This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Low-Rank Preserving Projections

Yuwu Lu, Zhihui Lai, Yong Xu, *Senior Member, IEEE*, Xuelong Li, *Fellow, IEEE*, David Zhang, *Fellow, IEEE*, and Chun Yuan

***Abstract*—As one of the most popular dimensionality reduction techniques, locality preserving projections (LPP) has been widely used in computer vision and pattern recognition. However, in practical applications, data is always corrupted by noises. For the corrupted data, samples from the same class may not be distributed in the nearest area, thus LPP may lose its effectiveness. In this paper, it is assumed that data is grossly corrupted and the noise matrix is sparse. Based on these assumptions, we propose a novel dimensionality reduction method, named low-rank preserving projections (LRPP) for image classification. LRPP learns a low-rank weight matrix by projecting the data on a low-dimensional subspace. We use the L_{21} norm as a sparse constraint on the noise matrix and the nuclear norm as a low-rank constraint on the weight matrix. LRPP keeps the global structure of the data during the dimensionality reduction procedure and the learned low rank weight matrix can reduce the disturbance of noises in the data. LRPP can learn a robust subspace from the corrupted data. To verify the performance of LRPP in image dimensionality reduction and classification, we compare LRPP with the state-of-the-art dimensionality reduction methods. The experimental results show the effectiveness and the feasibility of the proposed method with encouraging results.**

***Index Terms*—Face recognition, image classification, locality preserving projections (LPP), low-rank representation (LRR).**

Manuscript received January 27, 2015; revised May 22, 2015; accepted July 6, 2015. This work was supported in part by the Natural Science Foundation of China under Grant 61203376, Grant 61300032, Grant 61375012, Grant 61362031, Grant 61170253, and Grant 61370163, in part by the National Significant Science and Technology Projects of China under Grant 2013ZX01039001-002-003, and in part by the Shenzhen Municipal Science and Technology Innovation Council under Grant JCYJ20130329151843309 and Grant JCYJ20140904154630436. This paper was recommended by Associate Editor X. He. *(Yuwu Lu and Zhihui Lai contributed equally to this work.) (Corresponding author: Yong Xu.)*

Y. Lu is with the Tsinghua-CUHK Joint Research Center for Media Sciences, Technologies and Systems, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China, and also with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: luyuwu2008@163.com).

Z. Lai is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518055, China, and also with the Institute of Textiles & Clothing, Hong Kong Polytechnic University, Hong Kong (e-mail: lai_zhi_hui@163.com).

Y. Xu is with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: yongxu@ymail.com).

X. Li is with the Center for OPTical IMagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xuelong_li@opt.ac.cn).

D. Zhang is with Biometrics Research Center, Hong Kong Polytechnic University, Hong Kong (e-mail: csdzhang@comp.polyu.edu.hk).

C. Yuan is with the Tsinghua-CUHK Joint Research Center for Media Sciences, Technologies and Systems, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China (e-mail: yuanc@sz.tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2015.2457611

I. Introduction

In MANY real applications, such as machine learning, data mining, image processing, and pattern recognition, the original data is always very high dimensional. The high-dimensional data needs high memory requirements and is computationally expensive. This problem is the so-called "curse of dimensionality" [1]. To solve this problem, many methods are proposed for dimensionality reduction. Principal component analysis (PCA) [2], [3] aims at preserving the global variance with the minimum reconstruction error by projecting data on a linear subspace spanned by principal component vectors. Linear discriminant analysis (LDA) [4], [5] searches the projection axes on which data from the same class are as close as to each other and requires data points from different classes are as far as possible. LDA has the over-fitting problem. A novel algorithm is proposed to overcome this problem without increasing the number of training samples [6]. However, for the undersample problem, LDA is not suited to traditional image representation. In [7], a novel method named general tensor discriminant analysis is proposed as a preprocessing step for LDA. Tao *et al.* [8] studied the geometric mean for subspace selection and analyze three criteria which can reduce the class separation problem.

However, linear dimensionality reduction may fail to discover essential data structure that is nonlinear. Thus, manifold learning algorithms are proposed to uncover the hidden semantics and simultaneously preserve the intrinsic geometric structure of the data. Among the manifold learning methods, locally linear embedding (LLE) [9], ISOMAP [10], and Laplacian eigenmaps (LE) [11] are the most popular manifold learning techniques. LLE not only represents the local geometry by linear coefficients and reconstructs a given sample by its neighbors but also seeks a low-dimensional embedding which is suitable for reconstruction [12]. However, all of these nonlinear methods suffer from the out-of-sample problem [13]. The simplest and frequently used technique is to learn the explicit linear mappings of the corresponding nonlinear manifold learning methods [14]. Locality preserving projections (LPP) [15], [16], a linearization of LE, neighborhood preserving embedding (NPE) [17], neighborhood preserving projections (NPP) [18], [19], and the linearization of LLE are proposed to address the out-of-sample problem. LPP constructs a certain affinity graph by the data

2168-2267 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

and preserves the local geometry of the original data. In order to improve the performance in ranking applications of LPP, a novel method named semi-supervised LPP (SSLPP) is proposed [20]. SSLPP introduces the relevance degree information to the affinity graph in LPP. NPE also aims at keeping the local neighborhood structure of the data as LPP. However, NPE uses local squares approximation to construct the affinity weight matrix. NPP introduces a linear transform matrix to approximate LLE. Thus, NPP has the neighborhood property similar to LLE. NPP utilizes local neighborhood relations to learn the global structure. Kokiopoulou and Saad [21] pointed out that adding the orthogonality constraint on the projection directions is more effective for preserving the intrinsic geometrical structure of the data. Thus, orthogonal LPP (OLPP) [22] adds the orthogonality constraint to produce orthogonal basis functions and has more locality preserving power and discriminating power than LPP. Liu *et al.* [23] proposed orthogonal NPE (ONPE) which also requires the basis functions to be orthogonal. Sparsity preserving projections (SPP) [24] preserves the sparse reconstructive relationship of the data by minimizing a L_1 regularization term which different from many existing dimensionality reduction methods, such as LPP and NPE. To introduce ranking information into dimensionality reduction, a novel method named ranking graph embedding (RANGE) is proposed in [25]. RANGE models the global structure and the local relationships among different relevance degree sets. Pang *et al.* [26] proposed a novel feature description framework named distributed object detection. To extract robust features for human detection, a novel framework is proposed in [27]. Geng *et al.* [28] proposed a novel metric learning method named domain adaptation metric learning for domain adaptation settings. In practical applications, multiview learning can obtain more information than an individual view, thus many related works are proposed, such as large-margin multiview information bottleneck [29], multiview stochastic neighbor embedding [30], and multiview intact space learning [31].

However, in real applications, most of the aforementioned methods preserve the local neighborhood information and ignore the global structure of the data. Local information of the data is easily effected by illumination, corruptions, or noises. Thus, the recognition rate of these methods may degrade in clustering or classifying tasks with noises or corruptions. Fortunately, the recently proposed low-rank representation (LRR) methods have gain attentions for its robustness on the noise/corrupted data. In the past several years, many LRR methods were proposed for robust classification tasks [32]–[35]. Robust PCA (RPCA) introduces the nuclear norm to recover the subspace structure from the data corrupted by noises or occlusions [32], [33]. Liu *et al.* [34] proposed LRR to recover the lowest rank representation of the data. LRR segments the data drawn from subspaces and can better captures the global structure of the data. Zhuang *et al.* [35] proposed non-negative low rank and sparse graph (NNLRS) for semi-supervised learning. NNLRS constructs an informative graph by introducing sparsity and low rankness of high-dimensional data. NNLRS-graph also can better capture the global structure of the data as LRR.

Conceptually, manifold learning methods construct the graph by local patches and rely on pair-wise Euclidean distances. Thus, these methods are sensitive to noises or errors in data [35]. Therefore, how to learn a robust representative coefficient matrix for feature extraction is a key problem. Fortunately, LRR can capture the global structure of data and is robust to noises and outliers. To effective improve the robustness of the preserving projection-based methods, we propose to harness local preserving projections, sparsity, and low rankness of high-dimensional data to build a graph. Therefore, in this paper, we propose a novel dimensionality reduction method, named low-rank preserving projections (LRPP) for image classification. First, we assume that data is grossly corrupted and the noise matrix is sparse. The L_{21} norm is used as a sparse constraint on the noise matrix and the nuclear norm as a low-rank constraint on the weight matrix. Then, LRPP learned a low-rank weight matrix which projecting the data on a low-dimensional subspace. LRPP keeps the global structure of the data during the dimensionality reduction procedure and the learned low rank weight matrix can lower the disturbance of noises in the data. To verify the performance of LRPP in image dimensionality reduction and classification, we compare LRPP with the state-of-the-art dimensionality reduction techniques on six public image databases. The experimental results show the effectiveness and the feasibility of the proposed method with encouraging results.

The main contributions of this paper are as follows.

1) This paper integrates the graph learning and the projection learning into a seamless model. Different from conventional locality or globality preserving projection methods, LRPP can learn a novel weight graph which not only obtains a balance between globality and locality of the data but also obtains strong robustness to the noisy data by using low rank learning.
2) LRPP shares some advantages of both LPP and other dimensionality reductions methods. LRPP can learn projection vectors which transform the data to a new subspace, which is robust to noises.
3) The technique proposed in this paper can be easily extended to supervised and semi-supervised scenarios. According to [36]–[39], the proposed method can also be extended to other scenarios.

The rest of this paper organized as follows. For the ease of reading, we review related works in Section II. The details of the proposed method are introduced in Section III. Section IV introduces the convergence and the computational complexity of LRPP and its connection with some related works. Section V shows the experiments results. Finally, we conclude this paper in Section VI.

For convenience, we present in Table I the important notations used in the rest of this paper.

II. Reviews of Related Works

In order to the ease of reading, in this section, we briefly review some related works including LPP, NPE, SPP, LRR, and RPCA.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
IMPORTANT NOTATIONS USED IN THIS PAPER

m, n	data dimensionality, number of the training points
H_{ij}	weight coefficient
X	data matrix of size $m \times n$
E	additive error matrix
A	base matrix
W	coefficient matrix
x_i, x_j	data points
y_i, y_j	low-dimensinal representations of x_i and x_j
p	projective vector
i, j	$i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$

A. LPP

Similar to PCA, LPP is also an unsupervised dimensionality reduction approach. However, PCA preserves the global structure of the data whereas LPP keeps the local structure of the data. The objective function of LPP is defined as follows:

$$\frac{1}{2}\sum_{ij}^{n} H_{ij}\left\|y_i - y_j\right\|_2^2 \tag{1}$$

where $y_i = p^T x_i$ and $y_j = p^T x_j$, $i = 1, 2, \ldots, n$, T represents the transpose of a vector. H_{ij} is a weight coefficient defined as follows:

$$H_{ij} = \begin{cases} \exp\left(-\left\|x_i - x_j\right\|^2 \Big/ t\right) \\ 0. \end{cases} \tag{2}$$

The objective function defined in (1) can be rewritten as

$$\begin{aligned}\frac{1}{2}\sum_{ij}^{n} H_{ij}\left\|y_i - y_j\right\|_2^2 &= \frac{1}{2}\sum_{ij}^{n} H_{ij}\left\|p^T x_i - p^T x_j\right\|_2^2 \\ &= p^T X(D-H)X^T p = p^T X L X^T p \end{aligned} \tag{3}$$

where D is a diagonal matrix with its entries being the row sums of H, i.e., $d_{ii} = \sum_j H_{ij}$, $L = D - H$.

B. NPE

Similar to LPP, NPE is also to keep the local neighborhood structure of the data. NPE measured the local approximation error by minimizing the function [17]

$$\phi(N) = \sum_i \left\| x_i - \sum_j N_{ij} x_j \right\|^2 \tag{4}$$

where x_j are k neighbors of x_i. According to [17], a reasonable criterion for choosing good projection which minimizing the cost function is

$$\phi(p) = \sum_i \left(p^T x_i - \sum_j \tilde{N}_{ij} p^T x_j \right)^2 \tag{5}$$

where $\tilde{N}_{ij}$ is the optimal solution of (4). To minimize (5), it can be lead to

$$\min_p \frac{p^T X \tilde{M} X^T p}{p^T X X^T p} \tag{6}$$

where $\tilde{M} = (I - N)^T (I - N)$.

C. SPP

LPP and NPE keep local neighborhood information during the dimensionality reduction procedure. To preserve the sparse reconstructive relationship of the data, Qiao *et al.* [24] proposed SPP which used the L_1 regularization to minimize the following objective function:

$$\max_p \frac{p^T X S_\beta X^T p}{p^T X X^T p} \tag{7}$$

where $S_\beta = S + S^T - S^T S \cdot S$ is constructed as

$$\min_{s_i} \|s_i\|_1, \ \text{s.t. } x_i = X s_i, 1 = I_1^T s_i \tag{8}$$

where $s_i = [s_{i1}, \ldots, s_{i,i-1}, 0, s_{i,i+1}, \ldots, s_{in}]^T$ is an n-dimensional vector in which the ith element is equal to zero. $I_1 \in R^n$ is a vector of all ones.

For more details about SPP, the readers are referred to [24].

D. LRR

As a subspace clustering method, the basic idea of LRR is to capture the lowest rank representation in the combination of the bases in the given dataset. This problem can be formulated as

$$\min_W \|W\|_*, \ \text{s.t. } X = AW \tag{9}$$

where $A = [a_1, \ldots, a_d]$ is the bases matrix, $\|\cdot\|_*$ represents the nuclear norm of a matrix. In real applications, we often choose the data matrix as the base matrix, thus (9) can be rewritten as

$$\min_W \|W\|_*, \ \text{s.t. } X = XW. \tag{10}$$

With the noise in the data, the matrix always divided into two parts

$$\min_W \|W\|_* + \lambda \|E\|_p, \ \text{s.t. } X = XW + E \tag{11}$$

where $E \in R^{m \times n}$ is a sparse additive error matrix. $\|\cdot\|_p$ denotes certain norm regularization, $\|E\|_p$ can be written in $\|E\|_F^2$, $\|E\|_1$, and $|E\|_{21}$ in different cases.

E. RPCA

PCA has been used in many fields as one of the most important dimensionality reduction techniques. However, PCA is sensitive to outliers or gross corruptions of the data. To overcome these shortcomings of PCA, RPCA [32], [33] is proposed to recover the subspace structure from the original data including gross errors [40], [41].

RPCA assumed that the observed data matrix can be decomposed as $X = \tilde{X} + E \in R^{m \times n}$, where $\tilde{X} \in R^{m \times n}$ is an original

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

low rank data matrix and $E \in R^{m\times n}$ is a sparse additive error matrix. The objective function of RPCA is defined as follows:

$$\min_{\tilde{X},E} \operatorname{rank}(\tilde{X}) + \alpha\|E\|_0, \text{ s.t. } X = \tilde{X} + E \tag{12}$$

where α is a positive parameter, and $\|\cdot\|_0$ is the L_0 norm of a matrix. Due to the optimization problem (12) is non-convex and no efficient solution can obtained, (12) is usually transformed to the following convex problem by relaxing the rank function and the L_0 norm into the nuclear norm and the L_1 norm, respectively:

$$\min_{\tilde{X},E} \|\tilde{X}\|_* + \alpha\|E\|_1, \text{ s.t. } X = \tilde{X} + E \tag{13}$$

where $\|\cdot\|_*$ is the nuclear norm, $\|\cdot\|_1$ denotes the L_1 norm of a matrix and $\alpha > 0$ is a parameter.

III. Low-Rank Preserving Projections

In this section, we introduce the details of the proposed method, i.e., LRPP. We first present the motivation of LRPP, and then describe the objective function of LRPP and its solutions.

A. Motivations

LPP constructs an affinity graph to incorporate neighborhood information of the data and transform data points into a new space and ensure that points are in close proximity in the original space remained as in the new space. Similar to LPP, NPE also builds a graph to preserve the local structure of the data. Adding the orthogonal constraint between the projection directions can improve the effectiveness for preserving the intrinsic structure of the data [21]. ONPP and OLPP encode the orthogonal information and effectively improve the classification performance. However, these unsupervised dimensionality reduction methods mainly rely on pair-wise Euclidean distances, they are very sensitive to noises and errors in data [35]. These methods only capture the local structure of the data and ignore the global structure. For corrupted samples which belong to the same class may distribute far away from the nearest area. Under this circumstance, the recognition rate of LPP, NPE, and their modifications will be greatly degraded. Therefore, how to deal with this problem is of vital importance.

Fortunately, recent research indicates that LRR can capture the global structure of the data and the low-rank properties have great robustness to corruptions and even can recover the true data [34]. Therefore, these low-rank properties can be integrated into dimensionality reduction to address the sensitivity of the locality-based or neighborhood-based dimensionality reduction methods such as LPP, NPP, and their modifications. The key idea is to effectively utilize the low-rankness of high-dimensional data to build an informative graph and capture the global structure of the noisy data for dimensionality reduction. Thus, in this paper, we propose a novel dimensionality reduction method, named LRPP for image feature extraction and classification. Assuming the data is grossly corrupted, we use the L_{21} norm as a sparse constraint on the noise matrix and the nuclear norm as a low-rank constraint on the weight matrix. LRPP learns a low rank weight matrix to capture the globality of the data, which is preserved when the data is projected on a low-dimensional subspace. LRPP effective captures the global information of the data and also keeps the low rank information of the data during the dimensionality reduction procedure.

B. Objective Function of LRPP

The conventional LPP and its modifications methods mainly use the pair-wise Euclidean distances to capture the locality of the data. These methods are very sensitive to noises and errors in the data. However, LRR can better capture the global subspace structure of the data and is more robust to noises and outliers [34], [35]. To improve the robustness of LPP, we use low rankness of the data to construct an affinity graph, with the assumption that noises of the data are sparse.

Let $X = [x_1, \ldots, x_n] \in R^{m\times n}$ be a data matrix in which each column represents a sample. Each sample of X can be represented by a linear combination of basis matrix $A = [a_1, \ldots, a_d]$

$$X = AW \tag{14}$$

where $W = [w_1, \ldots, w_n]$ is the coefficient matrix. However, in real applications, data is usually corrupted by noises or outliers. We suppose noise matrix E is sparse. Thus, (14) can be reformulated as

$$X = AW + E. \tag{15}$$

In order to preserve the low rankness of the dataset for feature extraction so as to uncover the true representative relationship of the data and be robust to noises, we propose the following rank and L_{21} minimization problem:

$$\begin{aligned} \min_{p,W,E} \ & \frac{1}{2}\sum_{i,j}^{n}(W_{ij} + W_{ji})\|p^T x_i - p^T x_j\|_2^2 \\ & + \alpha\operatorname{rank}(W) + \beta\|E\|_{21} \\ \text{s.t. } & X = AW + E \end{aligned} \tag{16}$$

where p is a projection vector which transform data into a new subspace, and α, β are positive parameters, and $\|E\|_{21} = \sum_{j=1}^{m}\sqrt{\sum_{i=1}^{n}(E_{ij})^2}$.

The first term in model (16) (i.e., $\sum_{i,j}^{n}(W_{ij} + W_{ji})\|p^T x_i - p^T x_j\|_2^2$) is used to learn a novel weight matrix and project data on a new subspace. In this process, noises in the data are weakened to ensure samples of the same class distribute in the same area. The second term of (16) [i.e., $\alpha\operatorname{rank}(W)$] ensures that the learned weight matrix has a low rank. In this way, the global structure of the data is well preserved. The last term of (16) is used to ensure that the noise matrix is sparse. The proposed model (16) is totally different from LPP, NPE, and their modifications which only considering the local information of the data. LRPP effectively encodes the global structure of the data and the learned weight matrix enhances the separate ability. The problem (16) is NP-hard and not easy to solve, so we relax $\operatorname{rank}(W)$ into its nuclear norm $\|W\|_*$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

(i.e., the sum of singular values of W) [33]. The problem (16) can be rewritten as [34]

$$\min_{p,W,E} \ \frac{1}{2}\sum_{i,j}^{n}\left(W_{ij}+W_{ji}\right)\left\|p^T x_i - p^T x_j\right\|_2^2 + \alpha\|W\|_* + \beta\|E\|_{21} \quad \text{s.t.} \ \ X = AW + E. \tag{17}$$

Suppose there is a linear mapping such that $y_i = p^T x_i$ and $y_j = p^T x_j$, then (17) can be formulated as

$$\min_{p,W,E} \ \frac{1}{2}\sum_{i,j}^{n}\left(W_{ij}+W_{ji}\right)\left\|y_i - y_j\right\|_2^2 + \alpha\|W\|_* + \beta\|E\|_{21} \quad \text{s.t.} \ \ X = AW + E. \tag{18}$$

In the following section, we will show how to solve (18) by using the linearized alternating direction method with adaptive penalty (LADMAP) [42].

C. Optimization of LRPP

We use the LADMAP [42] to solve (18). In the model of LRPP, we choose A as X itself. We first introduce an auxiliary variable J in order to make the objective function separable

$$\min_{J,W,E,p} \ \frac{1}{2}\sum_{i,j}^{n}\left(J_{ij}+J_{ji}\right)\left\|y_i - y_j\right\|_2^2 + \alpha\|W\|_* + \beta\|E\|_{21} \quad \text{s.t.} \ \ X = XW + E, \ \ W = J. \tag{19}$$

The augmented Lagrangian function of problem (19) is

$$\begin{aligned} &L(W, J, E, p, M_1, M_2, \mu) \\ &\quad = \frac{1}{2}\sum_{i,j}^{n}\left(J_{ij}+J_{ji}\right)\left\|y_i - y_j\right\|_2^2 + \alpha\|W\|_* + \beta\|E\|_{21} \\ &\qquad + \langle M_1, X - XW - E\rangle + \langle M_2, W - J\rangle \\ &\qquad + \frac{\mu}{2}\left(\|X - XW - E\|_F^2 + \|W - J\|_F^2\right) \\ &\quad = \frac{1}{2}\sum_{i,j}^{n}\left(J_{ij}+J_{ji}\right)\left\|y_i - y_j\right\|_2^2 + \alpha\|W\|_* + \beta\|E\|_{21} \\ &\qquad + \frac{\mu}{2}\left(\left\|X - XW - E + \frac{M_1}{\mu}\right\|_F^2 + \left\|W - J + \frac{M_2}{\mu}\right\|_F^2\right) \\ &\qquad - \frac{1}{2\mu}\left(\|M_1\|_F^2 + \|M_2\|_F^2\right) \end{aligned} \tag{20}$$

where M_1 and M_2 are Lagrange multipliers, $\mu > 0$ is a penalty parameter, and $\|\cdot\|_F$ denotes the Frobenious norm of a matrix. The problem (20) is unconstrained and it can be minimized with respect to W, J, E, p by fixing other variables, respectively.

In the next sections, we give the updating rules of each variable.

D. Computation of W

We show that the updating rule about W that solves the optimization problem in (20) can be expressed in terms of J and E.

Given J, E, and p, the terms in (20) that depend on W are

$$\alpha\|W\|_* + \frac{\mu}{2}\left\|X - XW - E + \frac{M_1}{\mu}\right\|_F^2 + \frac{\mu}{2}\left\|W - J + \frac{M_2}{\mu}\right\|_F^2. \tag{21}$$

Then, the updating rule about W in (20) can be rewritten as

$$\arg\min_{W} \ \alpha\|W\|_* + \frac{\mu}{2}\left\|X - XW - E + \frac{M_1}{\mu}\right\|_F^2 + \frac{\mu}{2}\left\|W - J + \frac{M_2}{\mu}\right\|_F^2 \tag{22}$$

which can be solved by LADMAP [42] and singular value thresholding (SVT) operator [43].

E. Computation of J

When computing J, we can rewrite (20) as

$$\min_{J} \frac{1}{2}\sum_{i,j}^{n}\left(J_{ij}+J_{ji}\right)\left\|y_i - y_j\right\|_2^2 + \frac{\mu}{2}\left\|W - J + \frac{M_2}{\mu}\right\|_F^2. \tag{23}$$

To obtain the updating rule for J, we first fix W and E. Let $B_{ij} = \|y_i - y_j\|_2^2$ and $M = W + \frac{M_2}{\mu}$, (23) can be rewritten as

$$\min_{J} \frac{1}{2}\sum_{i,j}^{n}\left(J_{ij}+J_{ji}\right)B_{ij} + \frac{\mu}{2}\|M - J\|_F^2. \tag{24}$$

The model (24) can be rewritten as

$$\begin{aligned} &\frac{1}{2}\sum_{i,j}^{n}\left(J_{ij}+J_{ji}\right)B_{ij} + \frac{\mu}{2}\|M - J\|_F^2 \\ &\quad = \frac{1}{2}\mathrm{Tr}\left(I_1\left(\left(J + J^T\right)\odot B\right)\right) + \frac{\mu}{2}\|M - J\|_F^2 \\ &\quad = \frac{1}{2}\mathrm{Tr}(I_1(J\odot B)) + \frac{1}{2}\mathrm{Tr}\left(I_1\left(J^T\odot B\right)\right) + \frac{\mu}{2}\mathrm{Tr}\|M - J\|_F^2 \\ &\quad = \frac{1}{2}\sum_i B_{i:}J_{:i} + \frac{1}{2}\sum_i B_{i:}J_{:i}^T + \frac{\mu}{2}\sum_i (M_{i:} - IJ_{i:})_2^2 \\ &\quad = \frac{1}{2}\sum_i B_{i:}J_{:i} + \frac{\mu}{4}\sum_i (M_{i:} - IJ_{:i})_2^2 + \frac{1}{2}\sum_i B_{i:}J_{:i}^T \\ &\qquad + \frac{\mu}{4}\sum_i \left(M_{i:} - IJ_{:i}^T\right)_2^2 \end{aligned} \tag{25}$$

where $B_{i:}$ represents the ith row of matrix B, $B_{:i}$ represents the ith column of matrix B, I_1 represents the matrix whose elements are all 1, $\odot$ is a Hadamard product operator of matrices.

For the ease of reading, let $B_1 = B_{i:}$, $M_1 = M_{i:}$, $J_1 = J_{i:}$, and $J_2 = J_{:i}^T$. Then, we iteratively solve the following minimization problems across different classes:

$$\sum_i \left(\min_{J_1} \frac{1}{2}B_1 J_1 + \frac{\mu}{4}(M_1 - IJ_1)_2^2\right) \tag{26}$$

and

$$\sum_i \left(\min_{J_2} \frac{1}{2}B_1 J_2 + \frac{\mu}{4}(M_1 - IJ_2)_2^2\right). \tag{27}$$

Both problems (26) and (27) are convex and smooth, they have analytic solutions. We can obtain the optimal solution

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

J_1^* of (26) and the optimal solution J_2^* of (27) by solving the derivatives with respect to J_1 and J_2, respectively. We define the weight matrix J as

$$J = \left(J_1^* + J_2^*\right)/2.$$

F. Computation of E

Given W, p, and J, the terms of E in (20) are

$$\min_E \beta\|E\|_{21} + \frac{\mu}{2}\left\|X - XW - E + \frac{M_1}{\mu}\right\|_F^2. \tag{28}$$

Let $V = X - XW + \frac{M_1}{\mu}$, (28) can be rewritten as

$$\min_E \beta\|E\|_{21} + \frac{\mu}{2}\|V - E\|_F^2. \tag{29}$$

According to [44], we can solve (29) via the following lemma.

Lemma 1: Let V be a given matrix. If the optimal solution to $\min_W \alpha\|E\|_{21} + (1/2)\|E - V\|_F^2$ is E^*, then the ith column of E^* is

$$\left[E^*\right]_{:,i} = \begin{cases} \dfrac{\left\|[V]_{:,i}\right\|_2 - \alpha}{\left\|[V]_{:,i}\right\|_2} V_{:,i} & \text{if } \left\|[V]_{:,i}\right\|_2 > \alpha \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

G. Computation of p

For the given W, J, and E, after some simplifications and elimination, (20) can be formulated as

$$\min_p \frac{1}{2}\sum_{i,j}^{n}\left(J_{ij} + J_{ji}\right)\left\|y_i - y_j\right\|_2^2. \tag{31}$$

Let $H = (1/2)(J + J^T)$, (31) can be reformulated as

$$\min_p \sum_{i,j}^{n} H_{ij}\left\|y_i - y_j\right\|_2^2. \tag{32}$$

Model (32) can be rewritten as

$$\begin{aligned} \sum_{i,j}^{n} H_{ij}\left\|y_i - y_j\right\|_2^2 &= \sum_{i,j}^{n} H_{ij}\left\|p^T x_i - p^T x_j\right\|_2^2 \\ &= 2\left(\sum_{i,j}^{n} p^T x_i H_{ij} x_i^T p - \sum_{i,j}^{n} p^T x_i H_{ij} x_j^T p\right) \\ &= p^T X(D - H)X^T p \\ &= p^T XLX^T p \end{aligned} \tag{33}$$

where $D_{ii} = \sum_j H_{ij}$, $L = D - H$. The matrix D provides a natural measure on the data points and the importance of y_i is corresponding to the value of D_{ii} [15]. Therefore, we impose a constraint as follows:

$$y^T Dy = 1 \Rightarrow p^T XDX^T p = 1. \tag{34}$$

Then, the minimization problem (33) reduces to finding

$$\arg\min_{p^T XDX^T p = 1} p^T XLX^T p. \tag{35}$$

To solve the optimal solution of (35) is identical to solve the following generalized eigenvalue problem:

$$XLX^T p = \lambda XDX^T p. \tag{36}$$

Algorithm 1 LRPP

Input: Training set X; and parameter α, β in (16);
Initialization: $W = 0, J = 0, E = 0, M_1 = 0, M_2 = 0$, $\mu > 0,\ \rho > 0,\ \lambda_1 > 0,$ and $\lambda_2 > 0$.
repeat

1. Update W by (22):

$$\arg\min_W \alpha\|W\|_* + \frac{\mu}{2}\left\|X - XW - E + \frac{M_1}{\mu}\right\|_F^2 + \frac{\mu}{2}\left\|W - J + \frac{M_2}{\mu}\right\|_F^2$$

2. Update J by (26) and (27):

$$\sum_i\left(\min_{J_1} \frac{1}{2}B_1 J_1 + \frac{\mu}{4}(M_1 - IJ_1)_2^2\right)$$
$$\sum_i\left(\min_{J_2} \frac{1}{2}B_1 J_2 + \frac{\mu}{4}(M_1 - IJ_2)_2^2\right)$$

3. Update E by (29)

$$\min_E \beta\|E\|_{21} + \frac{\mu}{2}\|V - E\|_F^2$$

4. Update p by (36)

$$XLX^T p = \lambda XDX^T p$$

5. Update Lagrange multipliers as follows:

$$M_1 = M_1 + \mu(X - XW - E);$$
$$M_2 = M_2 + \mu(W - J).$$

6. Update μ by $\mu = \min(\rho\mu, \max\mu)$.
7. Update $t = t + 1$
8. Obtain the optimal solution (W, J, E, p)

Output: The projection vector p.

The definitions of L and D in our method are different of LPP. The algorithm steps of LRPP are outlined in Algorithm 1.

IV. Analysis and Comparison

In this section, we analyze the convergence and the computational complexity of LRPP and its connection with some related methods.

A. Convergence Analysis

The convergence of the exact LADMAP algorithm has been generally proven in [42]. There are four blocks (including p, W, J, and E) in Algorithm 1 and the objective function (19) is not smooth, it would be not easy to prove the convergence in theory. Step 1 in Algorithm 1 is obtained by SVT operator, its convergence has been proven in [43]. To derive the solution of (25), we iteratively solve (26) across different classes i. During each iteration of minimizing (26), the variable to be minimized is $J_{:i}$, while the remaining variables are fixed. For each iteration, (26) and (27) are convex, so they have analytic solutions and convergent. The convergence of step 3 is also proven in [44]. With four variables,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

the optimization problem (25) is convergence, and thus a local (or even a global) minimizer can be expected.

B. Computational Complexity

In this section, we give a computational analysis of Algorithm 1. The major computation of LRPP algorithm is in steps 1 and 4. Step 1 requires computing the singular value decomposition [45] of a $n \times n$ matrix. Step 4 requires solving a generalized eigenvalue problem. Thus, we just give the complexity about these two steps in LRPP. We use $A = X$ in this paper, the computational complexity of step 1 is at most $O(m^2n)$. The complexity of step 4 is $O(m^3)$. Considering the number of iterations, the complexity of Algorithm 1 is $O(t(m^3+m^2n))$, where t is the number of iterations. Therefore, assuming that $m \geq n$, the upper bound of the complexity of Algorithm 1 is $O(tm^3)$.

C. Comparison With Related Works

As a popular dimensionality reduction method, LPP obtains widely attentions. LPP preserves the local structure of the data by recovering the intrinsic nonlinear manifold structure of the original space. The weight graph constructed in LPP is simply used the nearest neighbor points based on the Euclidean distances as the metric. Thus, LPP is sensitive to noises or corruptions in the data. LRPP not only introduces the low-rank property to keep the global structure of the data but also weakens the disturbance of the noise in the projection subspace. When $\alpha = \beta = 0$, the main part of the optimization problem (16) is degraded to the model of LPP, thus LPP is a special case of LRPP.

SPP and NPE have similar objective functions. Both of them are related to LLE. In fact, NPE is a linearized version of LLE and SPP actually is a regularized extension of NPE. While LRPP constructs the weight matrix in a completely different manner from NPE and SPP, which use the least squared reconstruction error and sparse reconstruction error, respectively. In particular, LRPP constructs the weight graph using all the training samples with low-rank constraint instead of k nearest neighbors.

Both OLPP and ONPE add the orthogonality constraint to improve the locality preserving power and reconstructive ability. However, OLPP and ONPE construct the weight matrix using the k nearest neighbors the same as LPP and NPE, which are completely different LRPP. LRPP proposes to harness local preserving projections, sparsity, and low rankness of high-dimensional data to build a graph. The sparsity and local preserving projections can capture the local relationships of the data, and the low rankness can better recover the global structure of the data. LRPP requires that the coefficient vectors of all data samples from a low-rank matrix.

LRR use the lowest rank representation to construct the affinities of an undirected graph to capture the global structure of the whole data. The solution of LRR can be solved by a nuclear norm minimization problem. Similar to LRR, LRPP also use a nuclear norm minimization to solve the optimization solution of the weight matrix. Without the first term, (16) is degraded to the optimization model of LRR. Thus, we can take LRR as a special case of LRPP. However, LRPP learns the optimal weight matrix and the projection at the same time, which is totally different from LRR.

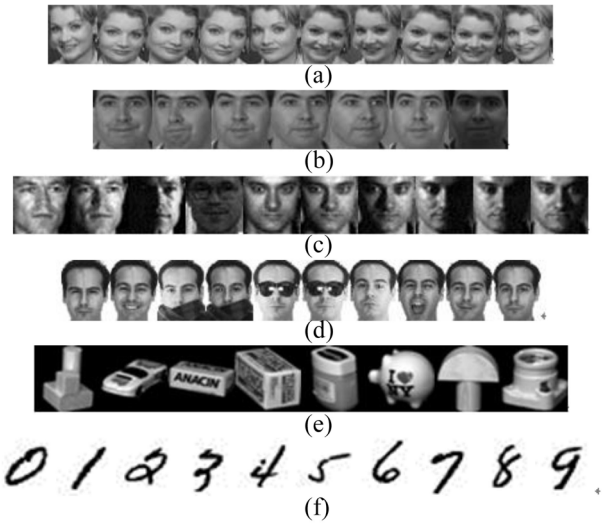


Fig. 1. Some images from (a) ORL, (b) FERET, (c) CMU PIE, (d) AR, (e) COIL20, and (f) MNIST.

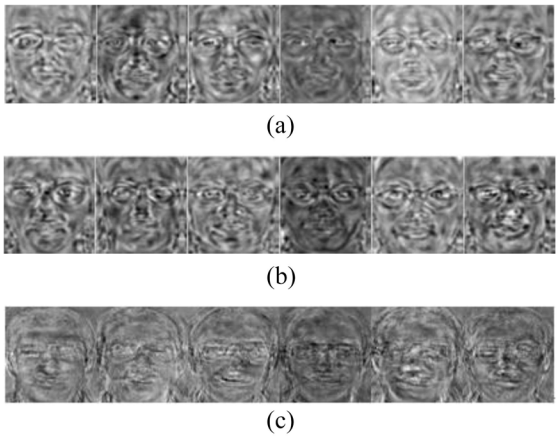


Fig. 2. Six basis vectors of LPP, NPE, and LRPP calculated from the training set of the ORL database. (a) LPPfaces. (b) NPEfaces. (c) LRPPfaces.

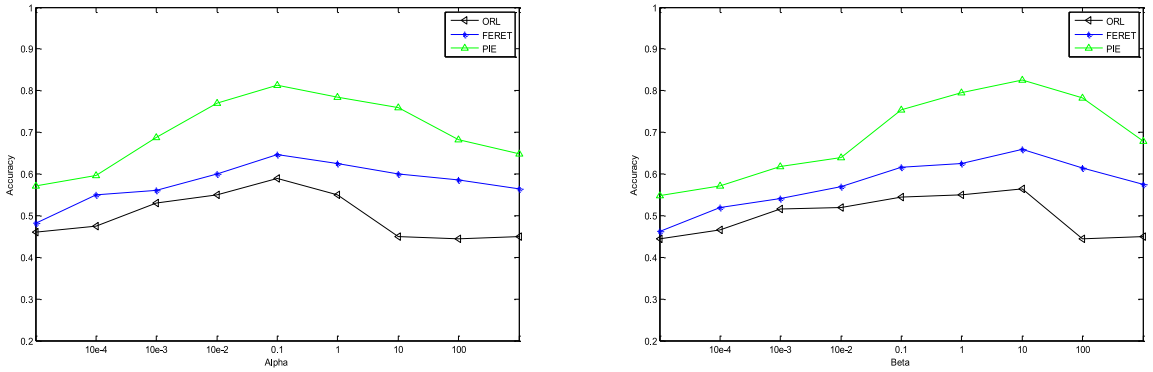
V. Experiments Analysis

In this section, to systematical evaluate the performance of the proposed method, i.e., LRPP, we verify the classification performance of LRPP in terms of random pixel corruptions, various level contiguous occlusions and the realistic occlusions (i.e., sunglasses and scarf).

A. Databases

Six different publicly available databases were used in our experiments, i.e., ORL [46], CMU PIE [47], FERET [48], AR [49], COIL20 [50], and MNIST [51] databases. Fig. 1 shows some example images from the above six databases. The first five databases are used for testing the performance of LRPP in image classification. The MNIST database is used for testing the performance of LRPP in

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b)

Fig. 3. Classification accuracy versus (a) with fixed and (b) with fixed on the ORL, FERET, and PIE database with the salt and pepper noise.

handwritten digit recognition. The corresponding experiments on the MNIST database are given in Section V-G.

The ORL face database composes of 40 distinct subjects. All subjects are in up-right, frontal position (with tolerance for some side movement). The CMU PIE database contains 41 368 face images collected from 68 subjects. Each subject has 13 images of different poses, 43 different illumination conditions, with four different expressions. In our experiment, a subset of five near frontal poses (C05, C07, C09, C25, and C29) and illuminations indexed as 08 and 11 was used. Therefore, each subject has ten images. The FERET database [48] used for evaluating face recognition algorithms displays diversity across gender, ethnicity, and age. The image set was acquired without any restrictions imposed on facial expression and with at least two frontal images shot at different times during the same photo session. For the FERET database, we randomly choose 70 people and six images of each subject to construct a subset. Thus, the total number of images is 420. The AR [49] face database 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 most persons were taken in two sessions (separated by two weeks). Each section contains 13 color images and 120 individuals (65 men and 55 women) participated in both sessions. In our experiments, we choose a subset of the AR database consisting of 50 men and 50 women. The COIL20 [50] database contains 20 objects and each object has 72 gray images which are taken from different view directions. All images in these five databases were normalized to 32×32 pixels and each image was reshaped to a vector.

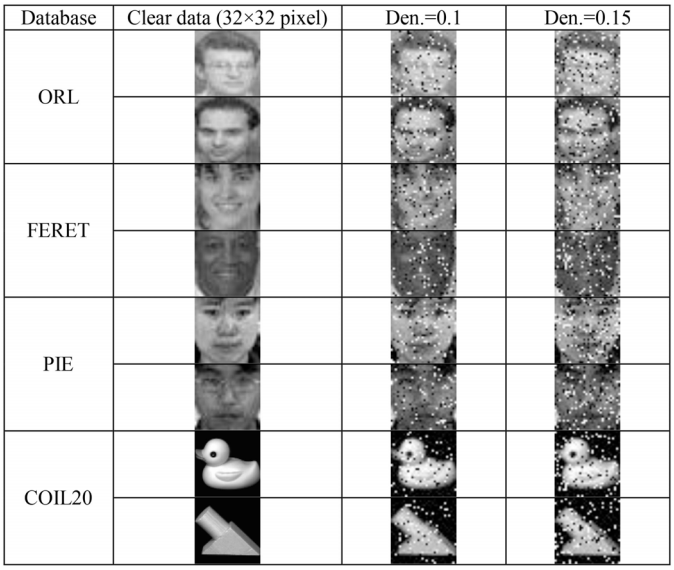


Fig. 4. Some examples of the original and corrupted images under different densities (Den.) of the salt and pepper noise from the ORL, FERET, PIE, and COIL20 databases.

B. Baselines and Parameter Setting

In our experiments, we test the performance of the proposed method on random pixel corruptions, various level contiguous occlusions and the realistic occlusions. We used the ORL, PIE, FERET, and COIL20 databases to test the performance of LRPP with the data containing random pixel corruptions and various level contiguous occlusions. For the ORL, PIE, and FERET databases, we randomly select half of the images per class as training samples, i.e., five, five, and three images per subject, respectively. For the COIL20 database, we randomly select 30 images per class to construct the training set, the rest samples are constructed the test set. The AR database was used for testing the classification accuracy of LRPP about face images with sunglasses and scarf occlusions. We compare our method with the algorithms of LPP [1], NPE [17], OLPP [22], ONPE [23], SPP [24], LRR [34], and RPCA [33]. In our experiments, we selected values for α and β in the range [0, 100]. According to the experiments, we can obtain the best classification accuracy when the parameters values are chosen in the range [0.001, 10]. Fig. 3 shows classification accuracy versus α with β fixed, and β with α fixed on the ORL, FERET, and PIE database with the "salt and pepper" noise. In this experiment, we randomly selected half images from each subject as training samples and the rest as test samples. These trails were independently conducted ten times, and the average classification accuracy was reported. The nearest neighbor classifier based on the Euclidean distance as the metric is used to calculate the percentage of samples in the test

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

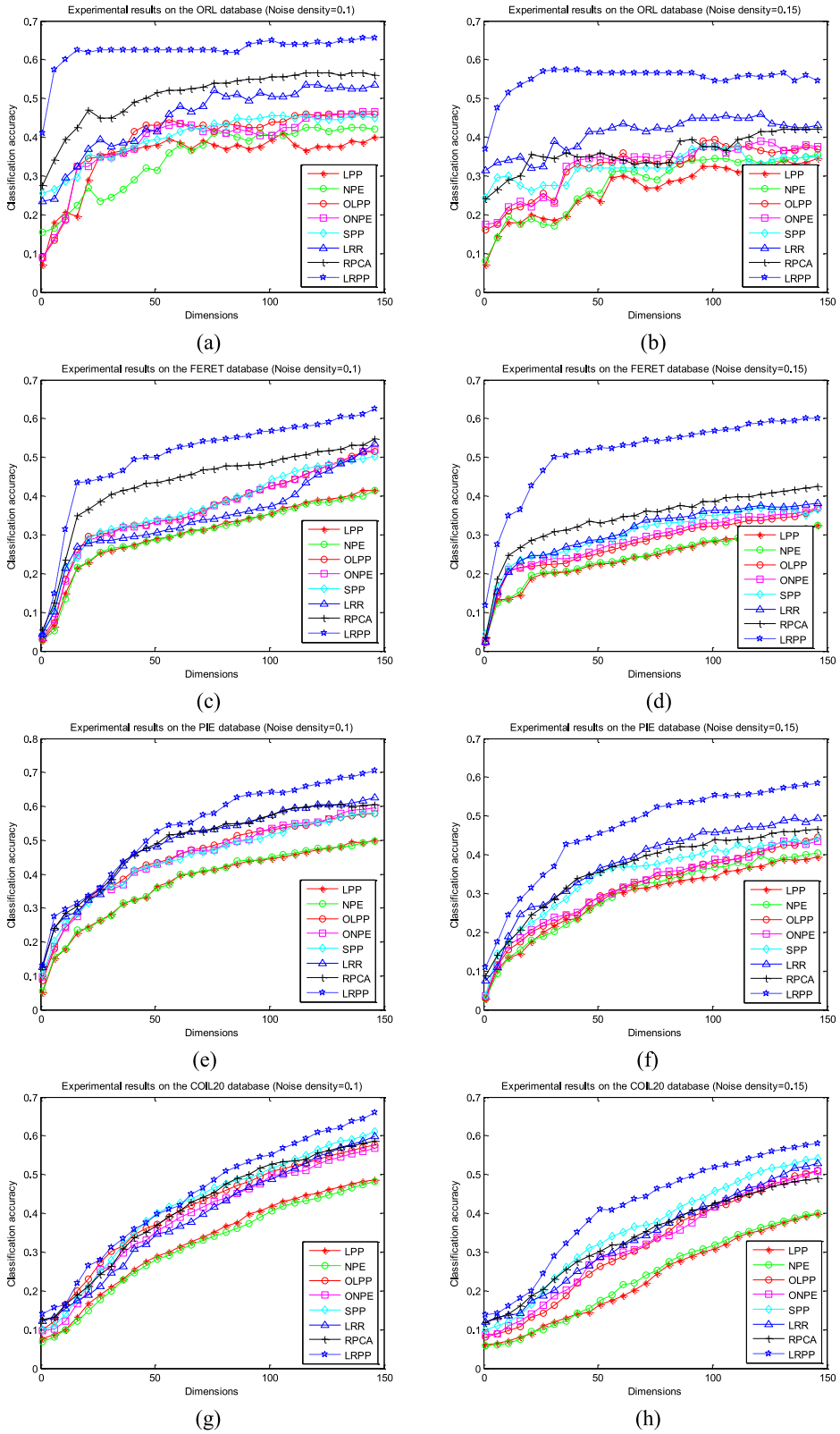


Fig. 5. Classification accuracies of LPP, NPE, OLPP, ONPE, SPP, LRR, RPCA, and LRPP on the ORL, FERET, PIE, and COIL20 databases with the salt and pepper noise. Experimental results on (a) ORL database (noise density = 0.1), (b) ORL database (noise density = 0.15), (c) FERET database (noise density = 0.1), (d) FERET database (noise density = 0.15), (e) PIE database (noise density = 0.1), (f) PIE database (noise density = 0.15), (g) COIL20 database (noise density = 0.1), and (h) COIL20 database (noise density = 0.15).

set that were correctly classified. Based on the training set of the ORL database, six LPPfaces, NPEfaces, and LRPPfaces are shown in Fig. 2.

C. Robustness Test With Random Pixel Corruptions

In order to test the robustness of LRPP of the data with random pixel corruptions, we add the salt and pepper noise

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

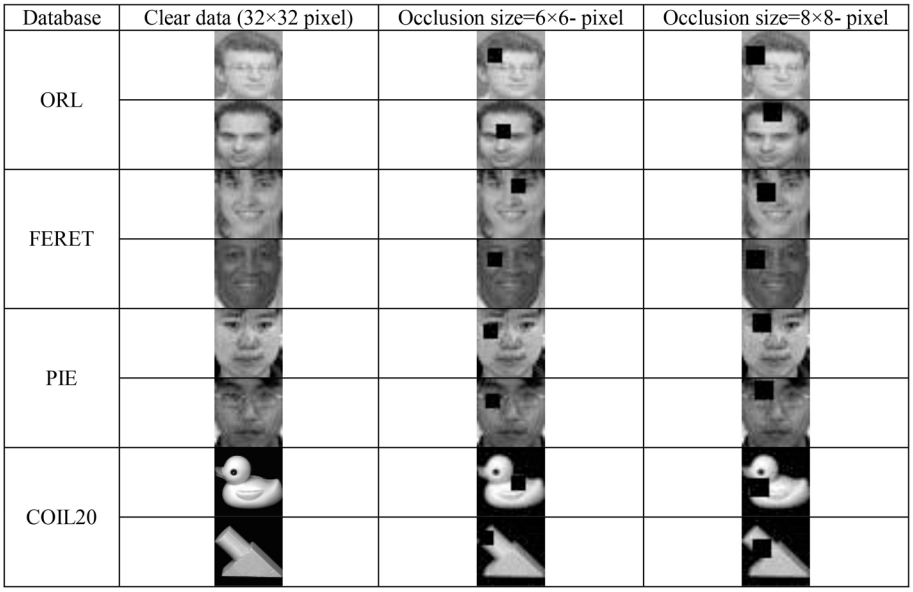


Fig. 6. Some examples of the original and corrupted images under different levels of contiguous occlusions from the ORL, FERET, PIE, and COIL20 databases.

into ORL, FERET, PIE, and COIL20 databases. The salt and pepper noise added into the data with two different densities: 0.1 and 0.15. Fig. 4 shows some images of the original and corrupted images under two different densities of the salt and pepper noise from the ORL, FERET, PIE, and COIL20 databases. Fig. 5 shows the classification accuracies of LPP, NPE, OLPP, ONPE, SPP, LRR, RPCA, and LRPP with different dimensions on random pixel corruptions with two different densities of the salt and pepper noise.

It can be found that LRPP outperforms the other comparison methods in the subspaces with different dimensions.

D. Robustness Test With Various Level Contiguous Occlusions

To test the robustness of LRPP of the data with various level contiguous occlusions, we randomly add some blocks to different locations in images and keep the remaining part unchanged. The sizes of blocks in images were set with different sizes: 6×6 and 8×8. Fig. 6 shows some examples of the original and corrupted images under different levels of contiguous occlusions from the ORL, FERET, PIE, and COIL20 databases. Fig. 7 shows the classification accuracies of LPP, NPE, OLPP, ONPE, SPP, LRR, RPCA, and LRPP with different dimensions on the various level contiguous occlusions.

As can be seen from Fig. 7, LRPP obtains the best recognition rates of the four groups of experiments, which shows the robustness for block occlusions when there are variations in illumination and expressions.

E. Robustness Test With Sunglasses Occlusion

In this section, to test the robustness of the proposed method on the sunglasses occlusion of face images, we conduct experiments on the AR database. We randomly choose three neutral images plus three images with sunglasses from session 1 as training samples. The test set is constructed by seven neutral images plus three images with sunglasses from session 2. Fig. 8 shows face images of the first subject used in our experiments. Images in the first row are training samples from session 1 and images in the second row are test samples from session 2. Fig. 9 illustrates the classification accuracy in the case of sunglasses occlusion.

As can be seen from Fig. 9, LRPP has the highest recognition rate among all the comparison methods. Thus, LRPP is more robust than the other methods of face images with the sunglasses occlusion.

F. Robustness Test With Scarf Occlusion

We mainly test the robustness of the proposed method on the scarf occlusion of face images in this section. Experiments were also conducted on the AR database. The training set is constructed by three randomly chose neutral images plus three images with scarf from session 1. The test set is constructed by seven neutral images plus three images with scarf from session 2. Fig. 10 shows face images of the first subject used in experiments. Images in the first row are training samples from session 1. Images in the second row are test samples from session 2.

G. Experiments on the MNIST Database

In this section, we conduct experiments on the MNIST database [51] to verify the performance of LRPP in handwritten digit recognition. The MNIST database contains 60 000 training samples and 10 000 test samples. The size of each image is 28×28. The task is to classify each image into one of the ten digits. The writers of the training set and test set are different. Fig. 12 shows some images of the original and corrupted images under two different densities of the salt

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

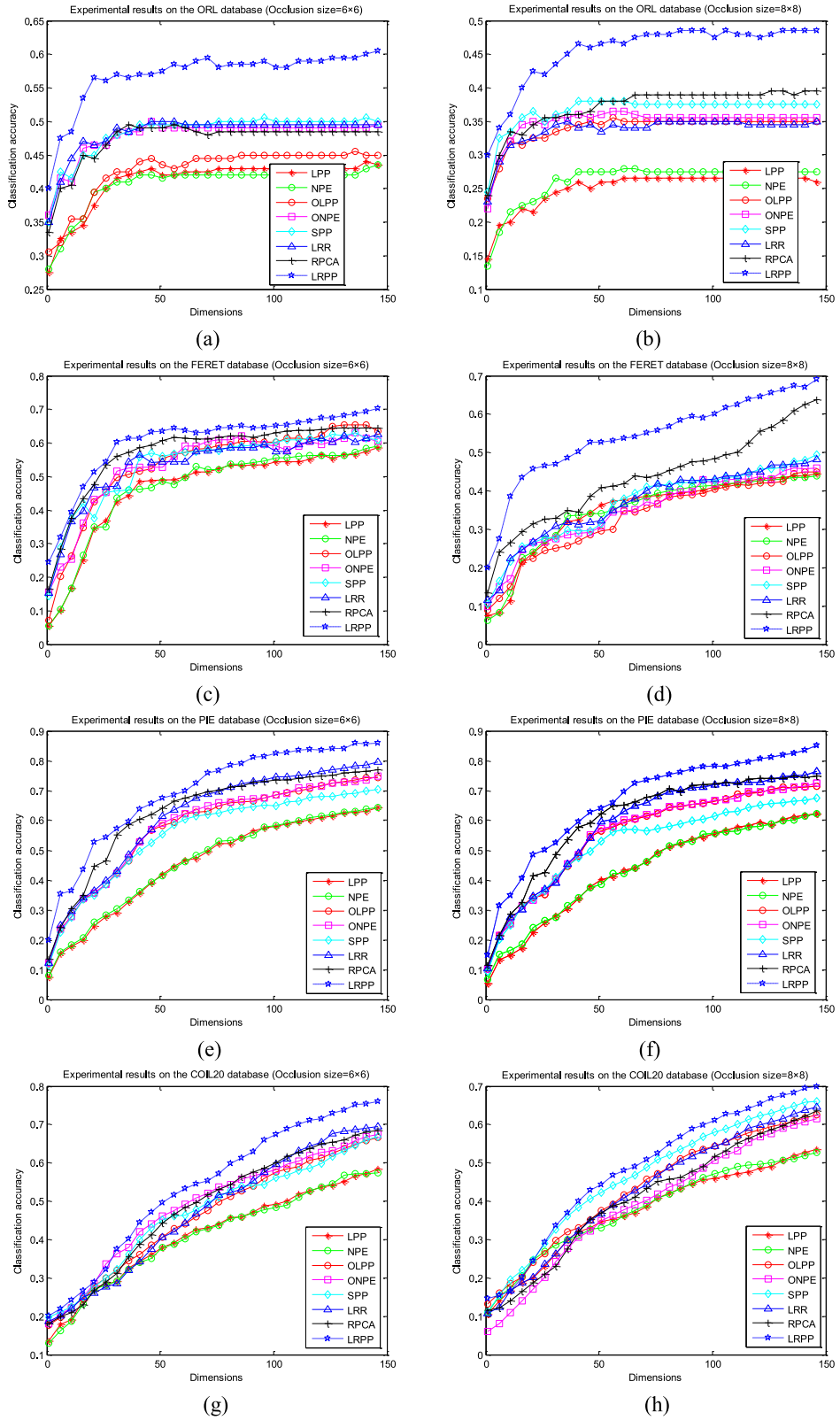


Fig. 7. Classification accuracies of LPP, NPE, OLPP, ONPE, SPP, LRR, RPCA, and LRPP on the ORL, FERET, PIE, and COIL20 database with different block occlusions, respectively. Experimental results on (a) ORL database (occlusion size $= 6 \times 6$), (b) ORL database (occlusion size $= 8 \times 8$), (c) FERET database (occlusion size $= 6 \times 6$), (d) FERET database (occlusion size $= 8 \times 8$), (e) PIE database (occlusion size $= 6 \times 6$), (f) PIE database (occlusion size $= 8 \times 8$), (g) COIL20 database (occlusion size $= 6 \times 6$), and (h) COIL20 database (occlusion size $= 8 \times 8$).

and pepper noise from the MNIST database. In our experiments, we randomly select 3000 images from 60 000 training samples to construct the training set and randomly select 5000 test samples to construct the test set. Fig. 13 illustrates the classification accuracy of the MNIST database with the salt and pepper noises.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

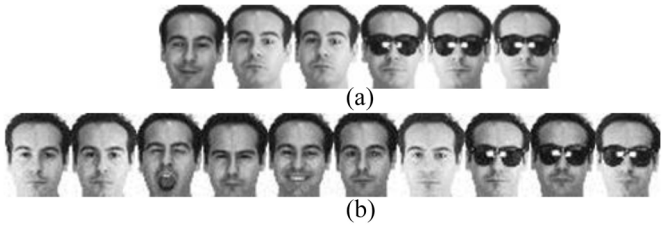


Fig. 8. Images examples of training and test in the sunglasses occlusion experiments. (a) Training samples. (b) Test samples.

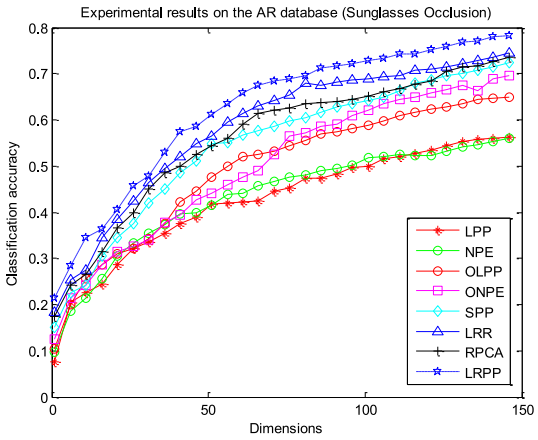


Fig. 9. Classification accuracies of LPP, NPE, OLPP, ONPE, SPP, LRR, RPCA, and LRPP on the AR database with sunglasses occlusion.

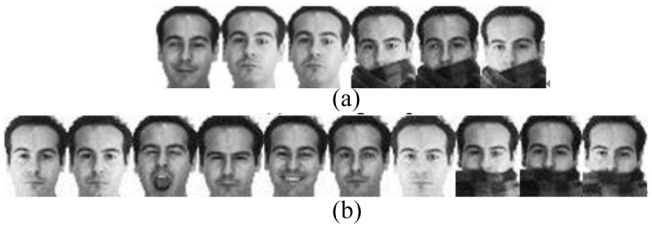


Fig. 10. Images examples of training and test in the scarf occlusion experiments. (a) Training samples. (b) Test samples.

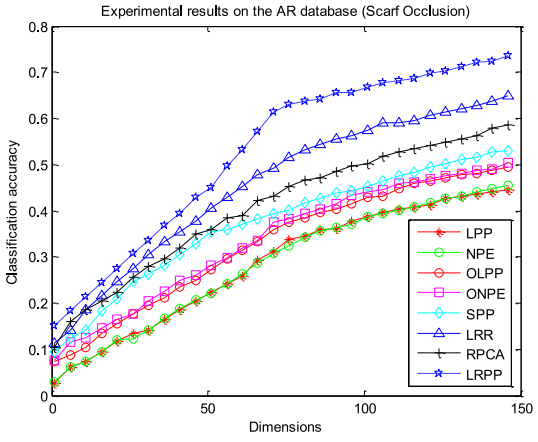


Fig. 11. Classification accuracies of LPP, NPE, OLPP, ONPE, SPP, LRR, RPCA, and LRPP on the AR database with scarf occlusion.

H. Observations and Discussions

From the experimental results, we can obtain the following observations and discussions.

1) Although the orthogonality information was introduced in OLPP and ONPE, LRPP still performed better than OLPP and ONPE, which indicates that combining projections and LRR for robust images classification provides more discriminative information than orthogonality constraint on the projections.
2) LRPP was more robust than the other compared methods and both OLPP and ONPE outperformed LPP and NPE in different dimensional subspace. That is, with increasing number of dimensions, OLPP and ONPE are more effective than LPP and NPE.
3) LRPP performed better than LPP and NPE, which indicates that globality is more important than locality in the robust image classification. In addition, LPP and

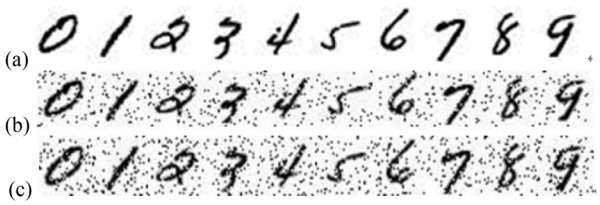
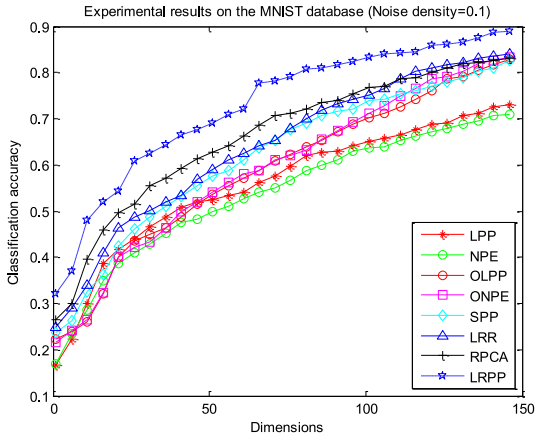
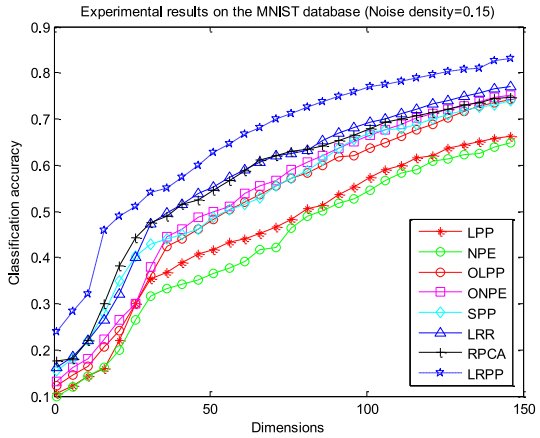


Fig. 12. Some example images from the MNIST database. (a) Original images. Images with the salt and pepper noise which density are (b) 0.1 and (c) 0.15.



(a)



(b)

Fig. 13. Classification accuracies of LPP, NPE, OLPP, ONPE, SPP, LRR, RPCA, and LRPP on the MNIST database with the salt and pepper noise. Density of the salt and pepper noise are (a) 0.1 and (b) 0.15.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

NPE almost have the similar classification accuracy, which indicates that using the Euclidean distances as the metric to measure the local geometric structure could not always improve the performance of robust image classification.

4) From Figs. 5, 7, 9, 11, and 13, we can see that LPP and NPE have lower classification accuracies among the comparison methods. This is because that the performance of LPP and NPE depend on the pair-wise Euclidean distances which is sensitive to the data corrupted with noises or errors. LRPP obtains the best classification performance due to the fact that it captures the global structure of the data to explore the latent discriminant information.

VI. Conclusion

In this paper, we encode preserving projections, sparsity, and low rankness of images data to build a graph and propose a novel dimensionality reduction method, named LRPP for image feature extraction and classification. LRPP well captured the global structure of the data and learned a low-rank weight matrix by projecting the data on a low-dimensional subspace. The L_{21} norm is introduced as a sparse constraint on the noise matrix and the nuclear norm as a low rank constraint on the weight matrix. Experimental results on six well-known databases showed the excellent performance of LRPP against the state-of-the-art preserving projection methods in robust image classification. It is shown that LRPP is robust to contiguous occlusions, random pixel corruptions, sunglasses and scarf occlusion of face images.

References

[1] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[2] P. N. Belhumeur, J. P. Hepanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[3] J. Yang, D. Zhang, and J. Y. Yang, "Constructing PCA baseline algorithms to reevaluate ICA-based face-recognition performance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 4, pp. 1015–1021, Aug. 2007.

[4] A. M. Martinez and A. C. Kak, "Principal components analysis versus linear discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[5] W. Zuo, D. Zhang, J. Yang, and K. Wang, "BDPCA plus LDA: A novel fast feature extraction technique for face recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 4, pp. 946–953, Aug. 2006.

[6] Y. Pang, S. Wang, and Y. Yuan, "Learning regularized LDA by clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2191–2201, Dec. 2014.

[7] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[8] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.

[9] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[10] 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.

[11] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[12] T. Zhang, K. Huang, X. Li, J. Yang, and D. Tao, "Discriminative orthogonal neighborhood-preserving projections for classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 1, pp. 253–263, Feb. 2010.

[13] Y. Bengio *et al.*, "Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2004, pp. 1–8.

[14] Z. Lai, W. K. Wong, Y. Xu, C. Zhao, and M. Sun, "Sparse alignment for robust tensor learning," *IEEE Trans. Neural Netw. Learn. Syst.,* vol. 25, no. 10, pp. 1779–1792, Oct. 2014.

[15] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacian faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[16] J. Lu and Y. P. Tan, "Regularized locality preserving projections and its extensions for face recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40. no. 3, pp. 958–963, Jun. 2010.

[17] X. He, D. Cai, S. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proc. Int. Conf. Comput. Vis. (ICCV),* Beijing, China, 2005, pp. 1208–1213.

[18] Y. Pang, N. Yu, H. Li, R. Zhong, and Z. Liu, "Face recognition using neighborhood preserving projections," in *Advances in Multimedia Information Processing* (LNCS 3768). Berlin, Germany: Springer, 2005, pp. 854–864.

[19] Y. Pang, L. Zhang, Z. Liu, N. Yu, and H. Li, "Neighborhood preserving projections (NPP): A novel linear dimension reduction method," in *Advances in Intelligent Computing* (LNCS 3644). Berlin, Germany: Springer, 2005, pp. 117–125.

[20] Z. Ji, Y. Pang, Y. He, and H. Zhang, "Semi-supervised LPP algorithms for learning-to-rank-based visual search reranking," *Inf. Sci.*, vol. 302, pp. 83–93, May 2015.

[21] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 29, no. 12, pp. 2143–2156, Dec. 2007.

[22] D. Cai, X. He, J. Han, and H. Zhang, "Orthogonal Laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3608–3614, Nov. 2006.

[23] X. Liu, J. Yin, Z. Feng, J. Dong, and L. Wang, "Orthogonal neighborhood preserving embedding for face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 1. San Antonio, TX, USA, 2007, pp. I-133–I-136.

[24] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognit.,* vol. 43, no. 1, pp. 331–341, 2010.

[25] Y. Pang, Z. Ji, P. Jing, and X. Li, "Ranking graph embedding for learning to rerank," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 8, pp. 1292–1303, Aug. 2013.

[26] Y. Pang, K. Zhang, Y. Yuan, and K. Wang, "Distributed object detection with linear SVMs," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2122–2133, Nov. 2014.

[27] Y. Pang, H. Yan, Y. Yuan, and K. Wang, "Robust CoHOG feature extraction in human-centered image/video management system," *IEEE Trans. Syst., Man, Cybern. B, Cybern.,* vol. 42, no. 2, pp. 458–468, Apr. 2012.

[28] B. Geng, D. Tao, and C. Xu, "DAML: Domain adaptation metric learning," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2980–2989, Oct. 2011.

[29] C. Xu, D. Tao, and C. Xu, "Large-margin multi-view information bottleneck," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 36, no. 8, pp. 1559–1572, Aug. 2014.

[30] B. Xie, Y. Mu, D. Tao, and K. Huang, "m-SNE: Multiview stochastic neighbor embedding," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 1088–1096, Aug. 2011.

[31] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell*., to be published.

[32] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Neural Inf. Process. Syst.*, Whistler, BC, Canada, 2009, pp. 1–9.

[33] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," *J. ACM*, vol. 58, no. 3, pp. 1–17, 2009.

[34] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. ICML*, Haifa, Israel, 2010, pp. 663–670.

[35] L. Zhuang *et al.*, "Non-negative low rank and sparse graph for semi-supervised learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2012, pp. 2328–2335.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[36] Y. Song, F. Nie, C. Zhang, and S. Xiang, "A unified framework for semi-supervised dimensionality reduction," *Pattern Recognit.*, vol. 41, no. 9, pp. 2789–2799, 2008.
[37] D. Zhang, Z. Zhou, and S. Chen, "Semi-supervised dimensionality reduction," in *Proc. SIAM Int. Conf. Data Min.*, vol. 15. Pittsburgh, PA, USA, 2007, pp. 629–634.
[38] S. Yan *et al.*, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach.*, vol. 29, no. 1, pp. 40–51, 2007.
Y. Lu, C. Lu, M. Qi, and S. Wang, "Graph embedding and extensions: A general framework for dimensionality reduction," (LNCS 6059), 2010, pp. 28–37.
[39] Z. Zheng, F. Yang, W. Tan, J. Jia, and J. Yang, "Gabor feature-based face recognition using supervised locality preserving projection," *Signal Process.*, vol. 87, pp. 2473–2483, Oct. 2007.
[40] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., San Francisco, CA, USA, Jun. 2010, pp. 1791–1798.
[41] B. K. Bao, G. Liu, C. Xu, and S. Yan, "Inductive robust principal component analysis," *IEEE Trans. Image Process*., vol. 21, no. 8, pp. 3794–3800, Aug. 2012.
[42] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low rank representation," in *Proc. NIPS*, Granada, Spain, 2011, pp. 612–620.
[43] J. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[44] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 569–592, 2009.
[45] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: Wiley, 2000.
[46] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. 2nd IEEE Int. Workshop Appl. Comput. Vis.*, Sarasota, FL, USA, 1994, pp. 138–142.
[47] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 25, no. 12, pp. 1615–1618, Dec. 2003.
[48] P. J. Phillips. (2004). *The Facial Recognition Technology (FERET) Database*. [Online]. Available: http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
[49] A. M. Martinez and R. Benavente, "The AR face database," Centre de Visio per Computador, Univ. Auton. Barcelona, Barcelona, Spain, Tech. Rep. 24, Jun. 1998.
[50] S. Nene, S. Nayar, and H. Murase, "Columbia object image library (COIL-20)," Dept. Comput. Sci., Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-006-96, 1996.
[51] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

Yuwu Lu received the B.S. degree in mathematics from the XingTai University in 2008 and the M.S. degree in mathematics from the Inner Mongolia University of Technology in 2011. He is currently pursuing the Ph.D. degree in computer science and technology with the Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China.

His current research interests include pattern recognition and machine learning. He has published eight scientific papers.

Zhihui Lai received the B.S. degree in mathematics from South China Normal University, Guangzhou, China, in 2002, the M.S. degree in mathematics from Jinan University, Guangzhou, in 2007, and the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, Nanjing, China, in 2011.

He has been a Research Associate with the Hong Kong Polytechnic University, Hong Kong, since 2010. His current 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. He is the author of over 30 scientific papers in pattern recognition and computer vision.

Yong Xu (M'06–SM'15) was born in Sichuan, China, in 1972. He received the B.S. and M.S. degrees from the PLA University of Science and Technology, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, Nanjing, China, in 2005.

He is currently with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China. His current research interests include pattern recognition, biometrics, machine learning, image processing, and video analysis.

Xuelong Li (M'02–SM'07–F'12) received the Ph.D. degree from the University of Science and Technology of China, Hefei, China. He is a Full Professor with the Center for OPTical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.

David Zhang (F'08) received the bachelor's degree in computer science from Peking University, Beijing, China, in 1974, the M.Sc. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1982 and 1985, respectively, and the second Ph.D. degree in electrical and computer engineering from the University of Waterloo, ON, Canada, in 1994.

From 1986 to 1988, he was a Post-Doctoral Fellow with Tsinghua University, Beijing, and an Associate Professor with the Academia Sinica, Beijing. He is currently the Chair Professor with the Hong Kong Polytechnic University, Hong Kong, where he is also the Founding Director of the Biometrics Technology Centre (UGC/CRC), supported by the Hong Kong SAR Government in 1998. He also serves as the Visiting Chair Professor with Tsinghua University, and an Adjunct Professor with Shanghai Jiao Tong University, Shanghai, China, Peking University, Harbin Institute of Technology, and the University of Waterloo. He has authored over ten books and 200 journal papers. He is the editor of the book entitled *International Series on Biometrics* (Springer).

Dr. Zhang is the Founder and an Editor-in-Chief of the *International Journal of Image and Graphics*, an Associate Editor of over ten international journals including the IEEE TRANSACTIONS AND PATTERN RECOGNITION. He was also the Organizer of the First International Conference on Biometrics Authentication, and is the Technical Committee Chair of IEEE Computational Intelligence Society. He is a Croucher Senior Research Fellow, the Distinguished Speaker of the IEEE Computer Society, and a fellow of International Association of Pattern Recognition.

Chun Yuan received the M.S. and Ph.D. degrees in computer science and technology from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 1999 and 2002, respectively.

He is currently an Associate Professor with the Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University, Shenzhen, China. He was a Post-Doctoral Research Fellow with the INRIA-Rocquencourt, Paris, France, from 2003 to 2004. In 2002, he was with Microsoft Research Asia, Beijing, China, as an Intern. His current research interests include computer vision, machine learning, video coding and processing, cryptography, and digital rights management.