Sparse bilinear preserving projections

Lai Zhihui[1,2], Chen Qingcai[1],
[1]Bio-Computing Research Center,
Shenzhen Graduate School, Harbin Institute of Technology,
Shenzhen, China
Lai_zhi_hui@163.com; Qingcai.chen@gmail.com

Zhong Jin[2]
[2]Schlol of computer Science,
Nanjing University of Science and Technology,
Nanjing, Jiangsu, China
jinzhong@patternrecognition.cn

***Abstract*—The techniques of linear dimensionality reduction have been attracted widely attention in the fields of computer vision and pattern recognition. In this paper, we propose a novel framework called Sparse Bilinear Preserving Projections (SBPP) for image feature extraction. We generalized the image-based bilinear preserving projections into sparse case for feature extraction. Different from the popular bilinear linear projection techniques, the projections of SBPP are sparse, i.e. most elements in the projections are zeros. In the proposed framework, we use the local neighborhood graph to model the manifold structure of the data set at first, and then spectral analysis and L_1-norm regression by using the Elastic Net are combined together to iteratively learn the sparse bilinear projections, which optimal preserve the local geometric structure of the image manifold. Experiments on some databases show that SBPP is competitive to some state-of-the-art techniques.**

Keywords-sparse projections; Elastic Net; subspace learing; feature extraction; manifold learing

I. INTRODUCTION

In recent years, dimensionality reduction techniques have been an active research field in computer vision and patter recognition. In classical dimensionality reduction techniques, the image matrices are always first transformed into high-dimensional vectors. For the vector-based dimensionality reduction methods, the most well-known ones are principal component analysis [1] and linear discriminant analysis [2]. However, the image matrix-to-vector transformation causes some problems. On the one hand, the structure information embedded in the image matrix is destroyed, and one the other hand, the so called “curse of dimensionality” [3] is also emerged. The high-dimensional small sample size problems [4] directly lead to the over fitting or unstable problems in classical methods. In order to address these problems, researchers proposed a large number of methods, of which the image matrix based methods became the important parts. The matrix-based extensions of the classical methods include two-dimensional PCA (2DPCA) [5] and two-dimensional LDA (2DLDA) [6,7]. With the development of the manifold learning techniques [8-10], many local geometry based method, such as locality preserving projections (LPP) [11,12], neighborhood preserving embedding (NPE) [13], etc., were proposed for feature extraction. Since the image vector based and manifold learning based linear projection methods also encounter the small samples size problem, researchers borrowed the idea of 2DPCA and explored the 2D extensions of the vector based projection methods. By using the locality preserving criterion, 2DLPP [14,15] and its supervised vision [16,17] were proposed and used for image feature extraction and classification. However, these 2D methods can only learn the single projection matrix for dimensionality reduction from only one side of the image matrix. In order to enhance the ability of dimensionality reduction, the bilinear projection methods was recently proposed. The representative bilinear projection methods are generalized low rank approximation matrix (GLRAM) [18], bilinear projection 2DLDA [19] and tensor subspace analysis (TSA) [20], which obtain a good image compression effect with less coefficients.

In recent years, sparse learning methods [21,22], which use the sparse representation as the classifier for object recognition, have attracted much attention in the fields of signal processing, statistic learning, and the related theories and algorithms began to be used in the pattern recognition. Efron et al. [23] proposed the least angle regression (Lars) for sparse regression and feature selection. By combining the L_1- and L_2-norm for regression, Zhou and Hastie proposed the Elastic Net [24] to conquer some potential drawbacks of the Lasso [25]. Zhou et al. proposed sparse PCA (SPCA) [26] by using the Elastic Net for regular principle component learning. Cai et al [27] proposed a unified sparse subspace learning framework with spectral regression on special weighted neighborhood graph. However, two significant disadvantages of these vector based sparse learning methods are that the image-to-vector transform procedure causes the loss of some useful structural information embedding in the original images and that the high-dimensional vector based sparse subspace learning algorithms are very time consuming by using the possible larger cardinalities. As a result, they might not suitable for some real word applications though they can discover the important factor for feature extraction.

Motivated by the bilinear projection methods for image matrix dimensionality reduction, we propose the unsupervised sparse bilinear preserving projections based on the TSA and sparse subspace learning. Different from the existing bilinear projection methods, SBPP can learn the optimal sparse bilinear projection matrices that preserve the local geometric structure.

The rest of the paper is organized as follows. SBPP algorithm is described in Section Ⅱ. In Section Ⅲ, experiments are carried out to evaluate our SBPP algorithm. Finally, the conclusions and future research directions are given in Section Ⅳ.

978-1-4577-0121-4/11/$26.00 ©2011IEEE

II. SPARSE BILINEAR PRESERVING PROJECTIONS

A. Bilinear Dimensionality Reduction

Assume that $X_i \in R^{n\times n}$ $(i=1,2,...,m)$ are the 2D image matrices of the training images. Suppose U is an $n\times d_U$ -dimensional matrix and V is an $n\times d_V$ -dimensional matrix, where each column of U and V is a unitary column vector. Usually, each column vector in U and V are not sparse, i.e. there are almost no zero elements in each projection matrix. The purpose of bilinear sparse dimensionality reduction is to seek two optimal sparse projection matrixes $U=(\phi_1,\phi_2,...,\phi_{d_V})$ and $V=(\varphi_1,\varphi_2,...,\varphi_{d_V})$ to map a 2D image from $n\times n$ -dimensional image space into an $d_U\times d_V$ -dimensional Euclidean space by the following bilinear sparse projections:

$$Y_i = U^T X_i V \qquad (i=1,2,\cdots,m) \tag{1}$$

B. The model of the SBPP

The locality preserving criterion as in TSA is still used in the sparse bilinear preserving projections framework, but additional sparseness constraints are imposed on the objective function of TSA. Of cause, other criterions such as those in [28] can also be used. But in this paper, we only focus on the unsupervised learning. The model of SBPP can be stated as follows:

$$\begin{aligned}&\arg\min_{U,V}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|Y_i-Y_j\right\|_F^2 W_{ij}\\ =&\arg\min_{U,V}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|U^T X_i V-U^T X_j V\right\|_F^2 W_{ij}\end{aligned} \tag{2}$$

Subject to

$$\sum_i D_{ii}\left\|U^T X_i V\right\|_F^2 = 1 \tag{3}$$

$$Card(u)\le K_u \tag{4}$$

$$Card(v)\le K_v \tag{5}$$

where the local neighborhood graph is defined as

$$W_{ij}=\begin{cases} e^{-\frac{\|X_i-X_j\|_F^2}{t}}, & if\ X_i\in N_K(X_j)\ or\ X_j\in N_K(X_i),\\ 0 & , otherwise.\end{cases}$$

v and u are the any column vectors in V and U, and $Card(\varphi)$ (i.e. cardinality) denotes the number of non-zero elements of v or u. $N_K(X_i)$ indicates the set of samples in the K nearest neighbors of the sample X_i, t is a suitable constant and $\|\cdot\|_F$ is the Frobenius norm of the 2D image matrix.

Obviously, directly solving the optimization problem is very difficult. To the best of our knowledge, there has not such method that can directly solve the above optimization problem. In the following paragraphs, we propose an iteration method to obtain its optimal sparse solutions. At first, we reformulate the TSA and then give the iterative optimal sparse solutions of the model. The idea proposed in this paper is: first fix U to obtain sparse V and then fix V to obtain sparse U. Iterating these steps until convergent, we can obtain the optimal sparse bilinear projections. The details are described in the following subsections.

C. Fix U to obtain the sparse V

For a fixed sparse matrix U, denote $U^T X_i = X_{U,i}$, from (2) we have

$$\begin{aligned}&\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|U^T X_i v-U^T X_j v\right\|_F^2 W_{ij}\\ =&\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|X_{U,i} v-X_{U,j} v\right\|_F^2 W_{ij}\\ =& v^T X_U^T[(D-W)\otimes I]X_U v\end{aligned} \tag{6}$$

where $X_U=[X_{U,1}^T,X_{U,2}^T,\cdots,X_{U,m}^T]^T$ is the 2D image training sample matrix, I_n is an identity matrix of order n, operator $\otimes$ is the Kronecher product of the matrices.

In order to remove an arbitrary scaling factor in the embedding, the imposed constraint can be reformulated as the following:

$$\sum_i D_{ii}X_{U,i}^T X_{U,i}=1 \Rightarrow v^T X_U^T(D\otimes I_n)X_U v=1 \tag{7}$$

Now the minimization problem with a fixed u is reduced to be:

$$\arg\min_{v^T X_U^T(D\otimes I)X_U v=1} \quad v^T X_U^T[(D-W)\otimes I]X_U v \tag{8}$$

The transformation vector v (when U is fixed) that minimizes the objection is given by the minimum eigenvector solution to the generalized eigenvalue problem:

$$X_U^T[(D-W)\otimes I]X_U v=\lambda X_U^T(D\otimes I_n)X_U v \tag{9}$$

With the sparseness constraint on v of (9), at last, the model becomes:

$$\begin{cases} X_u^T[(D-W)\otimes I]X_u v=\lambda X_u^T(D\otimes I_n)X_u v\\ Card(v)\le K_v\end{cases} \tag{10}$$

Obviously, if we set $K_v=n$ (i.e. without sparseness constraint), (10) degrades to (9). However, directly solving the generalized eigen-function of (9) can not obtain the sparse projections. Therefore, in order to avoid expensive computation and obtain the sparse solution, we need two theorems presented in Section E. With the two theorems, the spectral analysis and L_1-norm regression by using the Elastic Net can be combined together to learn the sparse projections. The details of obtaining the sparse solutions of (10) are presented in Section E.

D. Fix V to obtain the sparse U

Similarly, for a fixed V, denote $X_i V = X_{V,i}$, from Eq. (2) we have

$$\begin{aligned}&\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|u^T X_i V-u^T X_j V\right\|_F^2 W_{ij}\\ =&\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|u^T X_{V,i}-u^T X_{V,i}\right\|_F^2 W_{ij}\\ =& u^T X_V^T[(D-W)\otimes I]X_V u\end{aligned} \tag{11}$$

With the constraint of

$$\sum_i D_{ii} X_{V,i}^T X_{V,i} = 1 \Rightarrow u^T X_V^T (D \otimes I_n) X_V u = 1 \quad (12)$$

where $X_V = [X_{V,1}, X_{V,2}, \cdots, X_{V,m}]^T$, we have

$$X_V^T [(D-W) \otimes I] X_V v = \lambda X_V^T (D \otimes I_n) X_V v \quad (13)$$

At last, we get the following generalized eigen-equation with sparseness constraint:

$$\begin{cases} X_V^T [(D-W) \otimes I] X_V u = \gamma X_V^T (D \otimes I_n) X_V u \\ Card(u) \le K_u \end{cases} \quad (14)$$

Let $\varphi_{u,1}, \varphi_{u,2}, ..., \varphi_{u,d_U}$ be the sparse solutions of (14). Then $U = [\varphi_{u,1}, \varphi_{u,2}, ..., \varphi_{u,d_U}]$ is the optimal transformation matrix when V is fixed.

Since the process for obtaining the solutions of the proposed method is symmetric and parallel, we donnot distinguish the math symbol for either fixed V to obtain the U or fixed U to obtain the V hereafter.

E. Sparse bilinear preserving projections

In this subsection, we perform graph spectral analysis in our objective functions. With these preparations, L_1-norm regression by using the Elastic Net can be combined together to learn the sparse projections, which will be presented in next subsection. Note that the spectral analysis is suitable for either fixed V or U since we use the same graph and eigenvector for regression. Moreover, for the symmetric, let X_* denotes the X_U or X_V. By using the following two theorems, the sparse projections can be obtained with fast manner. The proof of the theorems, which is very easy, is omitted for saving space.

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

$$[(D-W) \otimes I] y = \lambda (D \otimes I) y \quad (15)$$

If there exist φ satisfying

$$y = X_* \varphi \quad (16)$$

then φ will be the eigenvactor corresponding eigenvalue λ of the (9) or (13).

By Theorem 1, instead of directly solving the generalized eigen-problem in (9) or (13), we can solve the eigen-problem in (9) or (13), and then find φ such that $y = X_* \varphi$. However, directly solving the eigen-problem in (15) is still computationally expensive due to the eigen-decomposition of a very high dimensional matrix. The following theorem will further help us reduce the cost of this task.

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

$$(D-W) z = \lambda D z \quad (17)$$

Let γ be any unit vector in R^n and $y = z \otimes \gamma$, then λ and y are the eigenvalue and corresponding eigenvector of the generalized eigen-problem in (15).

By using Theorem 2, instead of solving the generalized eigen-problem in (15), we can solve the generalized eigen-problem in (17) to save computational cost. As a result, we can obtain λ and z from (17) with lower computational cost because D is a diagonal matrix and the sizes of D and W are smaller than $(D-W) \otimes I$ and $D \otimes I$.

According to the above theorems, spectral regression with L_1-norm penalty combines the spectral graph analysis and regression to provide 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.

Denote $y^l = z \otimes v^l$ $(l = 1, ..., d)$, where z is the eigenvector associated with the first largest nontrivial eigenvalue (the trivial eigenvalue 1 is the largest eigenvalue of (9)) by solving the eigen-equation (9) and v^l 's are orthogonal vectors with $\|v^l\|_2 = 1$. Then, instead of directly solving the generalized eigen-equations with sparseness constraints of (10) or (14), the optimal SBPP projections can be obtained from the following two Elastic Net optimal problems:

$$\varphi_{v,l} = \arg\min_{\varphi_u} (\sum_{i=1}^{m \times n} (x_{u,i}\varphi_u - y_i^l)^2 + \alpha \sum_{j=1}^{n} \bar{\varphi}_{u,j}^2 + \beta \sum_{j=1}^{n} |\bar{\varphi}_{u,j}|) \quad (18)$$

$$\varphi_{u,l} = \arg\min_{\varphi_v} (\sum_{i=1}^{m \times n} (x_{v,i}\varphi_v - y_i^l)^2 + \alpha \sum_{j=1}^{n} \bar{\varphi}_{v,j}^2 + \beta \sum_{j=1}^{n} |\bar{\varphi}_{v,j}|) \quad (19)$$

where y_i^l $(l = 1, ..., d)$ denotes the i th element of y^l, $x_{u,i}$ ($x_{v,i}$)is the i th row of X_U (X_V).The optimal sparse solutions of (18-19) are called sparse bilinear preserving projections (SBPP). Thus, the two optimal sparse projection matrices of SBPP are

$$U = (\varphi_{v,1}, \varphi_{v,2}, ..., \varphi_{v,d}) \text{ and } V = (\varphi_{u,1}, \varphi_{u,2}, ..., \varphi_{u,d}) \quad (20)$$

We summarize SBPP 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 I. SBPP ALGORITHM

Step 1. Constuct the similarity matrix W and compute diagonal matrix D.
Step 2. Compute the first eigenvectors associated with the largest nontrivial eigenvalue by solving the eigenequation (17).
Step 3. Select d mutually orthogonal unit vectors $v_1, v_2, ..., v_d$.
Step 4. Compute $y^1, y^2, ..., y^d$ as $y^i = z \otimes v_i$ $(i = 1, 2, ..., d)$ and initialize $U = I$
Step 5. Iterate until achieving the iteration number set by user or the objective function convergences.
(a) Fix U, Compute the d sparse solutions according to (19) to construct transformation matrix $V = (\varphi_{u,1}, \varphi_{u,2}, ..., \varphi_{u,d})$.
(b) Fix V, Compute the d sparse solutions according to (18) to construct transformation matrix $U = (\varphi_{v,1}, \varphi_{v,2}, ..., \varphi_{v,d})$.
Step 6. Compute $Y_i = U^T X_i V$ as the low-dimension feature.

From the algorithm procedures, it can be found that the objective function of SBPP

$$J(U,V)=\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|U^T X_i V - U^T X_j V\right\|_F^2 W_{ij} \tag{21}$$

will convergence to the local minimum since in each iteration we have

$$J(U^t,V^t)\geq J(U^t,V^{t+1})\geq J(U^{t+1},V^{t+1})>0 \tag{22}$$

where t is the number of iterations. Our experience research indicates that SBPP achieves to the local minimum within several iterations.

III. EXPERIMENTAL RESULTS

To evaluate the proposed SBPP algorithms, we compare it with the unsupervised methods, i.e. 2DPCA, 2DLPP and the bilinear projection method GLRAM and TSA, for image feature extraction and classification. The Yale and CMU PIE face databases was used in the experiments. The Yale face database was used to examine the performance when both facial expressions and illumination were varied. The CMU PIE face database was used to test the performance of the proposed method when there were variations on illumination and pose. Nearest neighborhood classifier with Euclidean distance are used in all the experiments.

A. Experiments on 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 40×40 pixels. Fig.1 shows sample images of one person in the Yale database.

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

In this experiment, 5 images were randomly selected from the image gallery of each individual to form the training sample set. The remaining $11-5=6$ images were used for test. The experiments were repeated 20 times. 2DPCA, GLRAM, 2DLPP, TSA and the proposed SBPP were used for feature extraction. The number of the nearest neighbors was set to be 5, 10 15…. The cardinalities are set to be 1, 2, 4, 6….The maximal average recognition rate of each method and the corresponding dimension are given in Table Ⅱ. Fig. 2 shows the variations of the recognition rates vs. the dimensions of the 5 methods on Yale face database. As it is shown in Table Ⅱ and Fig. 3, the top recognition rates of SBPP are significantly higher than the other methods. Moreover, the dimensions (i.e. the size of the low-dimensional matrix) required to represent the 2D images are also less than the other methods'.

TABLE II. THE AVERAGE RECOGNITION RATES (PERCENT) AND THE CORRESPONDING DIMENSIONS OF 5 METHODS ON THE YALE FACE DATABASE.

Methods	2DPCA	2DLPP	GRLAM	TSA	SBPP
Recognition rate (%)	90.22	91.05	90.72	93.50	95.44
Dimension	40×35	40×22	22×22	20×20	15×15

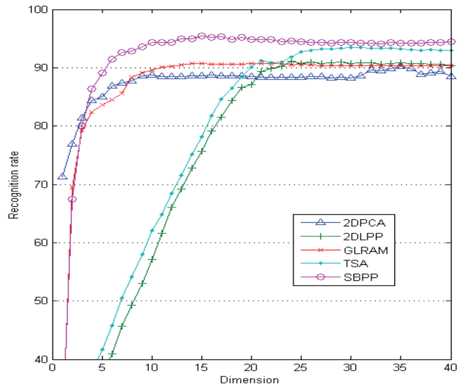

Fig. 2 The variation of the recognition rates vs. the dimensions on Yale face database

B. Experiment on CMU PIE face database

The CMU PIE database [29] contains 68 people, and each person has 13 pose variations that ranged from right to left profile images and 43 different light lighting conditions, which have 21 flashes with ambient light on or off. We chose 23 frontal-view images in our experiments. Original images were aligned, cropped, and then resized to 46×46 pixels. Fig. 3 shows some sample images in CMU PIE face database.

In the experiments, we randomly selected 10 images from each individual for training, while the remaining images of each individual were selected for test. The parameters were set as in Section A. The experiments were repeated 20 times and the average recognition rates and its corresponding dimensions are reported in Table Ⅲ.

Fig.3. Some sample images of one person in the CMU PIE face database.

TABLE III. THE AVERAGE RECOGNITION RATES (PERCENT) AND THE CORRESPONDING DIMENSIONS OF 5 METHODS ON THE CMU PIE FACE DATABASE.

Methods	2DPCA	2DLPP	GRLAM	TSA	SBPP
Recognition rate (%)	57.67	57.24	57.62	59.32	65.30
Dimension	46×46	40×22	46×46	46×46	44×44

It can be seen from the Table III that the first 3 compared methods have almost the same top recognition rates. By integrating the local geometric and bilinear compress for feature extraction, TSA outperforms the first 3 methods, i.e. 2DPCA, 2DLPP and GRLAM. With the sparsity constraint on the projection vectors, SBPP is superior to all the other compared methods. Moreover, in fact, we found in experiments again that the dimensions for representing the low-dimensional features of SBPP are significantly lower than the other methods' when it achieved the best recognition rates of the compared methods. This indicates that the sparse bilinear projections do extract/select the important discriminant variables for image low-dimensional representations.

IV. CONCLUSIONS

In this paper, a novel method called sparse bilinear preserving projections is proposed for bilinear feature extraction from the image matrix. Differing from the existing bilinear projection method such as GLRAM and TSA is that the proposed method can learn the sparse projection matrices for feature extraction. Since the model of the sparse bilinear preserving projections has not close form solutions, we proposed to use the Elastic Net for iterating according to the graph spectral to obtain the optimal sparse solutions. Experiments on the Yale and CMU PIE face databases show that the proposed SBPP is superior to some of the other bilinear feature extraction methods. In the feature, we plan to investigate the supervised sparse bilinear discriminant projection method to further enhance the performance of the sparse bilinear projection framework in image recognition.

ACKNOWLEDGMENT

This work is partially supported by the Natural Science Foundation of China under grant Nos. 61005005, 61005008, 60973098, 60632050, 60705006, 60873151, the Hi-Tech Research and Development Program of China under grant No.2006AA01Z119 and Fu Jian Provincial Department of Science and Technology of China under grant No.JK2010046.

REFERENCES

[1] M. Turk, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 10, Jan. 1991, pp. 358-86.

[2] J.P. Hespanha and D.J. Kriengman, with P.N. Belhumeur, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 19, 1997, p. 711–720.

[3] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: a literature suvey," *ACM Computing Surveys*, vol. 35, 2003, pp. 399-458.

[4] A.K.J. S.J.Raudys, "Small sample size effects in statistical pattern recognition: recommendations for practitioners," *IEEE transactions on pattern analysis and machine intelligence*, vol. 13, 1991, pp. 252-246.

[5] J. Yang, D. Zhang, A.F. Frangi, and J.-yu Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, Jan. 2004, pp. 131-137.

[6] J. Yang, D. Zhang, X. Yong, and J.-yu Yang, "Two-dimensional Discriminant Transform for Face Recognition," *Pattern Recognition*, vol. 38, 2005, pp. 1125-1129.

[7] M. Li and B. Yuan, "2D-LDA: a statistical linear discriminant analysis for image matrix," *Pattern Recognition Letters*, vol. 26, 2005, pp. 527-532.

[8] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15, Jun. 2003, pp. 1373-1396.

[9] S. Roeeis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, 2000, pp. 2323-2326.

[10] J.B. Tenenbaum, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, 2000, pp. 2319-2323.

[11] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, Mar. 2005, pp. 328-40.

[12] Y. Xu, A. Zhong, J. Yang, and D. Zhang, "LPP solution schemes for use with face recognition," *Pattern Recognition*, vol. 43, 2010, pp. 4156-4176.

[13] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," *10th IEEE International Conference on Computer Vision*, vol. 2, 2005, pp. 1208-1213.

[14] S. Pal, with Ben. Yang, "Two-dimensional Laplacianfaces method for face recognition, Pattern Recognition," *Pattern Recognition*, vol. 41, 2008, pp. 3237-3243.

[15] S. Chen, H. Zhao, M. Kong, and B. Luo, "2DLPP: a two-dimensional extension of locality preserving projections," *Neurocomputing*, vol. 70, 2007, pp. 912-921.

[16] Y. Xu, G. Feng, and Y. Zhao, "One improvement to two-dimensional locality preserving projection method for use with face recognition," *Neurocomputing*, vol. 73, 2009, pp. 245-249.

[17] M. Wan, Z. Lai, J. Shao, and Z. Jin, "Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and Palm Biometrics," *Neurocomputing*, vol. 73, 2009, pp. 197-203.

[18] J. Ye, "Generalized low rank approximations of matrices," *Machine Learning*, vol. 61, 2005, pp. 167-191.

[19] J. Ye, R. Janardan, and Q. Li, "Two-Dimensional Linear Discriminant Analysis," *In Proceedings in advaced Neural Information Process Systems 17*, Jul. 2004, pp. 1569-1576.

[20] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," *in Advances in Nerual Information Processing Systems 18*, 2005.

[21] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, Feb. 2009, pp. 210-27.

[22] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Transactions on Circuits and Systems for Video Technology, to be appeared*.

[23] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, 2004, pp. 407-499.

[24] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, 2005, pp. 301-320.

[25] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, 1996, pp. 267-288.

[26] H. Zou, T. Hastie, and R. Tibshirani, "Sparse Principal Component Analysis," *Journal of Computational and Graphical Statistics*, vol. 15, Jun. 2006, pp. 265-286.

[27] D.Cai, "Spectral regression: a unified approach for sparse subspace learning," *in: Proceedings of 2007 International Conference on Data Mining (ICDM07), Omaha, NE*, pp. 73-82.

[28] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, Jan. 2007, pp. 40-51.

[29] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, 2003, pp. 1615-1618.