Orthogonal locality minimizing globality maximizing projections for feature extraction

Feiping Nie
Shiming Xiang
Yangqiu Song
Changshui Zhang
Tsinghua University
Department of Automation
Beijing 100084, China
E-mail: feipingnie@gmail.com

Abstract. Locality preserving projections (LPP) is a recently developed linear-feature extraction algorithm that has been frequently used in the task of face recognition and other applications. However, LPP does not satisfy the shift-invariance property, which should be satisfied by a linear-feature extraction algorithm. In this paper, we analyze the reason and derive the shift-invariant LPP algorithm. Based on the analysis of the geometrical meaning of the shift-invariant LPP algorithm, we propose two algorithms to minimize the locality and maximize the globality under an orthogonal projection matrix. Experimental results on face recognition are presented to demonstrate the effectiveness of the proposed algorithms.
[DOI: 10.1117/1.3067869]

Subject terms: locality preserving projections; subspace learning; shift invariant; feature extraction; face recognition.

Paper 080745R received Sep. 24, 2008; revised manuscript received Nov. 21, 2008; accepted for publication Nov. 22, 2008; published online Jan. 26, 2009.

1 Introduction

Linear-feature extraction and dimensionality-reduction techniques are very important approaches to deal with high-dimensional data, such as texts, images, and videos. In the past decades, many supervised algorithms have been proposed for the purpose of classification. Linear-discriminate analysis (LDA)[1] is one of the most popular ones. It has been successfully applied in many classification tasks such as face recognition. Recently, an algorithm called locality-preserving projections (LPP)[2] was developed and is becoming a frequently used linear-feature extraction technique for many applications.

For a linear-feature extraction algorithm, data shifted by an arbitrary constant vector should not influence the learned projection matrix. However, the original LPP algorithm proposed in Ref. 2 does not satisfy this shift-invariance property. We analyze the reason from the derivation of LPP. LPP is derived from Laplacian eigenmap,[3] while in Laplacian eigenmap, there is an implicit relationship which is not considered in the derivation of the original LPP algorithm. Taking the relationship into account, we derive a shift-invariant LPP algorithm that satisfies the shift-invariance property.

The geometrical meaning of the derived shift-invariant LPP is also revealed in this paper. We reveal that it is to minimize the sum of the Euclidean distances between data pairs which are local to each other, while the weighted covariance matrix of data is fixed to a constant matrix. Yang et al. recently proposed an unsupervised, discriminant projection algorithm,[4] which is to solve the following optimization problem:

$$\mathrm{W} = \arg \min_{\mathrm{W}^T\mathrm{X}\mathrm{L}_t\mathrm{X}^T\mathrm{W}=\mathrm{I}} \mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W}), \tag{1}$$

where X is the data matrix, L is a Laplacian matrix, and $\mathrm{L}_t = \mathrm{I} - (1/n)\mathbf{1}\mathbf{1}^T$ is the centralization matrix. The algorithm is said to have extended LPP to obtain a globally maximizing and locally minimizing projection. In fact, from the geometrical meaning of the shift-invariant LPP, we know that, although the shift-invariant LPP is derived from the motivation of locality preserving, it preserves the local and global property simultaneously. We see in Section 3 that the optimization problem (1) is just a simplified case of the optimization problem (11) in the shift-invariant LPP. Similar to LPP, the learned projection matrix is not orthogonal.

Recently, the orthogonal method has been attracting a great deal of interest as the orthogonality is desirable and often demonstrates good performance empirically.[5–9] An orthogonal LPP algorithm is proposed in Ref. 10 to solve the following optimization problem:

$$\mathrm{W} = \arg \min_{\mathrm{W}^T\mathrm{W}=\mathrm{I}} \mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W}). \tag{2}$$

From the optimization problem, we can see that the algorithm only performs a locality-minimizing projection and does not take the global information into account, which might be insufficient to gain the discriminative power.

Cai et al. also proposed another orthogonal LPP algorithm recently.[11] They used a step-by-step procedure to obtain a set of orthogonal projections $\{w_1, w_2, \ldots, w_m\}$. After calculating the first $k-1$ projections $\{w_1, w_2, \ldots, w_{k-1}\}$, the k'th projection w_k is calculated by solving the following optimization problem

0091-3286/2009/$25.00 © 2009 SPIE

$$w_k = \arg \min_{W_{k-1}^T w_k = 0} \frac{w_k^T X L X^T w_k}{w_k^T X D X^T w_k}, \tag{3}$$

where D is the degree matrix on the graph and $W_{k-1} = [w_1, w_2, \ldots, w_{k-1}]$. The step-by-step procedure makes the algorithm computationally more expensive and makes the objective with regard to W to be optimized not clear. The shift-invariance property is also not taken into account in their algorithm.

Inspired by the geometrical meaning of the shift-invariant LPP, in this paper we propose two novel algorithms to minimize the locality and maximize the globality simultaneously under an orthogonal projection matrix. Experiments on face recognition are presented, and the experimental results demonstrate that the proposed algorithms are effective for feature extraction.

The rest of this paper is organized as follows. In Section 2, we revisit the derivation of LPP, and the shift-invariant LPP is derived in Section 3. In Section 4, we reveal the geometrical meaning of the derived shift-invariant LPP. In Section 5, we propose two novel algorithms to perform locality-minimizing, globalitymaximizing projections. In Section 6, we present the experiments on face recognition to verify the effectiveness of the proposed algorithms. Finally, we conclude this paper in Section 7.

2 LPP Revisited

LPP is a linear-feature extraction algorithm[2] derived from Laplacian eigenmap.[3] Given n data points $x_i \in \mathbb{R}^d (i = 1, 2, \ldots, n)$, in order to discover the corresponding $y_i \in \mathbb{R}^m$ in the low-dimensional manifold for x_i, Laplacian eigenmap is used to solve the following optimization problem:

$$Y = \arg \min_{YDY^T = I} \operatorname{tr}(YLY^T), \tag{4}$$

where $Y = [y_1, y_2, \ldots, y_n] \in \mathbb{R}^{m \times n}$, D is a diagonal matrix, $D_{ii} = \Sigma_j A_{ij}$, and $L = D - A$ is a Laplacian matrix defined on a graph constructed by the given data. The affinity matrix A could be defined by

$$A_{ij} = \begin{cases} e^{-\|x_i - x_j\|^2/t} & x_i \text{ and } x_j \text{ are neighbors} \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

where t is the parameter of the heat kernel and $\|\cdot\|$ denotes the 2-norm of the vector (i.e., $\|x\|^2 = x^T x$).

The map from $x \in \mathbb{R}^d$ to $y \in \mathbb{R}^m$ learned by Laplacian eigenmap is nonlinear, and the aim of LPP is to find a linear map to approximate this nonlinear map. Denoting the data matrix $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d \times n}$, LPP assumes the map from $x \in \mathbb{R}^d$ to $y \in \mathbb{R}^m$ is linear, and let

$$Y = W^T X, \tag{6}$$

where $W = [w_1, w_2, \ldots, w_m] \in \mathbb{R}^{d \times m}$ is a projection matrix.

Imposing linear relationship (6) on optimization problem (4), LPP is used to solve another optimization problem as follows:

$$W = \arg \min_{W^T X D X^T W = I} \operatorname{tr}(W^T X L X^T W) \tag{7}$$

where $tr(\cdot)$ denotes the trace operator of the matrix.

The solution to this optimization problem is finally reduced to solving the following eigen-decomposition problem:

$$XLX^T W = XDX^T W \Lambda \tag{8}$$

where Λ is the eigenvalue matrix and W is the corresponding eigenvector matrix of $(XDX^T)^{-1} XLX^T$.

3 Shift-Invariant LPP

For a linear-feature extraction algorithm, data shifted with an arbitrary constant vector should not influence the result of the learned projection matrix W. That is to say, when data are shifted by $\tilde{x} = x - c$, where c is an arbitrary constant vector, the learned projection matrix W should not be changed. The shift-invariant property can be well understood. When the given data are shifted with an arbitrary constant vector, the Euclidean distances between the data pairs are unchanged, and thus the structure of the data is unchanged. Therefore, the learned projection matrix should also remain unchanged.

Two popular linear-feature extraction methods, principal component analysis (PCA) and LDA, both satisfy this property.

When data are shifted by $\tilde{x} = x - c$, the data matrix becomes $\tilde{X} = X - c1^T$, where $1 = [1, 1, \ldots, 1]^T$. Then we have the following theorem:

Theorem 1: If L is a Laplacian matrix, then $\tilde{X} L \tilde{X}^T = XLX^T$, where $\tilde{X} = X - c1^T$.

Proof. According to the definition of the Laplacian matrix,[12] we know $L1 = 0$. Then, $\tilde{X} L \tilde{X}^T = (X - c1^T) L (X - c1^T)^T = XLX^T - XL1c^T - c1^T LX + c1^T L1c^T = XLX^T$.

It is well-known that the projection matrix W in PCA is formed by the eigenvectors of the total scatter matrix S_t, and the projection matrix W in LDA is formed by the eigenvectors of $S_w^{-1} S_b$, where S_w is the within-class scatter matrix and S_b is the between-class scatter matrix. From the view of graph embedding, we know that $S_t = XL_t X^T$, $S_w = XL_w X^T$, and $S_b = XL_b X^T$, where L_t, L_w, and L_b are all Laplacian matrices.[13] According to Theorem 1; if X is substituted with $X - c1^T$, the matrices S_t, S_w, and S_b, are all unchanged, and thus the learned W in LDA and in PCA are unchanged. Therefore, PCA and LDA both satisfy the shift-invariance property.

LPP is also a linear-feature extraction algorithm. However, the original LPP algorithm proposed in Ref. 2 does not satisfy this property. The reason is that LPP, which is used to solve optimization problem (7), is derived from optimization problem (4), while the solution to problem (4) implicitly satisfies the following relationship:

$$YD1 = 0 \tag{9}$$

Imposing Eq. (6) on optimization problem (4) cannot guarantee that implicit relationship (9) is satisfied.

In order to solve this problem and guarantee that implicit relationship (9) is satisfied in the solution to the derived problem of LPP, we replace the linear constraint in Eq. (6) with another linear constraint as follows:

$$\mathrm{Y} = \mathrm{W}^T\mathrm{X}\left(\mathrm{I} - \frac{1}{1^T\mathrm{D}1}\mathrm{D}11^T\right), \tag{10}$$

where I denotes an $n \times n$ identity matrix and $1 = [1, 1, \ldots, 1]^T$.

Obviously, Eq. (10) is also a linear map from $x \in \mathbb{R}^d$ to $y \in \mathbb{R}^m$, and it satisfies Relationship (9) in all cases.

Imposing Eq. (10) instead of Eq. (6) on the optimization problem (4), we can obtain a shift-invariant LPP algorithm, which is used to solve the following optimization problem:

$$\mathrm{W} = \arg \min_{\mathrm{W}^T\mathrm{X}\mathrm{L}_d\mathrm{X}^T\mathrm{W}=\mathrm{I}} \mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W}), \tag{11}$$

where $\mathrm{L}_d = \mathrm{D} - (1/1^T\mathrm{D}1)\mathrm{D}11^T\mathrm{D}$.

Theorem 2: L_d is a Laplacian matrix.

Proof. First, we prove that L_d is positive semidefinite. For any vector $x \in \mathbb{R}^n$, we have

$$\begin{aligned} x^T\mathrm{I}_d x &= x^T\mathrm{D}x - \frac{1}{1^T\mathrm{D}1}x^T\mathrm{D}11^T\mathrm{D}x \\ &= \sum_{i=1}^{n} \mathrm{D}_{ii}x_i^2 - \frac{1}{\Sigma_{i=1}^{n}\mathrm{D}_{ii}}\left(\sum_{i=1}^{n} \mathrm{D}_{ii}x_i\right)^2 \\ &= \frac{1}{\Sigma_{i=1}^{n}\mathrm{D}_{ii}} \sum_{i,j=1}^{n} \mathrm{D}_{ii}\mathrm{D}_{jj}(x_i - x_j)^2 \\ &\geqslant 0, \end{aligned} \tag{12}$$

where x_i is the i'th element in x. So L_d is positive semidefinite.

On the other hand, we can see that $\mathrm{L}_d 1 = \mathrm{D}1 - (1/1^T\mathrm{D}1)\mathrm{D}11^T\mathrm{D}1 = 0$. Therefore, L_d is a Laplacian matrix.□

As L and L_d both are Laplacian matrices, therefore, similarly to PCA and LDA, the shift-invariance property is naturally satisfied in Eq. (11).

4 Geometric Meaning of the Shift-Invariant LPP

In this section, we reveal the geometric meaning of the derived shift-invariant LPP algorithm and then, based on the geometric meaning, we propose two feature-extraction algorithms in the next section.

Denote $p_i = \mathrm{D}_{ii}/\Sigma_i\mathrm{D}_{ii}$, $\bar{x} = \Sigma_i p_i x_i$, and $\alpha = \Sigma_i \mathrm{D}_{ii}$. With some algebraic operations, one can easily verify the following two equations:

$$\mathrm{X}\mathrm{L}\mathrm{X}^T = \frac{1}{2}\sum_{i,j} \mathrm{A}_j(x_i - x_j)(x_i - x_j)^T \tag{13}$$

$$\mathrm{X}\mathrm{L}_d\mathrm{X}^T = \alpha\sum_{i=1}^{n} p_i(x_i - \bar{x})(x_i - \bar{x})^T. \tag{14}$$

According to Eqs. (13) and (14), we have

$$\mathrm{tr}(\mathrm{X}\mathrm{L}\mathrm{X}^T) = \frac{1}{2}\sum_{i,j} \mathrm{A}_{ij}\|x_i - x_j\|^2 \tag{15}$$

$$\mathrm{tr}(\mathrm{X}\mathrm{L}_d\mathrm{X}^T) = \alpha\sum_{i=1}^{n} p_i\|x_i - \bar{x}\|^2. \tag{16}$$

Under a projection matrix $\mathrm{W} \in \mathbb{R}^{d\times m}$, data point $x \in \mathbb{R}^d$ is transformed into $y \in \mathbb{R}^m$ by $y = \mathrm{W}^T x$. It is not difficult to verify the following two equations:

$$\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W}) = \frac{1}{2}\sum_{i,j} \mathrm{A}_{ij}\|y_i - y_j\|^2 \tag{17}$$

$$\mathrm{W}^T\mathrm{X}\mathrm{L}_d\mathrm{X}^T\mathrm{W} = \alpha\sum_{i=1}^{n} p_i(y_i - \bar{y})(y_i - \bar{y})^T. \tag{18}$$

Note that Eq. (11) is determined by Eqs. (17) and (18). According to Eqs. (17) and (18), the geometric meaning of the shift-invariant LPP algorithm becomes clear. The algorithm tries to find a projection matrix W, such that under this projection matrix, the sum of the Euclidean distances between data pairs which are local to each other is minimized, while the weighted covariance matrix is fixed to a constant matrix ($\mathrm{W}^T\mathrm{X}\mathrm{L}_d\mathrm{X}^T\mathrm{W} = \mathrm{I}$). In the original LPP, the meaning of $\mathrm{W}^T\mathrm{X}\mathrm{D}\mathrm{X}^T\mathrm{W}$ is not the weighted covariance matrix anymore.

5 Locality-Minimizing Globality-Maximizing Projections

From the previous analysis, we know that LPP is used to minimize $\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W})$, while minimizing $\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W})$ is equivalent to minimizing the distances of data pairs in locality. To gain more discriminative power, it is desirable to minimize the locality and maximize the globality simultaneously.

According to Eq. (18), we have

$$\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}_d\mathrm{X}^T\mathrm{W}) = \alpha\sum_{i=1}^{n} p_i\|y_i - \bar{y}\|^2, \tag{19}$$

so $\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}_d\mathrm{X}^T\mathrm{W})$ is the weighted variance of data, which can be seen as the global information in data. Therefore, it is desirable to minimize $\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W})$ and maximize $\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}_d\mathrm{X}^T\mathrm{W})$ simultaneously. To this end, it is reasonable to select the following two criteria:

$$\mathcal{M}_1(\mathrm{W}) = \mathrm{tr}[\mathrm{W}^T X(\lambda\mathrm{L} - \mathrm{L}_d)\mathrm{X}^T\mathrm{W}] \tag{20}$$

$$\mathcal{M}_2(\mathrm{W}) = \frac{\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}\mathrm{X}^T\mathrm{W})}{\mathrm{tr}(\mathrm{W}^T\mathrm{X}\mathrm{L}_d\mathrm{X}^T\mathrm{W})}, \tag{21}$$

where λ is a user-predefined constant. It is easy to see that the above two criteria both satisfy the shift-invariance property.

In order to obtain an orthogonal projection matrix W, we further add a constraint as $\mathrm{W}^T\mathrm{W} = \mathrm{I}$, then solve the following two optimization problems:

Table 1 Accuracy±dev. (%) on the AT&T database.

	2 Train	4 Train	6 Train
LDA	77.5±2.9	89.9±1.8	91.6±2.0
LPP	77.4±2.7	87.1±1.9	88.6±2.3
SI-LPP	79.1±2.6	88.1±1.9	89.1±2.2
LMGMP1	82.9±2.8	94.2±1.6	96.9±1.5
LMGMP2	84.6±2.6	94.9±1.7	97.2±1.3

Table 2 Accuracy±dev. (%) on the UMIST database.

	4 Train	6 Train	8 Train
LDA	85.0±3.5	90.9±2.4	93.6±2.1
LPP	79.6±3.7	86.3±2.6	90.2±2.4
SI-LPP	81.8±3.8	88.2±2.6	91.5±2.4
LMGMP1	86.5±3.7	93.5±2.1	96.3±1.4
LMGMP2	90.6±2.9	95.4±1.3	97.3±1.2

$$W = \arg \min_{W^TW=I} \operatorname{tr}[W^TX(\lambda L - L_d)X^TW] \tag{22}$$

$$W = \arg \min_{W^TW=I} \frac{\operatorname{tr}(W^TXLX^TW)}{\operatorname{tr}(W^TXL_dX^TW)}. \tag{23}$$

The solution to Eq. (22) can be easily obtained by eigen decomposition of $X(\lambda L - L_d)X^T$ as follows

$$X(\lambda L - L_d)X^TW = W\Lambda \tag{24}$$

where Λ is the eigenvalue matrix of $X(\lambda L - L_d)X^T$ and W is the corresponding eigenvector matrix.

Solving Eq. (23) is a little intractable. Fortunately, the solution to it can also be efficiently obtained by iterative procedure.[14,15]

Solving optimization problems (22) and (23) can provide a general graph-based feature-extraction framework. Different construction of the Laplacian matrices L and L_d leads to different unsupervised, semisupervised, or supervised feature-extraction algorithms, and the corresponding kernelization and tensorization extensions can also be easily derived from this framework.[13]

6 Experimental Results

In this section, we evaluate the performance of the proposed algorithms for face recognition. The algorithms corresponding to optimization problems (22) and (23) are denoted by LMGMP1 and LMGMP2, respectively. In the experiments, λ in Eq. (20) is set to 2. We compare the proposed algorithms with LDA, LPP, and shift-invariant LPP (SI-LPP).

LMGMP1, LMGMP2, LPP, and SI-LPP construct the same graph structure based on the label information. PCA is used as a preprocessing step before performing the algorithms.

In each experiment, we randomly select several samples per class for training, and the remaining samples are used for testing. The classification is based on 1-nearest neighbor classifier. The average accuracy rates and the standard deviations are recorded over 50 random splits.

6.1 *AT&T Database*

The AT&T face database includes 40 distinct individuals, and each individual has 10 different images. Each image is down-sampled to the size of 28×23. The training number *t* per class is 2, 4, and 6, respectively. The best results of each algorithm are reported in Table 1. For *t*=4, the results of accuracy rate versus dimension are shown in Fig. 1.

As can be seen in Table 1 and Fig. 1, the performances of the proposed algorithms are much better than those of LDA and LPP, especially when the reduced dimension is very low. SI-LPP also gains an improvement over LPP.

6.2 *UMIST Database*

The UMIST repository is a multiview database consisting of 575 images of 20 people, each covering a wide range of poses from profile to frontal views. We down-sample the size of each image to 28×23. The training number *t* per class is 4, 6, and 8, respectively. The best results of each

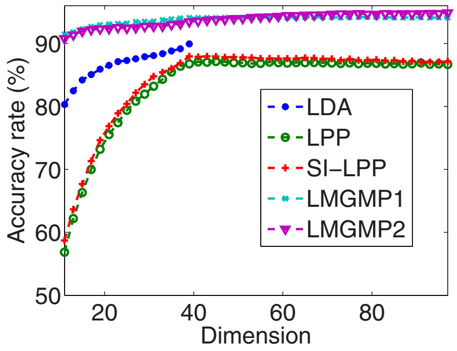


Fig. 1 Accuracy versus dimensions on the AT&T database (4 train).

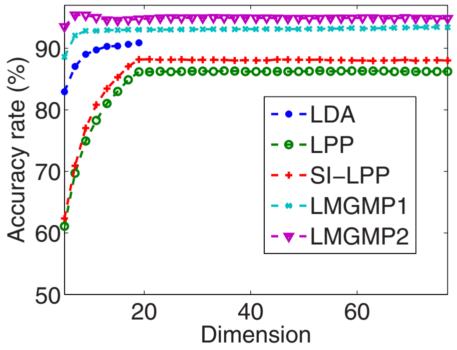


Fig. 2 Accuracy versus dimensions on the UMIST database (6 train).

algorithm are reported in Table 2. For $t=6$, the results of accuracy rate versus dimension are shown in Fig. 2.

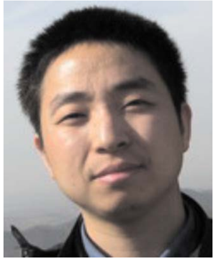
Similarly to the above experiments, the performances of the proposed algorithms are significantly better than those of LDA and LPP, especially when the reduced dimension is very low. SI-LPP also gains an improvement over LPP.

7 Conclusion

In this paper, we point out that the LPP algorithm, which was developed recently and has been frequently used in many applications, does not satisfy the shift-invariance property. The reason is analyzed, and the shift-invariant LPP algorithm is derived. We also reveal the geometrical meaning of the derived shift-invariant LPP. On the basis of the analysis of the geometrical meaning, we propose two novel algorithms to minimize the locality and maximize the globality simultaneously under an orthogonal projection matrix. Several recently proposed works related to our algorithms are discussed. Experimental results on face recognition demonstrate the effectiveness and superiority of the proposed algorithms.

References

1. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, Hoboken, NJ (2000).
2. X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, vol. 16, pp. 153–160, MIT Press, Cambridge, MA (2003).
3. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.* **15**(6), 1373–1396 (2003).
4. J. Yang, D. Zhang, J. Y. Yang, and B. Niu, "Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 650–664 (2007).
5. J. Yan, N. Liu, B. Zhang, S. Yan, Z. Chen, Q. Cheng, W. Fan, and W.-Y. Ma, "OCFS: optimal orthogonal centroid feature selection for text categorization," in *Proc. 28th Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 122–129, ACM, New York (2005).
6. J. Ye, "Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems," *J. Mach. Learn. Res.* **6**, 483–502 (2005).
7. C. H. Q. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix tri-factorizations for clustering," in *Int. Conf. on Knowledge Discovery and Data Mining*, pp. 126–135, ACM, New York (2006).
8. X. Liu, J. Yin, Z. Feng, J. Dong, and L. Wang, "Orthogonal neighborhood preserving embedding for face recognition," in *Intl. Conf. on Image Processing*, pp. I-133–I-136, IEEE, Piscataway, NJ (2007).
9. H. Wang, S. Chen, Z. Hu, and W. Zheng, "Locality-preserved maximum information projection," *IEEE Trans. Neural Netw.* **19**(4), 571–585 (2008).
10. E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(12), 2143–2156 (2007).
11. D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Trans. Image Process.* **15**(11), 3608–3614 (2006).
12. F. R. K. Chung, "Spectral graph theory," presented at *CBMS Regional Conf. Series in Mathematics*, vol. 92, American Mathematical Society (Feb. 1997).
13. S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(1), 40–51 (2007).
14. F. Nie, S. Xiang, and C. Zhang, "Neighborhood minmax projections," *Intl. Joint Conf. on Artificial Intelligence*, pp. 993–998 (2007).
15. H. Wang, S. Yan, D. Xu, X. Tang, and T. S. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ (2007).



Feiping Nie received his BS in computer science from North China University of Water Conservancy and Electric Power, China in 2000 and MS in computer science from Lanzhou University, China in 2003. He is currently a PhD candidate in the Department of Automation, Tsinghua University, China. His research interests include machine learning, pattern recognition, data mining, and image processing.

Shiming Xiang received his BS from the Department of Mathematics of Chongqing Normal University, China in 1993, MS from the Department of Mechanics and Mathematics of Chongqing University, China in 1996, and PhD from the Institute of Computing Technology, Chinese Academy of Sciences, China in 2004. He held a postdoctoral position in the Department of Automation, Tsinghua University until 2006. He is currently an associate researcher at the Institute of Automation Chinese Academy of Science. His interests include computer vision, pattern recognition, and machine learning.

Yangqiu Song received his BS from the Department of Automation, Tsinghua University, China in 2003. He is currently a PhD candidate in the Department of Automation, Tsinghua University. His research interests focus on machine learning and its applications.

Changshui Zhang received his BS in Mathematics from Peking University, China in 1986 and PhD from the Department of Automation, Tsinghua University in 1992. He is currently a professor of the Department of Automation, Tsinghua University. He is an associate editor of the journal *Pattern Recognition*. His interests include artificial intelligence, image processing, pattern recognition, machine learning, evolutionary computation, and complex system analysis.