feature Y_i. This clearly shows us that the raw 2D images are essentially projected on the profile subspace when using S2DLDP for feature extraction. As a result, classification is also performed on the profile subspaces. Assume that the 2D face images of different individuals lie on different submanifolds, then the low-dimensional feature Y_i's represent/reflect each submanifold's distributions on the profile subspaces. Using these face features for classification is exactly performed for face recognition on the profile subspaces. This gives us an insightful understanding for 2D face feature extraction and 2D face recognition. Thus, the sparse 2D profile subspaces, i.e. sparsefaces, explicitly show us what the learned subspace looks like and which subspace the 2D images are projected on. As a result, the sparsefaces overcome the drawback of lacking semantic properties in the projection matrices learned by the other 2D methods.

The following experiments will prove that the profile subspaces are more discriminative than the ones learned by the other 2D methods. The efficiency of S2DLDP will be shown in Section 4.3 on a larger database.

4.1.2. Robustness of the sparsefaces

In this experiment, we test the robustness of the sparsefaces. We focus on the case that there are outliers (left-light, right-light and surprised images can be viewed as outliers) in training set and text set. The experiment was performed using the first six images (i.e. center-light, with glasses, happy, left-light, without glasses, and normal) per class for training, and the remaining five images (i.e. right-light, sad, sleepy, surprised and winking) for test. Thus, there are outliers in both training set and test set. For feature extraction, we used 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed S2DLDP, respectively.

The maximal recognition rates of 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and S2DLDP with $Card(\varphi) = 4$ are 93.33%, 89.33%, 94.67%, 92.00% and 97.33%, respectively. The top recognition rate of S2DLDP with $Card(\varphi) = 4$ is significantly higher than the other methods. This indicates that S2DLDP is more robust than the other methods under the facial expressions and illumination variations. The recognition rates of S2DLDP versus the variation of $Card(\varphi)$ and dimension are shown in Fig. 3. From Fig. 3, we can also see that the recognition rates achieve its best result when $Card(\varphi) = 4$ and then decrease with the increase in the number of $Card(\varphi)$. The optimal sparseness is $Card(\varphi) = 4$with non-zero loadings occupying 10% of the elements in the projection axes. This experiment also shows that the sparseface when $Card(\varphi) = 4$, i.e. the first sparseface shown in Fig. 2 has more discriminative ability. With the further increase in the number of "*Card*", the discriminant ability of S2DLDP first decreases and then increases. This experiment indicates that the sparsefaces with a larger "*Card*" are not necessary to achieve higher recognition rates.

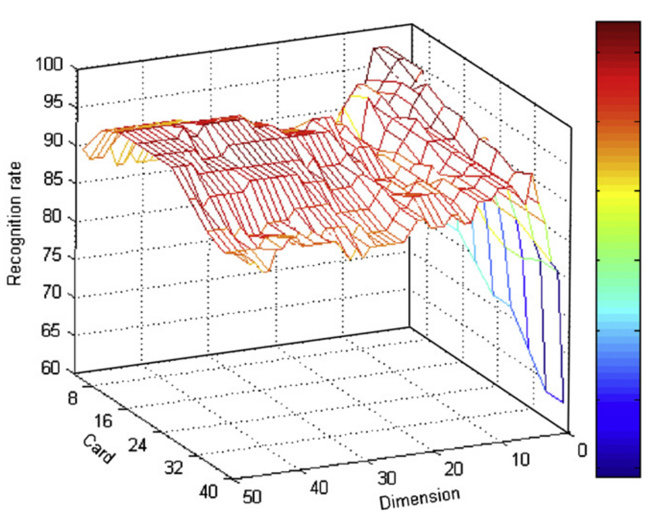


Fig. 3. The variation of *Card*, dimension and recognition rate of S2DLDP.

4.1.3. Performance of the sparsefaces

In this section, we evaluate the performance of the proposed method using the first 10 images of each person. For 2DPCA and 2DLDA, the only model parameter is the subspace dimension. For 2DLPP, 2DLGEDA and S2DLDP, the model parameters include neighborhood size, kernel width and cardinality. In our experiments, we simply set $k_w = l-1$, where l denotes the number of the training samples per class. The justification for this choice is that each sample should connect with the remaining $l-1$ samples of the same class such that within-class samples can be mapped together in the low-dimensional subspace. When using Gaussian kernel, we simply set kernel width $t = +\infty$(i.e. 0–1 pattern is used). The neighborhood size k in 2DLPP, k_p in 2DLGEDA and S2DLDP are selected from $\{1,2,4,\ldots,m-1\}$, and the cardinality K in S2DLDP is selected from $\{1,2,\ldots,n_1\}$ all by 5-fold cross-validation with one fold for training and the other 4 folds for validation. The average recognition rates of each method and the corresponding dimensions are given in Table 2. The average recognition rates (%) versus the dimensions are shown in Fig. 4. As it is shown in Table 2 and Fig. 4, the top recognition rates of S2DLDP are significantly higher than the other methods.

From the experimental results, we can draw a conclusion that S2DLDP effectively extract the distcriminant features of the face images and the face profile feature subspaces (i.e. sparsefaces) are the more discriminative feature subspaces for 2D face recognition. That is, S2DLDP not only can give an intuitive explanation that the

Table 2
The average recognition rates (percent) and the corresponding dimensions of five methods on the Yale face database.

Methods	2DPCA	2DLDA	2DLPP	2DLGEDA	S2DLDP
Recognition rate (%)	84.50	84.33	84.50	85.67	87.83
Dimension	40×34	40×14	40×3	40×26	40×44

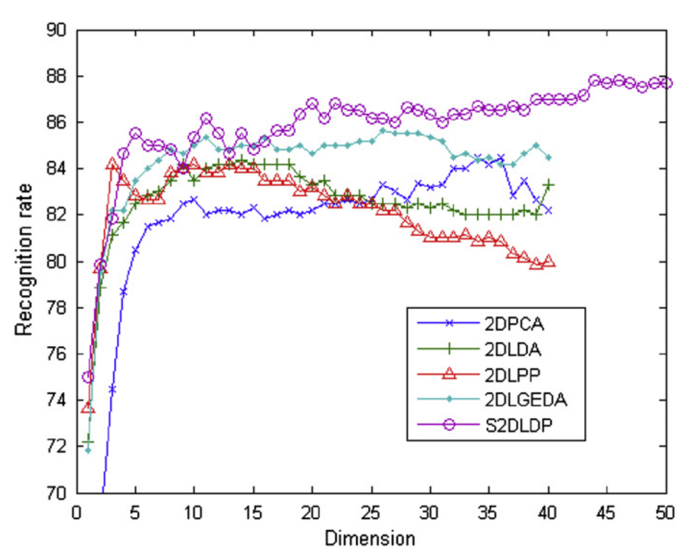


Fig. 4. The average recognition rates (%) versus the dimensions on the Yale face database.

low dimensional features are represented in profile subspaces, but also shows us that face profile subspaces are the more discriminative ones and of crucial importance for 2D face recognition.

4.2. Experiments on ORL face database

The ORL face database (http://www.uk.research.att.com/facedatabase.html) is used to evaluate the performance of S2DLDP under conditions where the pose and face expression vary. The ORL face database contains images from 40 individuals, each providing 10 different images. The facial expressions and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20^0. Moreover, there is also some variation in the scale of up to about 10%. All images normalized to a resolution of 56×46. Sample images of one person in the ORL face database are shown in Fig. 5.

In the experiments, the parameters are selected and set as in Section 4.1.3. We report the recognition rates of 5-fold cross-validation with 2 samples per class for training. The maximal average recognition rates of each method and the corresponding dimension are given in Table 3. The maximal average recognition rates (%) versus the dimensions are shown in Fig. 6. From Table 3 and Fig. 6, it can be found that S2DLDP obtains the highest average recognition rate. This indicates that sparsefaces, i.e. face profile subspaces, have more discriminant abilities than the image subspaces learned by other 2D methods.

4.3. Experiments on the AR face database

The AR face database [38] contains over 4000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of 120 individuals (65 men and 55 women) were taken in two sessions (separated by two weeks) and each section contains 13 color images. 20 images of 120 individuals are selected and used in our experiments. The face portion of each image is manually cropped and then normalized to 25×20 pixels for computational efficiency. The sample images of one person are shown in Fig. 7.

In the experiment, 10 images in the first section, i.e. the images on the first line in Fig. 7.) are selected from the image gallery of each individual to form the training sample set and the 10 images in the second section (the images on the second line in Fig. 7) are used for test. The parameters are selected and set as in Section 4.1.3. The neighborhood size k in 2DLPP, k_p in 2DLGEDA and S2DLDP are set as $1,2,4,\ldots,m-1$, and the cardinality K in S2DLDP is varied from 1 to n_1 to achieve the best results. The maximal recognition rate and corresponding dimensions are shown in Table 4. In addition, we also report the training time of 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed S2DLDP in Table 4 (CPU: Intel 3GHz, RAM: 2 GB). The recognition rates (%) versus the dimensions are shown in Fig. 8.

As it can be seen from Table 4 and Fig. 8, the top recognition rate of S2DLDP is the highest. The experimental results also support our analysis mentioned in Sections 4.1 and 4.2 and suggest that S2DLDP is more robust than 2DPCA, 2DLDA, 2DLPP and 2DLGEDA on facial expressions, lighting conditions and time variations. Moreover, compared with 2DLPP and 2DLGEDA, S2DLDP spends less time in learning the sparse projections. The Kronecker products of the large size matrices in 2DLPP and 2DLGEDA are rather time-consuming. Thus, spectral analysis combining with the Elastic Net greatly reduces the computation time since the computations of the Kronecker products of the large size matrices are avoided. This experiment indicates that S2DLDP is more effective and efficient than 2DLPP and 2DLGEDA.

4.4. Overall observations and discussions

According to the experiments performed on the three face databases, the following conclusions can also be drawn:

- Differing from 2DPCA and 2DLDA, which take the global Euclidean structure into account, 2DLPP, 2DLGEDA and S2DLDP focus on the local geometric structure. Local structure based manifold learning algorithms are superior to the methods based on global structure. Although 2DLPP, 2DLGEDA and S2DLDP all aim to discover the local geometric structure, S2DLDP consistently performs better than 2DLPP and 2DLGEDA, irrespective of the dimensional variation.
- Since the projection axes of S2DLDP are very sparse, the recognition rates of S2DLDP might not be higher than the other methods in the very low-dimensional subspace. But with a few more dimensions, the best recognition rates of S2DLDP are always significantly higher than the other methods'. This can be seen from Figs. 4, 6 and 8.

Table 3
The maximal average recognition rates (percent) and the corresponding dimensions of five methods on the ORL face database.

Methods	2DPCA	2DLDA	2DLPP	2DLGEDA	S2DLDP
Recognition rate (%)	83.94	84.19	84.25	85.06	86.75
Dimension	46×2	46×4	46×3	$46 \times \times 9$	46×41

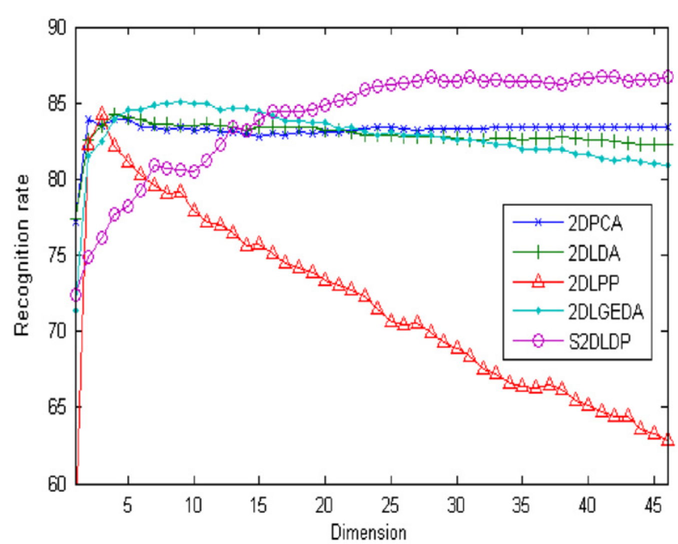


Fig. 6. The average recognition rates (%) versus the dimensions on the ORL face database.

Fig. 5. Sample images of one person in the ORL face database.

Fig. 7. Sample images of one person in the AR face database.

Table 4
The maximal recognition rates of five methods on the AR face database and the corresponding dimensions and training time (second).

Methods	2DPCA	2DLDA	2DLPP	2DLGEDA	S2DLDP
Recognition rate (%)	58.17	57.08	58.67	58.92	62.16
Dimension	20×18	20×20	20×15	20×20	20×24
Training time (s)	0.046	0.141	43.359	48.203	13.539

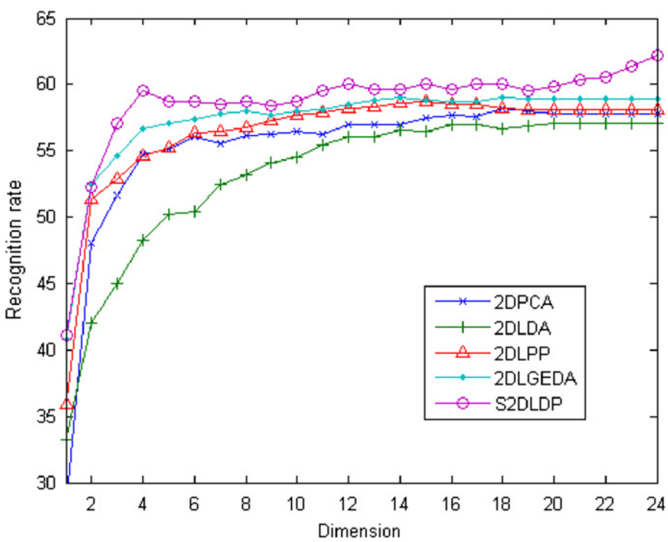


Fig. 8. The recognition rates (%) versus the dimensions on the AR face database.

- As it is shown in Fig. 2, the face subspaces (i.e. the transform matrices) learned by the existing 2D methods cannot provide semantic face images. Only the sparsefaces learned by S2DLDP can give us the face-like images.
- Theoretically speaking, all the 2D methods aim to preserve some kinds of geometric properties. But none of the compared 2D methods can give a meaningful and intuitive explanation on the learned subspace. However, S2DLDP not only gives an intuitively semantic interpretation of the learned subspace, but also shows us that the face profile subspace is the more discriminative feature subspace for face recognition. The explicit meaning using sparsefaces for feature extraction is to project the 2D face images on the face profile subspaces, which show us an insightful understanding of the appearance-based 2D face image representation and recognition.

5. Conclusions

In this paper, we develop a sparse learning technique called sparse two-dimensional locality discriminant projections (S2DLDP) for 2D image feature extraction. Similar to 2DDLPP and 2DLGEDA, S2DLDP also aims to discover the local geometry of the 2D image manifold. But significantly differing from the 2DDLPP and 2DLGEDA, the sparseness is imposed into the objective function. We extend the vector-based regression problem to the image-based (2D based) regression problems. Based on the spectral analysis and the L_1-norm regression using the 2D extension of the Elastic Net, we obtain a novel framework for sparse two-dimensional feature extraction.

The most important and interesting thing is that the sparse transformation matrices, i.e. sparsefaces, learned by S2DLDP provide us an insightful understanding for the feature extraction. That is, assume that the 2D face images of different individuals lie on different submanifolds, and then the low-dimensional features obtained from the proposed method essentially represent/reflect the submanifolds' distributions on the *profile subspaces* with maximal between-class separability and minimal within-class compactness. Using this low-dimensional features obtained by S2DLDP for classification is essentially performed classification on the *profile subspaces*. Experimental results on the Yale, ORL and AR face image databases show that S2DLDP not only can learn intuitively semantic face subspaces, but also can show us that face profile subspaces are more discriminative than the subspaces learned by other 2D methods and of crucial importance for 2D face recognition. Experiments show that S2DLDP is more effective and efficient than 2DLPP, 2DDLPP and 2DLGEDA.

Acknowledgments

This work is partially supported by the Natural Science Foundation of China under grant Nos. 60973098, 60632050, 60705006, 60873151 and Hi-Tech Research and Development Program of China under grant No. 2006AA01Z119.

References

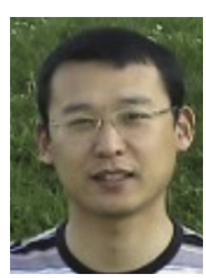
[1] A. Batur, M. Hayes, Linear subspaces for illumination robust face recognition, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 2001,pp. 296–301.
[2] M. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 586–591.
[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriengman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[4] M.E Tipping, C.M. Bishop, Probabilistic principal component analysis, Journal of the Royal Statistical Society, Series B 61 (3) (1999) 611–622.
[5] S. Yu, K. Yu, V. Tresp, H. Kriegel, M. Wu, Supervised probabilistic principal component analysis, in: Mark Lyle Ungar, Dimitrios Craven, Gunopulos, Tina Eliassi-Rad (Eds.), Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACMPress, New York, NY2006, pp. 464–473.
[6] S. Ioffe, Probabilistic linear discriminant analysis, in: Proceeings of the 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006, pp. 531–542.

[7] D. Tao, X. Li, X. Wu, S.J. Maybank, Geometric mean for subspace selection, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 260–274.
[8] D. Tao, X. Li, X. Wu, S.J. Maybank, General averaged divergences analysis, in: Proceedings of the IEEE International Conference on Data Mining, 2007.
[9] W. Bian, D. Tao, Harmonic mean for subspace selection, in: Proceedings of the 19th ICPR, 2008, pp. 1–4.
[10] M. Cevikalp, M. Neamtu, Wilkes, A. Barkana, Discriminative common vectors for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (1) (2005) 6–9.
[11] J. Yang, D. Zhang, A.F. Frangi, J.Y. 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.
[12] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recognition Letters 26 (5) (2005) 527–532.
[13] D.Q. Zhang, Z.H. Zhou, (2D)2PCA: two-directional two-dimensional PCA for efficient face representation and recognition, Neurocomputing 69 (2005) 224–231.
[14] P. Nagabhushan, D.S. Guru, B.H. Shekar, (2D)2FLD: an efficient approach for appearance based object recognition, Neurocomputing 69 (2006) 934–940.
[15] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, in: Proceedings of the Neural Information Processing Systems, 2005, pp. 1569–1576.
[16] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (10) (2007) 1700–1715.
[17] X. He, P. Niyogi, Locality Preserving Projections, Advances in Neural Information Processing System, vol. 16, MIT Press, 2004, pp. 153-160.
[18] X.F. He, S. Yan, Y. Hu, P. Niyogi, H.J. Zhang, Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.
[19] S. Yan, D. Xu, B. Zhang, H.-J. Zhang., Graph embedding: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence (Jan. 2007) 40–51.
[20] T. Zhang, D. Tao, et al., Patch alignment for dimension reduction, IEEE Transactions on Knowledge and Data Engineering 20 (2) (2009).
[21] Yang Ben, Pal Shiu, Two-dimensional Laplacianfaces method for face recognition, Pattern Recognition 41 (10) (2008) 3237–3243.
[22] D. Hu, G. Feng, Z. Zhou, Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognition 40 (1) (2007) 339–342.
[23] S. Chen, H. Zhao, M. Kong, B. Luo, 2DLPP: a two-dimensional extension of locality preserving projections, Neurocomputing 70 (4–6) (2007) 912–921.
[24] R. Zhi, Q. Ruan, Facial expression recognition base on two-dimensional discriminant locality preserving projections, Neurocomputing 70 (2007) 1543–1546.
[25] M. Wan, Z. Lai, J. Shao, Z. Jin, Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and Palm Biometrics, Neurocomputing 73 (2009) 193–203.
[26] Y. Xu, G. Feng, Y. Zhao, One improvement to two-dimensional locality preserving projection method for use with face recognition, Neurocomputing 73 (2009) 245–249.
[27] H. Zhou, T. Hastie, R. Tibshirani, Sparse Principle Component Analysis, Technical Report, Statistics Department, Stanford University, 2004.
[28] H. Zhou, T. Hastie. Regression Shrinkage and Selection via the Elastic Net, with Applications to Microarrays, Technical Report, Statistics Department, Stanford University, 2003.
[29] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Annals of Statistics 32 (2) (2004) 407–499.
[30] A. d'Aspremont, L.E. Chaoui, M.I. Jordan, G.R.G. Lanckriet, A direct formulation for sparse PCA using semidefinite programming, in: Advances in Neural Information Processing Systems, vol. 17, 2004.
[31] B. Moghaddam, Y. Weiss, S. Avidan, Spectral bounds for sparse PCA: exact and greedy algorithms, Advances in Neural Information Processing Systems 18 (2006) 915–922.
[32] B. Moghaddam, Y. Weiss, S. Avidan, Generalized spectral bounds for sparse LDA, in: Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 641–648.
[33] A. d'Aspremont, F. Bach, L.E. Chaoui, Optimal solutions for sparse principal component analysis, Journal of Machine Learning Research 9 (2008) 1269–1294.
[34] S. Haipeng, Z.H. Jianhua, Sparse principal component analysis via regularized low rank matrix approximation, Journal of Multivariate Analysis 99 (2008) 1015–1034.
[35] D. Cai, X. He, J.Han, Spectral regression: a unified approach for sparse subspace learning, in: Proceedings of the 2007 International Conference on Data Mining (ICDM 07), Omaha, NE, October 2007.
[36] D. Cai, X. He, J. Han, Sparse projections over graph, in: Proceedings of the 21st AAAI conference on artificial intelligence, 2008.
[37] T. Hastie, R. Tibshirani, J. Friedman., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag, New York, 2001.
[38] A.M. Martinez, R. Benavente, The AR face database, CVC Technical Report No. 24, June 1998.

Zhihui Lai received the BS degree in Mathematics from South China Normal University and MS degree from Ji'nan University, China, in 2002 and 2007, respectively. He is currently pursuing the PhD degree in the School of Computer Science from Nanjing University of Science and Technology (NUST). His research interests include face recognition, image processing and content-based image retrieval, pattern recognition, compressive sense, human vision modelization and applications in the fields of intelligent robot research.

Minghua Wan received the BS degree in Automated Institute from Nanchang University of Aviation in 2003 and the MS degree in Pattern Recognition and Intelligent Systems from Nanjing University of Science and Technology (NUST) in 2007. Now he is pursuing the PhD degree in Pattern Recognition and Intelligent Systems from Nanjing University of Science and Technology (NUST). His current research interests include face recognition and detection, and image processing.

Zhong Jin received the BS degree in Mathematics, the MS degree in Applied Mathematics and the PhD degree in Pattern Recognition and Intelligence System from Nanjing university of Science and Technology (NUST), China, in 1982, 1984, and 1999, respectively. He is a professor in the Department of Computer Science, NUST. He had a stay of 15 months as research assistant at the Department of Computer Science and engineering, the Chinese University of Hong Kong from 2000 to 2001. He visited the Laboratoire HEUDIASYC, Universite de Technologie de Compiegne, France, from October 2001 to July 2002. He visited the Centre de Visio per Computador, Universitat Autonoma de Barcelona, Spain, as the Ramon y Cajal program Research Fellow from September 2005 to October 2005. His current interests are in the areas of pattern recognition, computer vision, face recognition, facial expression analysis and content-based image retrieval.



Jian Yang received the BS degree in Mathematics from the Xuzhou Normal University in 1995. He received the MS degree in Applied Mathematics from the Changsha Railway University in 1998 and the PhD degree from the Nanjing University of Science and Technology (NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a Postdoctoral Researcher at the University of Zaragoza, and in the same year, he was awarded the RyC program Research Fellowship sponsored by the Spanish Ministry of Science and Technology. From 2004 to 2006, he was a Postdoctoral Fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a professor in the School of Computer Science and Technology of NUST. He is the author of more than 50 scientific papers in pattern recognition and computer vision. His current research interests include pattern recognition, computer vision and machine learning.