Open Access

Dimensionality Reduction Based on Low Rank Representation

Cheng Luo[*] and Yang Xiang

School of Electronic and Information Engineering, Tongji University, Shanghai, China, 201804

Abstract: Dimensionality Reduction is a common way to solve the problem of 'curse of dimensions', especially for image processing. Among all these methods, the linear methods are believed to have better performance in actual databases. This paper proposes a novel unsupervised linear dimensionality reduction method that based on low rank representation which aims at finding the subspace structure of the original data sets. This method named LRRDR tries to preserve the subspace structure of the original data therefore is better than the global dimensionality reduction methods like PCA. The experiments compare PCA, NPE, LRRDR and SPP and the results show that LRRDR outperforms the other methods.

Keywords: Dimensionality reduction, low rank representation, face recognition, subspace Structure.

1. INTRODUCTION

The problem of dimensionality reduction which concerns financial data, signal data, digital image and video data pre-processing has been taken into account for a long period of time. Up to now, most methods solving this problem share a common assumption: the high dimensional data is distributed in a linear or a manifold lower dimensional space. Based on this hypothesis, a lot of methods have been proposed, such as Principal Component Analysis(PCA) [1], Locality Preserving Projections(LPP) [2], Fisher Discriminant Analysis(FDA) [3], Neighborhood Preserving Embedding(NPE) [4, 5], Isomap [5], Locally Linear Embedding(LLE) [6], Laplacian Eigenmaps(LE) [7, 24] etc. In this paper, we focus on one sort of these methods named "linear dimensionality reduction" which contains PCA, LPP, FDA, NPE etc. The reason is nonlinear dimensional reduction methods cannot learn an obvious "transformation function" that could convert any input data into a lower dimensional space [8]. Furthermore, some researchers argued that nonlinear reduction methods always perform worse than linear methods in actual data samples, although they outperform in a lot of artificial data sets [9]. Based on this consideration, this paper proposes a linear dimensionality reduction method can be called LRRDR (Dimensionality Reduction by Low Rank Representation).

The originality of the idea of this method can be traced to NPE which try to keep the global Euclidean structure of sample data during dimensionality reduction [4]. However, NPE and a similar method LPP need a hyper-parameter ' k ' that indicates the size of the locality and need to be predefined. Furthermore, NPE assumes all the data points distributed in only one space which is not true for lots of real data samples.

Recently, some researchers proposed a new method named SPP (Sparsity Preserving Projections) [10]. This method is the combination of NPE and sparse representation, a nearly brought up quantization method which becomes one of the most influential machine learning paradigm rapidly. The authors argued that the l_1 regularization essentially encodes the prior knowledge of sparsity, allowing it to extract more discriminating information from the data than NPE does. After SPP, many of a method based on sparsity representation has been proposed [11-13]. However, all of the methods described previously did not consider the structure of the data which lies in the fundamental of dimensionality reduction as we stated earlier. The structure of objects in the images has been deeply studied in [14]. The authors claimed the objects in an image are actually lie in a low rank subspace of the origin, and happened to be in a higher one due to corruptions. Based on this consideration, LRR (Low Rank Representation) [15] has been proposed. Comparing to SR, LRR is better at capturing global structures of data. It does not need the number of subspaces and it does not need the number of dimensions of each subspace, and still this method can find a good reconstruct of the original data and consequentially it wipes out the noises and discovery the structure of the data.

In light of this consideration, we propose a novel dimensionality reduction method LRRDR. It has the following virtues:

1) LRRDR can preserve the global as well as the local structures of the original data. As is known to all, PCA is a global dimensionality reduction method and it can preserve the maximum variations of the original data when all the data points distributed in the same space [16]. But that is not hold for a large variety of data sets. LRRDR is a global method in nature and it also keeps the local structure in mind and will preserve it during the dimension reduce.

2) As LRR can recover the structure of data from noises [17], LRRDR is free from the sufferings of different of corruptions. Compare to PCA, LPP and NPE etc, LRRDR is robust and do not need some sophisticated processing and is

simple to implement. Details of kinds of noises can be found in [17].

3) LRRDR does not have the difficult in parameter selection. As has mentioned previously, LPP and NPE need to select the size of neighborhood and the type of kernel, which is very time consuming and requires a high professionalism of researcher.

The structure of this paper is organized as follows: section 2 reviews related work, to be concretely, we discuss PCA, NPE and SPP in detail because we will use them as a comparing work to LRRDR, section 3 introduces LRR which lies in the central of out reducing method and then we discuss the algorithm of LRRDR, section 4 is the experiment result and discussions, and section 5 is the conclusion.

2. RELATED WORK

In this section we review some existing dimensionality reduction methods and because we concentrate on linear unsupervised versions so we discuss this kind only. PCA is one of the most famous and popular dimensionality reduction method and most works are related to it directly or indirectly. It can be seen as a benchmark of the performance for most dimensionality reduction methods [8]. NPE is another well known dimensionality reduction method, different from PCA, it tries to keep the local structure of sample data during dimensionality reduction [4], so we will also discuss it. The last one we will talk about is SPP which can be seen as a development of NPE.

2.1. PCA

Given a set of data points (vectors) $x_1, x_2, L\ x_m$ of dimension n, PCA try to find a transforming matrix $U \in R^{n \times d}$ such that $y_i = U^T x_i$ ($i = 1, 2, L\ , m$), and the variance of every feature of y_i achieves maximum, or formally we can state it as follows:

$$\max_U \frac{1}{m} tr \sum_{i=1}^{m} (y_i - \bar{y})(y_i - \bar{y})^T \qquad (2.1)$$

where $\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_i$ is the mean of the data points. The solution to this problem is the eigenvectors of the covariance matrix of $\{x_i\}_{i=1}^{m}$, which we can refer to as Σ. Concretely, the transforming matrix U is made up of the are eigenvetors $u_1, u_2, L\ u_d$ of Σ which are corresponding to the largest eigenvalues of Σ.

2.2. NPE

NPE is another linear unsupervised dimensionality reduction method, and its main idea is to preserve the local neighborhood structure of the data. First it constructs an adjacency graph whose nodes are the points of the data sets, and two nodes on the graph are adjacent if and only if they are near enough. Here 'near enough' means either one of the two statements below: there exists a small number ε such that

the Euclidean distance of the two points smaller than ε or either one is among the other's k nearest neighbor [4]. Here ε and k are parameters that should be set in advance. Then a weight matrix W should be made subject to:

$$\min_W \sum_i \left\| x_i - \sum_{x_j \in B(i)} W_{ij} x_j \right\| \qquad (2.2)$$

where $B(i)$ is the neighborhood of x_i.

The essence of NPE is to preserve the weight matrix W when reduce the dimensionality. Concretely, the transform matrix U should satisfy the following constraint:

$$\min_U \sum_i \left\| y_i - \sum_{x_j \in B(i)} W_{ij} y_j \right\| \qquad (2.3)$$

where $y_i = U^T x_i$ ($i = 1, 2, L\ , m$).

The above problem can be transform to this generalized eigenvector problem:

$$XMX^T u = \lambda XX^T u \qquad (2.4)$$

where

$$M = (I - W)^T (I - W)$$

$$I = diag(1, 1, L\ 1)$$

And the columns of U is the eigenvectors of (2.4) which are corresponding to the smallest k eigenvalues.

2.3. SPP

SPP is based on NPE and sparsity reconstruction, the only different of SPP and NPE lies on the weight matrix W. Different from NPE, SPP does not need to construct the adjacency matrix and it derives W from the sparsity reconstruction. For SPP the columns of W should satisfy:

$$\min_{w_i} x_i = Xw_i + \lambda \| w_i \|_1 \qquad (2.5)$$

where w_i is the i th column of W, λ is a parameter and $\| \cdot \|_1$ is the l_1 norm. Then SPP need to solve the similar generalized eigenvector problem like NPE.

The authors of SPP argued that regularization essentially encodes prior knowledge of sparsity, allowing it to extract more discriminating information from the data than NPE does [10].

3. DIMENSIONALITY REDUCTION BASED ON LRR

3.1. Low Rank Representation

Low rank representation (LRR) is motivated from the subspace segmentation problem [18], and it seeks the lowest-rank representation among all the candidates that represent all vectors as the linear combination of the bases in a dictionary [15]. The assumption is that mixed data are modeled as a set of independent samples drawn from a mixture of probabilistic distributions. However, the dimensions of each subspace are unknown and the number of subspaces is also

remained to be discovered. The formal statement of the problem stated above is to find the reconstruction matrix W:

$$X = AW \tag{3.1}$$

where X is the data matrix, A is the dictionary. To find the global structure of the data set, we put a constraint of the weight matrix W that we want the rank of W to become the minimum, formally,

$$\min_W rank(W) \tag{3.2}$$

$$s.t. X = AW$$

The problem state above is the low rank representation of X, because we want the rank of the reconstruct matrix W to be 'low'. The above optimization problem is difficult to solve due to the discrete nature of the rank function [15, 25]. But we can transform it to a concave optimization problem:

$$\min_W \|W\|_* \tag{3.3}$$

$$s.t. X = AW$$

Where $\|\cdot\|_*$ is the nuclear norm [19], i.e. the sum of the singular value of a matrix. For real applications the data is always contaminated by different kinds of noises, and to promote the robustness of this algorithm, we can add a error term in the objective function:

$$\min_{W,E} \|W\|_* + \lambda \|E\|_{2,1} \tag{3.4}$$

$$s.t. X = AW + E$$

Where E is the error term, $\lambda > 0$ is a parameter to represent the estimation of the magnitude of noise, $\|E\|_{2,1} = \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} ([E]_{ij})^2}$ is $l_{2,1}$-norm, it encourage some of the columns to be 0, when they are disturbed severely. In some cases, when data contains stochastic error [17], we can replace $l_{2,1}$-norm with l_1-norm.

The optimization problem stated above is sound because it can be proved that when the data is not contaminated, the optimum solution to problem

$$\min_W \|W\|_* \tag{3.5}$$

$$s.t. X = XW$$

is diagonal block, and every block forms a subspace of the original data set.

3.2. Dimensionality Reduction based on LRR

The dimensionality reduction method based on LRR tries to preserve subspace structure of the origin data. The idea behind it is straight-forward, first use the low rank representation to reconstruct the data and discover the subspace structure of the data, and we assume every point (vector) can be represented by the points in the same subspace with it, then we reduce the dimensionality trying to preserve that structure.

Suppose $y = (y_1, y_2, L, y_m)^T$ is the set of one dimensional data reduced from the origin $X = (x_1, x_2, L, x_m)$. We use Mean Square Error (MSE) as the cost function:

$$E(y) = \sum_{i=1}^{m} (y_i - \sum_{j=1}^{m} W_{ij} y_j)^2 \tag{3.6}$$

The formula above can be vectorized as:

$$E(y) = (y - Wy)^T (y - Wy) \tag{3.7}$$

Suppose $y_i = u^T x_i$, where u the transform function and we have:

$$y^T = u^T X \tag{3.8}$$

Insert (3.8) to (3.7) we have:

$$E(y) = (y - Wy)^T (y - Wy) = y^T (I - W)^T (I - W) y$$

$$= u^T X (I - W)^T (I - W) X^T u$$

Note the scale of the cost function $E(y)$ can be arbitrarily large or small due to the scale of y, we need to add the constraint $y^T y = 1$, or

$$u^T XX^T u = 1 \tag{3.9}$$

Denote $D = (I - W)^T (I - W)$ and the question can be stated as:

$$\min_u E(u) = u^T XDX^T u \tag{3.10}$$

$$s.t. u^T XX^T u = 1$$

Question (3.10) can be solved by the generalized eigenvector problem stated below:

$$XDX^T u = \lambda XX^T u \tag{3.11}$$

To reduce x_i to a k dimensional vector, the transform function is matrix composed of the k smallest eigenvectors.

3.3. LRRDR Algorithm

The algorithm has two steps, the first is to compute the reconstruction matrix W of (3.4), and this problem can be solved by augmented Lagrange Method (ALM) or IALM (Inexact Augmented Lagrange Method) [20]. In this paper, we use IALM to accelerate the convergence speed. In (3.11) the generalized eigenvector problem often suffers from the computational degeneration of XX^T. When X is not row full rank, we can decompose X using singular value decomposition as $X = USV^T$, where $U \in R^{n \times r}$ is the eigenvectors of matrix XX^T, $S \in R^{r \times r}$ is a diagonal matrix whose elements in the diagonal is the square root of the eigenvalues of $X^T X$, $V \in R^{r \times m}$ is the eigenvectors of matrix $X^T X$. As S and V^T are both of full rank, we can use the full rank matrix

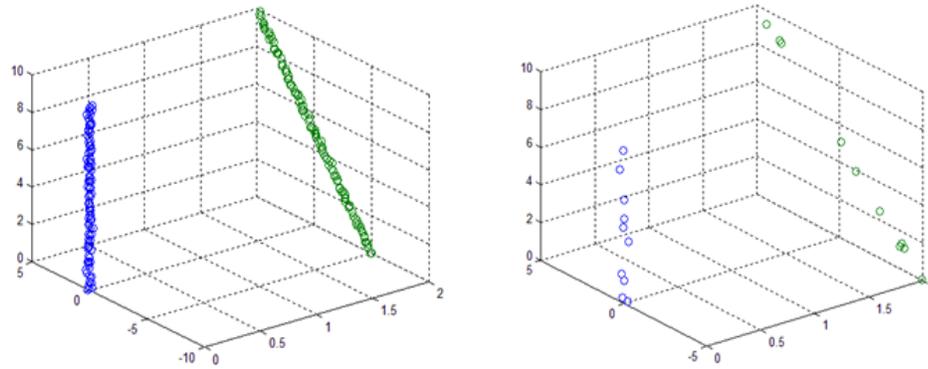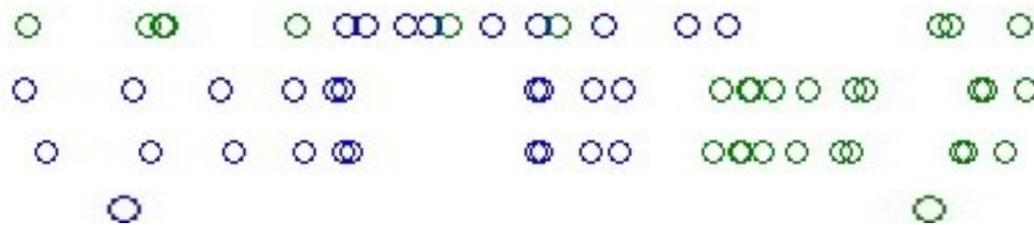**Fig. (1).** Three dim data, left: the data, right: the sample.

Fig. (2). Performance of different reduction methods, from top to bottom: PCA, NPE, SPP, LRRDR.

$\tilde{X} = U^T X = SV^T$ to replace X. The algorithm of LRRDR is stated below:

Algorithm 1
Input: data matrix $X \in \pmb{R}^{n \times m}$
Use IALM compute reconstruction matrix W of (3.4)
Compute k eigenvectors $U = [u_1, u_2, \text{L}, u_k]$ of (3.11) whose corresponding eigenvalues are of k least.
Reduce the dimensionality of X using $Y = UX$
Output: the reduced data set Y

4. EXPERIMENTS

4.1. Toy Experiments

This experiment is about transforming a set of three dimensional data into a set of one dimensional data. The three dimensional data is indeed lie in two different one dimensional subspaces, and we want to transform this data into one dimensional without breaking the subspace structure. We first construct the data set and then add white noise of standard variance 0.1 to it (the left of Fig. **1**). Then we randomly choose 10 points from each class, and get the data sets below (the right of Fig. **1**). We use four methods PCA, NPE, SPP and LRRDR to reduce the data. For NPE we empirically set $k = 5$ as the neighborhood size, and the result shows LRRDR successfully reduce the two classes without mixing up and the other three methods fail that. Fig. (**2**) shows the result of dimensionality reduction of each method.

The analysis of the result of the results these four methods are stated below.

1. PCA is a global reduction method, it assumes that all of the data is distributed in one space, and it seeks the direction whose variance is of the maximum, and it abandons the other two directions. In this example, the direction PCA seeks is actually the direction of the green class lie. Although it is the direction of the maximum variance, it mixes up the two classes.

2. Both NPE and SPP have better performance because these two methods preserve the local structure and consequently, the two classes are linear separable after being transformed. But they do not perform as well as LRRDR does because the structure they preserve is not the subspace structure we need.

3. The performance of LRRDR is rather impressive and the two classes are almost compressed into two points because it successfully discovered the two subspaces and preserves it during dimensional reduction. In fact, the reconstruction matrix is block diagonal and each block is corresponding to a subspace.

4.2. Face Recognition

4.2.1. Data Sets Description and Parameters Sets

In this paper, we use three prevalent data sets YaleB Face Database [21] and Yale Face database [22]. For YaleB, this data set contains 2432 photos of 38 individuals, and each person takes 64 photos under different illuminations. We list 22 photos of one person for example. To reduce the computing complexity, we use the cropped images with the resolution of 50 * 50. Except for that, we did not any other preprocessing. For Yale, this data set contains 165 photos of 15 individuals. Every person has been filmed under 11 different
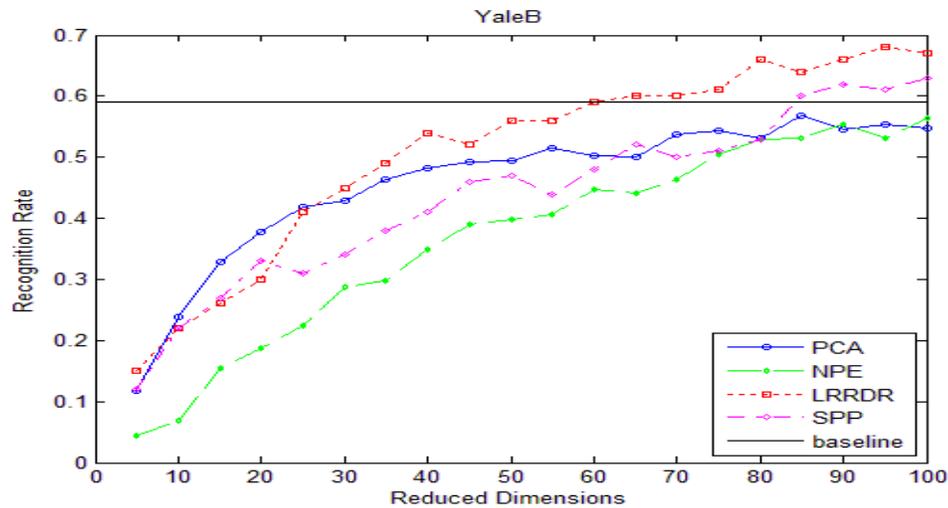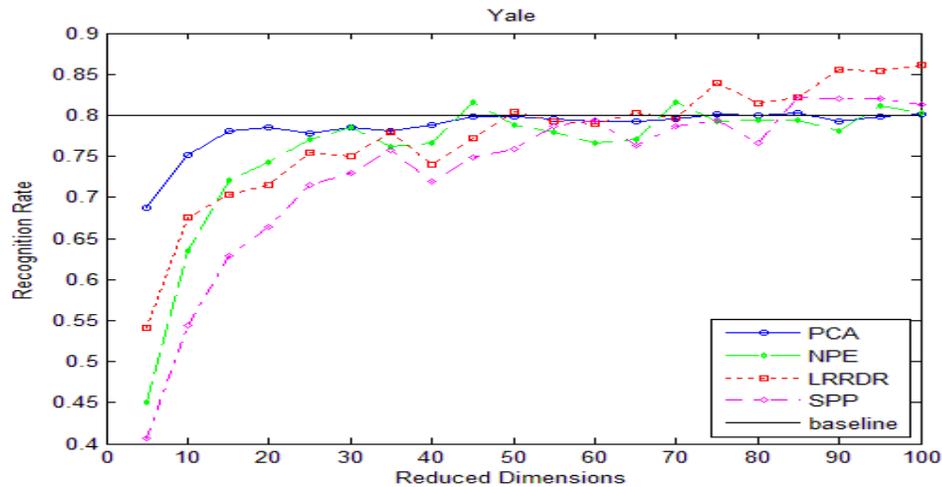
Fig. (3). Recognition rate for YaleB.

Fig. (4). Recognition rate for Yale.

conditions: happy, sleepy, left-light, wearing no glasses, normal, right-light, sad, center-light, surprised wearing glasses, and wink. The size of each photo is 30 * 30 and all the photos have been scaled to [0, 1]. When taking experiments, we randomly pick half of images of every individual as the training data for every data set and the rest are spared for tests. To prevent from being disturbed by random error, we repeat the computing of every setting of dimension 10 times and take the average as the result of recognition.

4.2.2. Parameter Selection

For PCA there is no parameter to be selected. For NPE, it has two pre-set parameters, the size of neighborhood k and the kernel width t. In our experiment, the parameters are selected by attempting every number of integers from 1 to 10 and the best one is kept for test. For SPP and LRRDR there is one parameter needs selected carefully is the error tolerant parameter λ, and we tried from 10^{-5} to 1 increased 3 times each time.

For classification, we make use of KNN with the number of adjacent neighbors set to 1 for simplicity. For every data set, we compute the baseline of recognition with KNN where $k = 1$ before dimensionality reduction.

4.2.3. Experiment Result and Discussion

We compute a baseline for every data set as a comparison; it is the result of 1-NN without any dimensionality reduction. Similar with other experiments, we randomly choose half of the images of every class (individual) as training, and the rest are used as test and we get the baseline for every data set.

PCA is one of the easiest to implement, it has no parameter to pre-define, and it converges very fast, in our experiment, it is faster than most of the other dimensionality reduction method, and the results of PCA float around the baseline. In our experiment, the performance of NPE differs a lot from the two data sets. For YaleB, NPE performs very bad, much

Table 1. Best Recognition Rate Based on 1-NN.

Methods		PCA	NPE	LRRDR	SPP	baseline
YaleB	Accuracy	0.56	0.56	0.68	0.63	0.59
	Dimensions	85	100	95	100	no
Yale	Accuracy	0.81	0.82	0.86	0.83	0.81
	Dimensions	80	70	100	90	no

lower than both the baseline and PCA due to the illumination changes greatly. But when it comes to the data set of Yale, NPE behaves very well, and the recognition rate is higher than PCA and the baseline. SPP has only one parameter to tune and consequently is easier to implement, however, it converges rather slow that it takes ten or hundred times slower than PCA and NPE.

LRRDR outperforms in both of the two datasets. It is robust to the noises caused by poor lighting conditions, and thus the recognition rate is higher than PCA as it is sensitive to noises [23]. Moreover, the essential reason that LRRDR outperforms the other dimensionality reduction methods is that it discovers the subspace structure and preservers it while reducing. The recognition rate of the data set YaleB for different dimensions is shown in Fig. (**3**) and the recognition rate of the data set Yale is shown in Fig. (**4**). The best recognition rate for these data sets is shown in Table **1**.

CONCLUSION

In this paper, we propose a novel dimensionality reduction method named LRRDR, which is an unsupervised linear dimensionality reduction method that is based on low rank representation. It is a global method and has only one parameter to tune and so it is easy to implement. It is robust to noises that influence other methods such as PCA. As the structure it discovers is global as well as local, it outperforms some local structure preserving methods such as NPE. Experiments show that for artificial data set and some face databases like YaleB and Yale, LRRDR outperforms PCA, NPE and SPP.

However, there are still some questions remain. As it is very hard to discover the subspaces when the data sets contain front-view faces as well as side faces, the recognition rate of LRRDR is not so good as it performs in YaleB or Yale. To overcome this shortcoming, it is need to consider the supervised information, and we leave it to future work.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   I.T. Jolliffe, "*Principal component analysis*", Springer-Verlag: New York, vol. 487, 1986.

[2]   P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, pp. 153-160, 2004.

[3]   S. Mika, R. Gunnar, W. Jason, S. Bernhard, and M. Klaus, "Fisher discriminant analysis with kernels", *Neural Networks for Signal Processing IX*, 1999: IEEE.

[4]   X. He, D. Cai, S. Yan, H. Zhang, "Neighborhood preserving embedding *Computer Vision*", ICCV. 10th *IEEE International Conference on*, vol. 2, 2005.

[5]   J.B. Tenenbaum, V. De Silva and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, vol. 290, no. 5500, pp. 2319-2323, 2000.

[6]   S.T. Roweis, and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding,'' *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000.

[7]   M. Belkin, and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, pp. 585-591, 2001.

[8]   L.K. Saul, K.Q.Weinberger, J.H. Ham, F. Sha, D.D. Lee, "Spectral methods for dimensionality reduction", *Semisupervised Learning*, pp. 293-308, 2006.

[9]   L. Van der Maaten, E.O. Postma and H.J. Van den Herik, "Dimensionality reduction: A comparative review," *Journal of Machine Learning Research*, vol. 10, pp. 1-41, 2009.

[10]   L.S. Qiao, S.C. Chen and X.Y. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331-341, 2010.

[11]   P. Niyogi, Locality preserving projections, "*Advances in Neural Information Processing Systems*", vol. 16, pp. 153-160, 2004.

[12]   L.M. Zhang, S.C. Chen and L.S. Qiao, "Graph optimization for dimensionality reduction with sparsity constraints," *Pattern Recognition*, vol. 45, no. 3, pp. 1205-1210, 2012.

[13]   P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, pp. 153-160, 2004.

[14]   L.M. Zhang, L.S. Qiao and S.C. Chen, "Graph-optimized locality preserving projections", *Pattern Recognition*, vol. 43, no. 6, pp. 1993-2002, 2010.

[15]   J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, "Robust face recognition *via* sparse representation", *Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227, 2009.

[16]   J.P. Costeira, and T. Kanade, "A multibody factorization method for independently moving objects", *International Journal of Computer Vision*, vol. 29, no. 3, pp. 159-179, 1998.

[17]   Liu, G., Z. Lin and Y. Yu. "Robust subspace segmentation by low-rank representation", *Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171-184, 2013.

[18]   R.H. Keshavan, A. Montanari and S. Oh, "Matrix completion from noisy entries", *The Journal of Machine Learning Research*, vol. 11, pp. 2057-2078, 2010,

[19]   G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, "Robust recovery of subspace structures by low-rank representation," *arXiv preprint* arXiv:1010.2955, 2010.

[20] G. Liu, H. Xu and S. Yan, "Exact subspace segmentation and outli-er detection by low-rank representation," *arXiv preprint* arXiv:1109.1646, 2011.

[21] M. Fazel, "Matrix rank minimization with applications", Ph.D. thesis, Stanford University, 2002.

[22] Z. Lin, M. Chen and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices", *arXiv preprint* arXiv:1009.5055, 2010.

[23] A.S. Georghiades, P.N. Belhumeur, and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.

[24] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces *vs.* fisherfaces: Recognition using class specific linear projection", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711-720, 1997.

[25] E.J. Candès, X. Li, Y. Ma, J. Wright, "Robust principal component analysis," *arXiv preprint* preprint arXiv:0912.3599, 2009.