

Discriminative Orthogonal Neighborhood-Preserving Projections for Classification

Tianhao Zhang, *Member, IEEE*, Kaiqi Huang, *Senior Member, IEEE*,
Xuelong Li, *Senior Member, IEEE*, Jie Yang, and Dacheng Tao, *Member, IEEE*

Abstract—Orthogonal neighborhood-preserving projection (ONPP) is a recently developed orthogonal linear algorithm for overcoming the out-of-sample problem existing in the well-known manifold learning algorithm, i.e., locally linear embedding. It has been shown that ONPP is a strong analyzer of high-dimensional data. However, when applied to classification problems in a supervised setting, ONPP only focuses on the intraclass geometrical information while ignores the interaction of samples from different classes. To enhance the performance of ONPP in classification, a new algorithm termed discriminative ONPP (DONPP) is proposed in this paper. DONPP 1) takes into account both intraclass and interclass geometries; 2) considers the neighborhood information of interclass relationships; and 3) follows the orthogonality property of ONPP. Furthermore, DONPP is extended to the semisupervised case, i.e., semisupervised DONPP (SDONPP). This uses unlabeled samples to improve the classification accuracy of the original DONPP. Empirical studies demonstrate the effectiveness of both DONPP and SDONPP.

Index Terms—Classification, dimensionality reduction, discriminative orthogonal neighborhood-preserving projection (DONPP), patch alignment.

I. INTRODUCTION

IN REAL classification tasks [11], [19], [24], [25], [27], [32], the raw data gathered from sensors, like cameras or radars, are often very high dimensional, so direct operations on them are computationally expensive and obtain less-than-optimal results. This is the so-called “curse of dimensionality” [4], which

Manuscript received January 25, 2008; revised April 7, 2008 and April 9, 2008. First published September 9, 2009; current version published October 30, 2009. This work was supported in part by the Nanyang Technological University SUG Grant under Project M58020010, by the National Science Foundation of China under Grant 60675023, by the China 863 High Tech. Plan under Grant 2007AA01Z164, and by the 100 Talents Program of The Chinese Academy of Sciences. This paper was recommended by Associate Editor R. Lynch.

T. Zhang is with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: z.tianhao@gmail.com).

K. Huang is with the National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing 100080, China (e-mail: kqhuang@nlpr.ia.ac.cn).

X. Li is with the State Key Laboratory of Transient Optics and Photonics, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, China (e-mail: xuelong_li@opt.ac.cn).

J. Yang is with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jieyang@sjtu.edu.cn).

D. Tao is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798, and also with the University of London, WC1E 7HX London, U.K., Xidian University, Xi’an 710071, China, and Wuhan University, Wuhan 430070, China (e-mail: dacheng.tao@gmail.com).

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/TSMCB.2009.2027473

is invariably dealt with by dimensionality reduction to reveal the intrinsic structure of such data [8], [10]. The problem of dimensionality reduction can be described as follows. Consider a data set X , which consists of N samples \vec{x}_i ($1 \leq i \leq N$) in a high-dimensional space \mathbb{R}^m . That is, $X = [\vec{x}_1, \dots, \vec{x}_N] \in \mathbb{R}^{m \times N}$. The objective of a dimensionality reduction algorithm is to compute the faithful low-dimensional representations $Y = [\vec{y}_1, \dots, \vec{y}_N] \in \mathbb{R}^{d \times N}$, where $d < m$, of X .

The most popular conventional algorithms for dimensionality reduction are principal component analysis (PCA) [16] and linear discriminant analysis (LDA) [9]. PCA is based on the computation of low-dimensional representations of high-dimensional data that maximize the total scatter. PCA is not suitable for classification problems since it does not use prior knowledge of class identities, that is, it is unsupervised. Unlike PCA, LDA is supervised. It finds the projection directions that maximize the trace of the between-class scatter matrix while at the same time minimizing the trace of the within-class scatter matrix. However, both are conventional algorithms that see only the global Euclidean structure and cannot discover the nonlinear structure hidden in the high-dimensional data.

Recently, a number of manifold learning algorithms have been developed, which promise to be useful for analyzing the high-dimensional data that lie on or near a submanifold of the observation space. Representative ones include locally linear embedding (LLE) [21], ISOMAP [26], Laplacian eigenmaps (LE) [2], and local tangent space alignment (LTSA) [30]. LLE first uses linear coefficients, which reconstruct a given sample by its neighbors, to represent the local geometry and then seeks a low-dimensional embedding, in which these coefficients are still suitable for reconstruction. ISOMAP, a variant of multidimensional scaling [8], preserves the global geodesic distances of all pairs of samples. LE preserves proximity relationships by manipulations on an undirected weighted graph, which indicates neighbor relations of pairwise samples. LTSA exploits the local tangent information as a representation of the local geometry, and this local tangent information is then aligned to provide a global coordinate. Unfortunately, all of these algorithms suffer from the out-of-sample problem [5].

He *et al.* address the out-of-sample problem by applying a linearization procedure to construct explicit maps over new samples. Examples of this approach include locality-preserving projections (LPP) [13], a linearization of LE and neighborhood-preserving embedding (NPE) [14], a linearization of LLE. In the recent research [18], Kokiopoulou and Saad point out that enforcing an orthogonality relationship between the projection directions is more effective for preserving the intrinsic

TABLE I
IMPORTANT NOTATIONS USED IN THIS PAPER

Notation	Description	Notation	Description
X	given dataset	L_i	representation of part optimization
Y	dimension-reduced dataset	S_i	selection matrix
N	size of the dataset X	F_i	indices for the i^{th} patch
m	dimension of original samples	I_d	$d \times d$ identity matrix
d	reduced dimension	\vec{e}_l	$[1, \dots, 1]^T \in \mathbb{R}^l$
\mathbb{R}^m	m -dimensional Euclidean space	\vec{c}_i	reconstruction coefficients
\vec{z}_k	$[0, \dots, 0]^T \in \mathbb{R}^k$	k	number of neighbors having different labels
X_i	i^{th} patch	β	scaling factor
\vec{x}_i	given sample in X	γ	control parameter
n_i	number of samples in class of \vec{x}_i	U	projection matrix

geometrical structure of raw data, and propose a new algorithm called orthogonal neighborhood-preserving projection (ONPP) as the linearization of LLE with the orthogonal constraint over the projection matrix. The empirical studies in [18] have shown that ONPP is more competitive than NPE (NPP in [18]). However, for classification problems, ONPP (in a supervised setting) only focuses on the intraclass geometrical information while ignores the interaction of samples from different classes.

In this paper, to enhance the classification performance of ONPP, we propose a new algorithm termed discriminative ONPP (DONPP) based on the perspective of patch alignment [31], [34]. DONPP offers three main benefits: 1) the algorithm takes into account both intraclass and interclass geometries so that it can achieve better performance in classification; 2) discriminability is effectively preserved in the algorithm because it takes into account neighbor samples having different labels; and 3) the algorithm produces the orthogonal projection matrix that follows from the ONPP property.

Moreover, this paper extends DONPP to the semisupervised case, i.e., semisupervised DONPP (SDONPP), by incorporating the additional unlabeled samples. SDONPP is more powerful since it can make use of unlabeled and labeled samples.

The rest of the paper is organized as follows: In Section II, we describe the new algorithm, i.e., DONPP. In Section III, the algorithm is extended to the semisupervised case, i.e., SDONPP. In Section IV, the experimental results are given, and Section V presents our conclusions.

II. DONPP

The proposed DONPP algorithm is based on the perspective of patch alignment [31], [34], which operates in two stages, i.e., part optimization and whole alignment. Part optimization involves building a patch using each sample and some of its related samples in the data set. On each of the built patches, the objective function is designed according to the algorithm. Since each sample can be seen as a part of whole samples, the stage is termed by “part optimization.” Whole alignment involves combining all the part optimizations to form the final global coordinate using the alignment trick [30]. The projection matrix can be obtained by solving the eigendecomposition problem. Table I shows the important notations used in this paper.

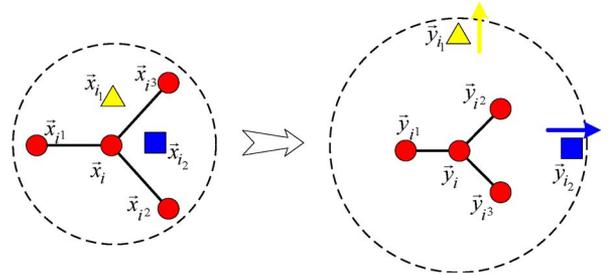


Fig. 1. Part optimization of DONPP. Samples with the same shape and color have the same label.

A. Part Optimization

For a given sample \vec{x}_i in X , we denote $n_i - 1$ samples having the same label as \vec{x}_i by $\vec{x}_{i1}, \dots, \vec{x}_{in_i-1}$, where n_i denotes the number of samples in the class of \vec{x}_i . We select its k nearest neighbors (NNs) having different labels from \vec{x}_i and denote them by $\vec{x}_{i1}, \dots, \vec{x}_{ik}$. By combining $\vec{x}_i, \vec{x}_{i1}, \dots, \vec{x}_{in_i-1}$ and $\vec{x}_{i1}, \dots, \vec{x}_{ik}$, we can build the patch for the sample \vec{x}_i as $X_i = [\vec{x}_i, \vec{x}_{i1}, \dots, \vec{x}_{in_i-1}, \vec{x}_{i1}, \dots, \vec{x}_{ik}]$.

The output of each patch in the low-dimensional space is denoted by $Y_i = [\vec{y}_i, \vec{y}_{i1}, \dots, \vec{y}_{in_i-1}, \vec{y}_{i1}, \dots, \vec{y}_{ik}]$. In the subspace, we expect to preserve the geometrical relationships between the given sample and the ones having the same label using linear reconstruction coefficients. In the meantime, we expect that the distances between the given sample and the neighbor samples having different labels are as large as possible. Fig. 1 illustrates how something happens in part optimization. The dotted circle on the left represents the i^{th} patch in an original high-dimensional space, and the patch consists of \vec{x}_i samples having the same label (i.e., $\vec{x}_{i1}, \vec{x}_{i2}$, and \vec{x}_{i3}) and neighbor samples having different labels (i.e., \vec{x}_{i4} and \vec{x}_{i5}). The expected results on the patch in the low-dimensional space are shown in Fig. 1 (right). The linear relationships between \vec{y}_i and $\vec{y}_{i1}, \vec{y}_{i2}, \vec{y}_{i3}$ are preserved invariant, whereas \vec{y}_{i4} and \vec{y}_{i5} are as far away as possible from \vec{y}_i .

DONPP assumes that each sample can be reconstructed using samples having the same label. That is, on the patch, \vec{x}_i can be linearly reconstructed from $\vec{x}_{i1}, \dots, \vec{x}_{in_i-1}$ as

$$\vec{x}_i = (\vec{c}_i)_1 x_{i1} + (\vec{c}_i)_2 x_{i2} + \dots + (\vec{c}_i)_{n_i-1} x_{in_i-1} + \vec{e}_i \quad (1)$$

where \vec{c}_i is an $n_i - 1$ -dimensional vector encoding reconstruction coefficients, and $\vec{\varepsilon}_i$ is the reconstruction error. Minimizing the error yields

$$\arg \min_{\vec{c}_i} \|\vec{\varepsilon}_i\|^2 = \arg \min_{\vec{c}_i} \left\| \vec{x}_i - \sum_{j=1}^{n_i-1} (\vec{c}_i)_j \vec{x}_{ij} \right\|^2. \quad (2)$$

Using the sum-to-one constraint $\sum_{j=1}^{n_i-1} (\vec{c}_i)_j = 1$, \vec{c}_i can be computed in closed form as

$$(\vec{c}_i)_j = \frac{\sum_{t=1}^{n_i-1} G_{jt}^{-1}}{\sum_{p=1}^{n_i-1} \sum_{q=1}^{n_i-1} G_{pq}^{-1}} \quad (3)$$

where $G_{jt} = (\vec{x}_i - \vec{x}_{ij})^T (\vec{x}_i - \vec{x}_{it})$ is called the local Gram matrix [22].

DONPP assumes that \vec{c}_i reconstructs both \vec{x}_i from $\vec{x}_{i1}, \dots, \vec{x}_{i n_i-1}$ in the high-dimensional space and \vec{y}_i from $\vec{y}_{i1}, \dots, \vec{y}_{i n_i-1}$ in the low-dimensional space. Based on this point, the cost function is written as

$$\arg \min_{\vec{y}_i} \left\| \vec{y}_i - \sum_{j=1}^{n_i-1} (\vec{c}_i)_j \vec{y}_{ij} \right\|^2. \quad (4)$$

Denoting the associated samples in (4) using the notation of patch Y_i , the equation is converted into

$$\arg \min_{Y_i} \text{tr} \left(Y_i \begin{bmatrix} -1 \\ \vec{c}_i \\ \vec{z}_k \end{bmatrix} [-1 \quad \vec{c}_i^T \quad \vec{z}_k^T] Y_i^T \right) = \arg \min_{Y_i} \text{tr} (Y_i L_i^n Y_i^T) \quad (5)$$

where $L_i^n = \begin{bmatrix} -1 \\ \vec{c}_i \\ \vec{z}_k \end{bmatrix} [-1 \quad \vec{c}_i^T \quad \vec{z}_k^T] = \begin{bmatrix} 1 & -\vec{c}_i^T & \vec{z}_k^T \\ -\vec{c}_i & \vec{c}_i \vec{c}_i^T & \vec{z}_k \vec{z}_k^T \\ \vec{z}_k & \vec{z}_k \vec{z}_k^T & I_k \end{bmatrix}$, and \vec{z}_k is the k -dimensional vector with all entries being 0.

Meanwhile, we expect that neighbor samples having different labels $\vec{y}_{i1}, \dots, \vec{y}_{ik}$ are as far away as possible from the given sample \vec{y}_i . To make this happen, we maximize the sum of the distances between \vec{y}_i and $\vec{y}_{i1}, \dots, \vec{y}_{ik}$ so that we have

$$\arg \max_{\vec{y}_i} \sum_{p=1}^k \|\vec{y}_i - \vec{y}_{ip}\|^2. \quad (6)$$

Using the notation Y_i , (6) is converted into

$$\arg \max_{Y_i} \text{tr} \left(Y_i \begin{bmatrix} -\vec{e}_k^T \\ \vec{z}_{n_i-1} \vec{z}_k^T \\ I_k \end{bmatrix} [-\vec{e}_k \quad \vec{z}_k \vec{z}_{n_i-1}^T \quad I_k] Y_i^T \right) = \arg \max_{Y_i} \text{tr} (Y_i L_i^{Ex} Y_i^T) \quad (7)$$

where $L_i^{Ex} = \begin{bmatrix} -\vec{e}_k^T \\ \vec{z}_{n_i-1} \vec{z}_k^T \\ I_k \end{bmatrix} [-\vec{e}_k \quad \vec{z}_k \vec{z}_{n_i-1}^T \quad I_k] = \begin{bmatrix} k & \vec{z}_{n_i-1}^T & -\vec{e}_k^T \\ \vec{z}_{n_i-1} & \vec{z}_{n_i-1} \vec{z}_{n_i-1}^T & \vec{z}_{n_i-1} \vec{z}_k^T \\ -\vec{e}_k & \vec{z}_k \vec{z}_{n_i-1}^T & I_k \end{bmatrix}$, $\vec{e}_k = [1, \dots, 1]^T \in \mathbb{R}^k$, and I_k is the $k \times k$ identity matrix.

Now, for one patch, we have the two optimizations described as (4) and (6). Considering them simultaneously, we can formulate the discriminator on the i th patch X_i as the linear form

$$\begin{aligned} & \arg \min_{\vec{y}_i} \left(\left\| \vec{y}_i - \sum_{j=1}^{n_i-1} (\vec{c}_i)_j \vec{y}_{ij} \right\|^2 - \beta \sum_{p=1}^k \|\vec{y}_i - \vec{y}_{ip}\|^2 \right) \\ &= \arg \min_{Y_i} \text{tr} (Y_i (L_i^n - \beta L_i^{Ex}) Y_i^T) \\ &= \arg \min_{Y_i} \text{tr} \left(Y_i \begin{bmatrix} 1 & -\vec{c}_i^T & \vec{z}_k^T \\ -\vec{c}_i & \vec{c}_i \vec{c}_i^T & \vec{z}_{n_i-1} \vec{z}_k^T \\ \vec{z}_k & \vec{z}_k \vec{z}_{n_i-1}^T & \vec{z}_k \vec{z}_k^T \end{bmatrix} \right. \\ & \quad \left. - \beta \begin{bmatrix} k & \vec{z}_{n_i-1}^T & -\vec{e}_k^T \\ \vec{z}_{n_i-1} & \vec{z}_{n_i-1} \vec{z}_{n_i-1}^T & \vec{z}_{n_i-1} \vec{z}_k^T \\ -\vec{e}_k & \vec{z}_k \vec{z}_{n_i-1}^T & I_k \end{bmatrix} Y_i^T \right) \\ &= \arg \min_{Y_i} \text{tr} (Y_i L_i Y_i^T) \end{aligned} \quad (8)$$

where

$$L_i = \begin{bmatrix} 1 - \beta k & -\vec{c}_i^T & -\beta \vec{e}_k^T \\ -\vec{c}_i & \vec{c}_i \vec{c}_i^T & \vec{z}_{n_i-1} \vec{z}_k^T \\ -\beta \vec{e}_k & \vec{z}_k \vec{z}_{n_i-1}^T & -\beta I_k \end{bmatrix} \quad (9)$$

encapsulates the information of objective for the i th patch, and β is a scaling factor in $[0, 1]$ to unify the two different measures.

B. Whole Alignment

For each patch X_i , $i = 1, \dots, N$, we have the part optimizations described in (8). In this section, these optimizations will be combined by assuming that the coordinate for the i th patch $Y_i = [\vec{y}_i, \vec{y}_{i1}, \dots, \vec{y}_{i n_i-1}, \vec{y}_{i1}, \dots, \vec{y}_{ik}]$ is selected from the global coordinate $Y = [\vec{y}_1, \dots, \vec{y}_N]$ such that

$$Y_i = Y S_i \quad (10)$$

where $S_i \in \mathbb{R}^{N \times (n_i+k)}$ is the selection matrix, and an entry is defined as

$$(S_i)_{pq} = \begin{cases} 1, & \text{if } p = (F_i)_q \\ 0, & \text{else} \end{cases} \quad (11)$$

where $F_i = \{i, i^1, \dots, i^{n_i-1}, i_1, \dots, i_k\}$ denotes the set of indexes for the i th patch. Then, (8) can be rewritten as

$$\arg \min_Y \text{tr} (Y S_i L_i S_i^T Y^T). \quad (12)$$

By summing over all the part optimizations described in (12), we can obtain the whole alignment as

$$\begin{aligned} & \arg \min_Y \sum_{i=1}^N \text{tr} (Y S_i L_i S_i^T Y^T) \\ &= \arg \min_Y \text{tr} \left(Y \sum_{i=1}^N S_i L_i S_i^T Y^T \right) \\ &= \arg \min_Y \text{tr} (Y L Y^T) \end{aligned} \quad (13)$$

where $L = \sum_{i=1}^N S_i L_i S_i^T \in \mathbb{R}^{N \times N}$ is the alignment matrix [30]. This is obtained using an iterative procedure as

$$L(F_i, F_i) \leftarrow L(F_i, F_i) + L_i \quad (14)$$

for $i = 1, \dots, N$ with the initialization $L = 0$.

To obtain the linear and orthogonal projection matrix U , such as $Y = U^T X$, we can impose $U^T U = I_d$, which is similar to that reported in [18], where I_d is the $d \times d$ identity matrix. Equation (13) is deformed as

$$\begin{aligned} & \arg \min_U \text{tr} (U^T X L X^T U) \\ & \text{s.t. } U^T U = I_d. \end{aligned} \quad (15)$$

Obviously, the solutions of (15) are obtained by solving a standard eigendecomposition problem

$$X L X^T \vec{u} = \lambda \vec{u}. \quad (16)$$

Let the column vectors $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_d$ be the solutions of (16) ordered according to eigenvalues $\lambda_1 < \lambda_2 < \dots < \lambda_d$. The optimal projection matrix U is given by $U = [\vec{u}_1, \vec{u}_2, \dots, \vec{u}_d]$.

The procedure for the proposed DONPP is described in the following (note that the first step is to use PCA [1], [17] to reduce noise):

- 1) Use PCA to project the data set X into the subspace to eliminate the useless information. We still denote by X the data set in the PCA subspace. We denote the PCA projection matrix by U_{PCA} .
- 2) For each sample \vec{x}_i in X , $i = 1, \dots, N$, build the patch $X_i = [\vec{x}_i, \vec{x}_{i_1}, \dots, \vec{x}_{i_{k_S-1}}, \vec{x}_{i_1}, \dots, \vec{x}_{i_k}]$, where $\vec{x}_{i_1}, \dots, \vec{x}_{i_{k_S-1}}$ are samples having the same label, and $\vec{x}_{i_1}, \dots, \vec{x}_{i_k}$ are k neighbor samples having different labels.
- 3) Compute L_i by (9) and construct the alignment matrix L using the iterative procedure described in (14).
- 4) Solve the standard eigenvalue problem $X L X^T \vec{u} = \lambda \vec{u}$ to obtain the DONPP projection matrix $U_{\text{DONPP}} = [\vec{u}_1, \vec{u}_2, \dots, \vec{u}_d]$, whose vectors are the eigenvectors corresponding to the d smallest eigenvalues. The final projection matrix is $U = U_{\text{PCA}} U_{\text{DONPP}}$.

III. SDONPP

Recent research [3], [7], [15], [29], [33], [35] has found that unlabeled samples may be helpful to improve the classification

performance. In this section, we generalize DONPP by introducing new part optimizations based on unlabeled samples and then incorporating them into the whole alignment stage as an SDONPP.

Unlabeled samples are attached to the original data set: $X = [\vec{x}_1, \dots, \vec{x}_N, \vec{x}_{N+1}, \dots, \vec{x}_{N+N_U}]$, where the first N samples are labeled, and the remaining N_U samples are unlabeled. The part optimization for each labeled sample is given by (8).

Unlabeled samples are valuable in improving the local geometry. For each unlabeled sample \vec{x}_i , $i = N + 1, \dots, N + N_U$, we search its k_S NNs $\vec{x}_{i_1}, \dots, \vec{x}_{i_{k_S}}$ in all training samples including both labeled and unlabeled samples. Let $X_i = [\vec{x}_i, \vec{x}_{i_1}, \dots, \vec{x}_{i_{k_S}}]$ denote the i th patch, and the associated index set is given by $F_i^U = \{i, i_1, \dots, i_{k_S}\}$.

We use linear reconstruction coefficients \vec{c}_i^U to represent the geometrical information on the patch. \vec{c}_i^U 's are obtained by

$$\arg \min_{\vec{c}_i^U} \left\| \vec{x}_i - \sum_{j=1}^{k_S} (\vec{c}_i^U)_j \vec{x}_{i_j} \right\|^2 \quad (17)$$

which can be solved similarly as described in Section II-A. In the low-dimensional space, we assume that \vec{y}_i can be reconstructed by $\vec{y}_{i_1}, \dots, \vec{y}_{i_{k_S}}$ with \vec{c}_i^U , i.e.,

$$\arg \min_{\vec{y}_i} \left\| \vec{y}_i - \sum_{j=1}^{k_S} (\vec{c}_i^U)_j \vec{y}_{i_j} \right\|^2. \quad (18)$$

According to $Y_i = [\vec{y}_i, \vec{y}_{i_1}, \dots, \vec{y}_{i_{k_S}}]$, the optimization on the i th patch is expressed as

$$\begin{aligned} & \arg \min_{Y_i} \text{tr} \left(Y_i \begin{bmatrix} -1 \\ \vec{c}_i^U \end{bmatrix} \begin{bmatrix} -1 & (\vec{c}_i^U)^T \end{bmatrix} Y_i^T \right) \\ &= \arg \min_{Y_i} \text{tr} (Y_i L_i^U Y_i^T) \end{aligned} \quad (19)$$

where

$$L_i^U = \begin{bmatrix} -1 \\ \vec{c}_i^U \end{bmatrix} \begin{bmatrix} -1 & (\vec{c}_i^U)^T \end{bmatrix} = \begin{bmatrix} 1 & -(\vec{c}_i^U)^T \\ -\vec{c}_i^U & \vec{c}_i^U (\vec{c}_i^U)^T \end{bmatrix}. \quad (20)$$

Considering all the samples including both labeled and unlabeled samples, we have the whole alignment

$$\begin{aligned} & \arg \sum_{i=1}^N \min_{Y_i} \text{tr} (Y_i L_i Y_i^T) + \gamma \arg \sum_{i=N+1}^{N+N_U} \min_{Y_i} \text{tr} (Y_i L_i^U Y_i^T) \\ &= \arg \min_Y \text{tr} \left(Y \left(\sum_{i=1}^N S_i^L L_i (S_i^L)^T \right. \right. \\ & \quad \left. \left. + \sum_{i=N+1}^{N+N_U} S_i^U \gamma L_i^U (S_i^U)^T \right) Y^T \right) \\ &= \arg \min_Y \text{tr} (Y L^S Y^T) \end{aligned} \quad (21)$$

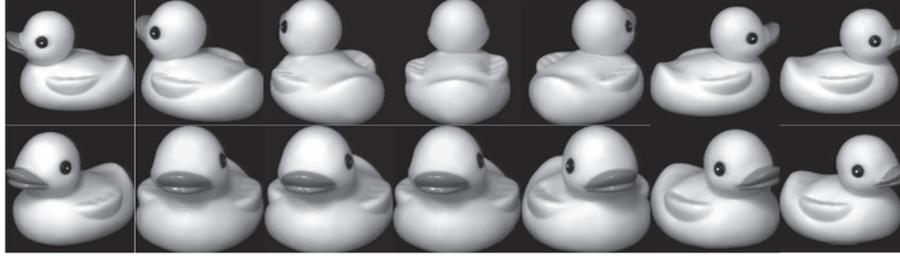


Fig. 2. Sample images from COIL-20.

where γ is a control parameter in $[0, 1]$, $S_i^L \in \mathbb{R}^{(N+N_U) \times (n_i+k)}$ and $S_i^U \in \mathbb{R}^{(N+N_U) \times (k_S+1)}$ are the selection matrices defined similarly as in Section II-B, and $L^S \in \mathbb{R}^{(N+N_U) \times (N+N_U)}$ is the alignment matrix constructed by

$$\begin{cases} L^S(F_i, F_i) \leftarrow L^S(F_i, F_i) + L_i \\ \text{for } i = 1, \dots, N \\ L^S(F_i^U, F_i^U) \leftarrow L^S(F_i^U, F_i^U) + \gamma L_i^U \\ \text{for } i = N+1, \dots, N+N_U \end{cases} \quad (22)$$

with the initialization $L^S = 0$.

Similar to Section II-B, the problem is ultimately converted into a standard eigendecomposition problem $XL^S X^T \bar{u} = \lambda \bar{u}$. The projection matrix U_{SDONPP} contains eigenvectors associated with the d smallest eigenvalues. Like DONPP, PCA is also utilized to reduce the sample noise, and the final projection matrix is $U = U_{PCA} U_{SDONPP}$.

IV. EXPERIMENTS

In this section, we first compare the proposed DONPP algorithm with representative dimensionality reduction algorithms and then compare the SDONPP algorithm with the original DONPP algorithm. We apply the usual three steps used in recognition problems. First, each algorithm is applied to training samples to learn the projection matrices. Second, each testing sample is projected onto a low-dimensional subspace via a projection matrix. Finally, the testing samples in the projected subspace are identified using the NN classifier.

A. Experiments on DONPP

In this section, we make use of three publicly available databases, namely, COIL-20 [20], YALE [1], and UMIST [12]. Among them, COIL-20 contains images of various objects, whereas YALE and UMIST are the face image databases. The face images are cropped with reference to the eyes, and all images from these three databases are normalized to 40×40 pixel arrays with 256 gray levels per pixel. Each image is reshaped to one long vector by arranging its pixel values in a fixed order. We implement six algorithms, namely, LDA [1], LPP (LPP1 in [6]), NPE [14], marginal Fisher analysis (MFA) [28], ONPP [18], and DONPP. Note that, LPP, NPE, and ONPP are all implemented in the supervised setting.

1) *COIL-20*: The COIL-20 database [20] contains images of 20 objects rotated on a turntable taken from a fixed camera,

resulting in 72 images per object. Fig. 2 shows some images of one object. For each of the 20 objects, we randomly selected a different number of samples per object for training (5, 10, 15, or 20), and the remainders were used for testing. All the tests were repeated ten times, and we then calculated the average recognition results. Fig. 3 gives comparisons of the recognition rates of six algorithms under different reduced dimensions. Table II lists the best recognition rates of the six algorithms along with the corresponding dimensions. It can be seen that DONPP outperforms the other algorithms. Table II also provides the optimal value of k , which is the number of neighbor samples having different labels, for DONPP. In Section IV-C, we will discuss k in detail.

2) *YALE*: The YALE database [1] consists of 15 subjects, and each subject has 11 samples with varying facial expression and configurations. Fig. 4 shows the image set for one individual. For training, we randomly selected different numbers of images per individual (3, 5, 7, or 9) and used the remaining images for testing. We repeated these trials independently ten times and computed the average recognition results. Fig. 5 and Table III give the results for the six algorithms on YALE.

3) *UMIST*: The UMIST database [12] consists of a total of 564 face images of 20 people. The individuals are a mix of race, sex, and appearance and are photographed in a range of poses from profile to frontal views. Fig. 6 shows some images of an individual. We randomly chose different numbers of images per individual (3, 5, 7, or 9) for training, and the rest are used for testing. We repeated these trials ten times and computed the average results. Fig. 7 and Table IV show the experimental results. It is clear that we can draw similar observations as for COIL-20 and YALE.

B. Experiments on SDONPP

We compare SDONPP and DONPP based on the UMIST database with setting $k = 0$ in (8). The averaged recognition rates were obtained from ten different random runs. For each turn, different numbers (3, 5, 7, or 9) of samples with labels and five samples without labels for each individual were selected at random to train SDONPP and DONPP, and the remaining samples for each individual were used for testing. It is worth noting that five samples without labels have no effects in training DONPP. Table V lists all the results and shows that the unlabeled samples are helpful in increasing the recognition rates.

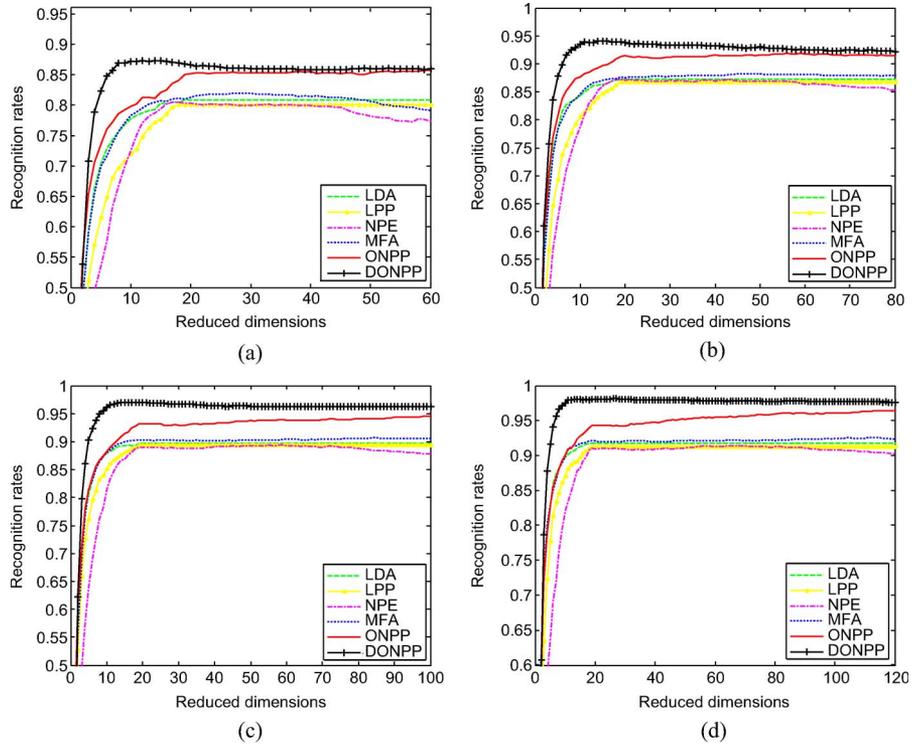


Fig. 3. Recognition rate versus subspace dimension on COIL-20. (a) Five samples for training. (b) Ten samples for training. (c) Fifteen samples for training. (d) Twenty samples for training.

TABLE II
BEST RECOGNITION RATES (IN PERCENT) OF SIX ALGORITHMS ON COIL-20

Number of Training	5	10	15	20
LDA	80.82 (19)	87.24 (19)	89.62 (19)	91.76 (19)
LPP	80.08 (19)	86.82 (19)	89.37 (19)	91.32 (19)
NPE	80.43 (18)	87.15 (18)	89.37 (55)	91.45 (67)
MFA	81.94 (27)	88.25 (45)	90.70 (84)	92.56 (114)
ONPP	85.82 (61)	91.83 (59)	94.81 (135)	96.44 (119)
DONPP	87.36 (14, 4)	94.11 (14, 2)	97.04 (15, 3)	98.13 (26, 3)

For LDA, LPP, NPE, MFA, and ONPP, the numbers in the parentheses are the subspace dimensions. For DONPP, the first numbers in the parentheses are the subspace dimensions and the second numbers are the parameter k .

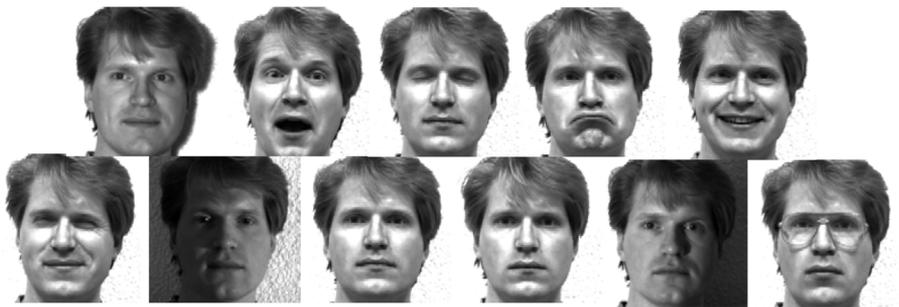


Fig. 4. Sample images from YALE.

C. Discussions

It is worth highlighting in the following a few points that arose from the experimental results:

- 1) DONPP is superior to ONPP since the algorithm considers not only the intraclass geometry but also the discriminative information derived from the interclass samples. Here, we use a simple toy example to confirm the point.

The toy data set plotted in Fig. 8 is employed. As shown, these discrete data are sampled from the two-layer Swiss roll, and each layer represents one class. Fig. 9 gives the 2-D outputs, yielded by ONPP and DONPP, respectively. Note that we perform the two algorithms without the PCA projection step since the dimension of the raw data is not high. It is shown that DONPP can map the two different class points separately and discovers the

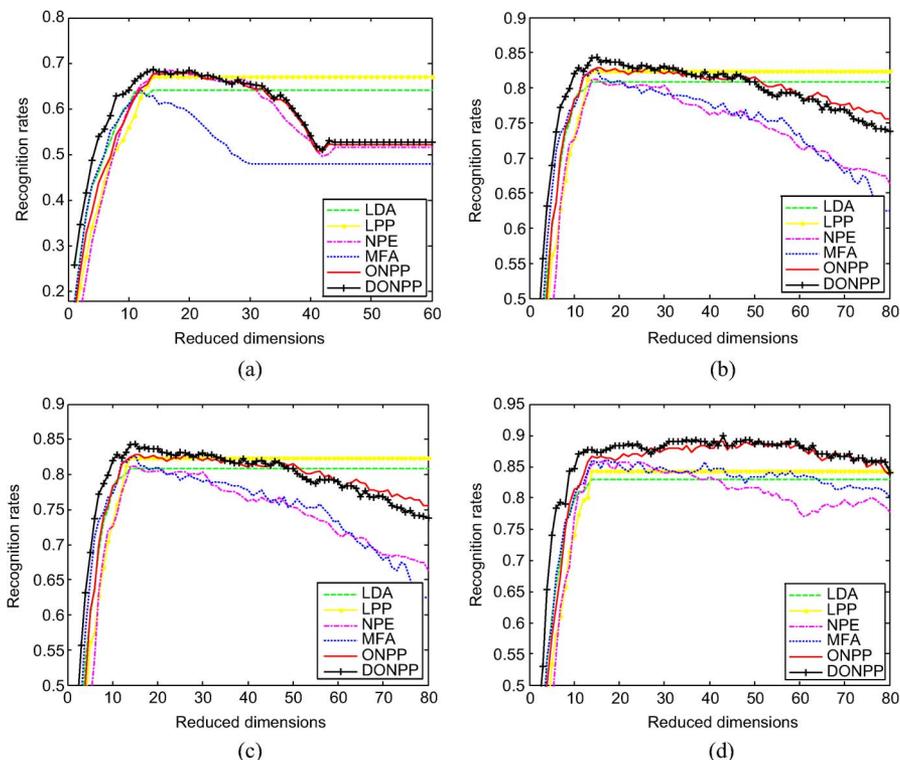


Fig. 5. Recognition rate versus subspace dimension on YALE. (a) Three samples for training. (b) Five samples for training. (c) Seven samples for training. (d) Nine samples for training.

TABLE III
BEST RECOGNITION RATES (IN PERCENT) OF SIX ALGORITHMS ON YALE

Number of Training	3	5	7	9
LDA	64.08 (14)	72.78 (14)	80.83 (14)	83.00 (14)
LPP	67.00 (14)	73.44 (14)	82.33 (14)	84.33 (14)
NPE	68.40 (20)	74.33 (15)	81.17 (14)	86.00 (17)
MFA	64.33 (12)	73.44 (15)	82.67 (15)	86.33 (17)
ONPP	67.90 (20)	77.00 (23)	82.83 (15)	89.33 (51)
DONPP	68.75 (14, 2)	77.44 (14, 3)	84.33 (15, 4)	90.00 (43, 2)

For LDA, LPP, NPE, MFA, and ONPP, the numbers in the parentheses are the subspace dimensions. For DONPP, the first numbers in the parentheses are the subspace dimensions and the second numbers are the parameter k .



Fig. 6. Sample images from UMIST.

intrinsic geometry of the raw data. However, ONPP fails to do so.

- 2) For patch building, k NNs of the sample having different labels can better characterize the discrimination than all samples having different labels. To examine this point, we studied the effects of k on the recognition rates based

on the UMIST database. In this experiment, three samples for each class were selected for training, and the reduced dimension was fixed to 9. As can be seen in part a of Fig. 10, by varying k from 0 to $N - n_i$ (57), we can obtain the recognition rate curve, and there is a peak on the curve when $k = 5$.

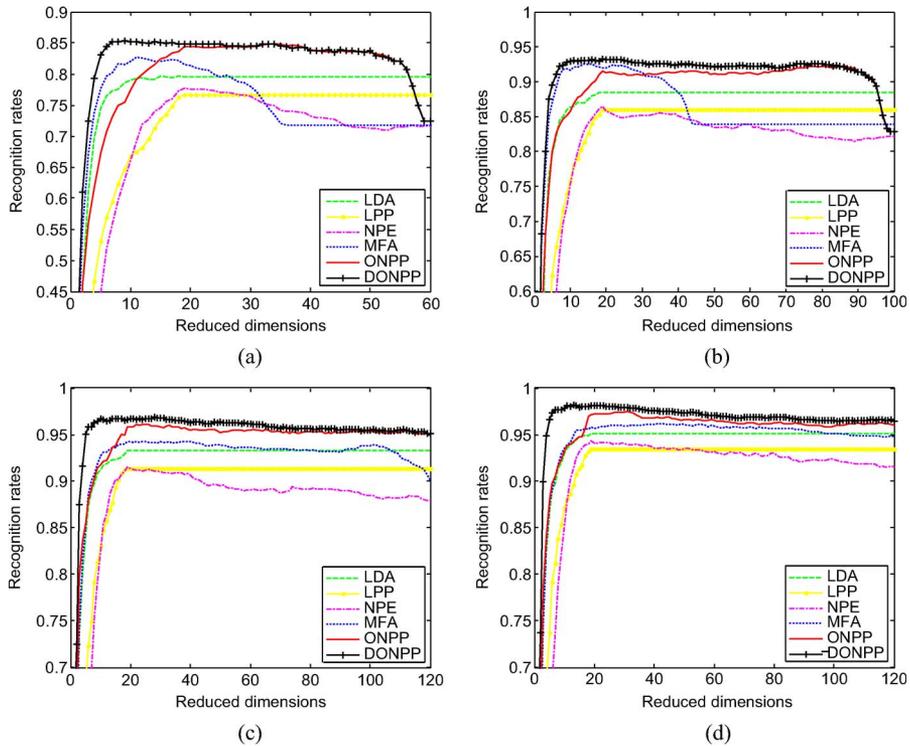


Fig. 7. Recognition rate versus subspace dimension on UMIST. (a) Three samples for training. (b) Five samples for training. (c) Seven samples for training. (d) Nine samples for training.

TABLE IV
BEST RECOGNITION RATES (IN PERCENT) OF SIX ALGORITHMS ON UMIST

Number of Training	3	5	7	9
LDA	79.71 (18)	88.51 (19)	93.31 (19)	95.14 (19)
LPP	76.58 (19)	86.06 (19)	91.36 (19)	93.44 (19)
NPE	77.65 (19)	86.53 (19)	91.52 (19)	94.30 (19)
MFA	82.64 (11)	92.61 (14)	94.28 (19)	96.20 (42)
ONPP	84.83 (34)	92.34 (86)	95.89 (25)	97.52 (31)
DONPP	85.86 (9, 5)	93.27 (19, 2)	96.85 (28, 1)	98.17 (13, 2)

For LDA, LPP, NPE, MFA, and ONPP, the numbers in the parentheses are the subspace dimensions. For DONPP, the first numbers in the parentheses are the subspace dimensions and the second numbers are the parameter k .

TABLE V
BEST RECOGNITION RATES (IN PERCENT) OF DONPP AND SDONPP ON UMIST

Number of Training (Labeled/Unlabeled)	3 / 5	5 / 5	7 / 5	9/5
DONPP	84.10	91.41	96.06	97.15
SDONPP	86.61	94.11	97.93	98.23

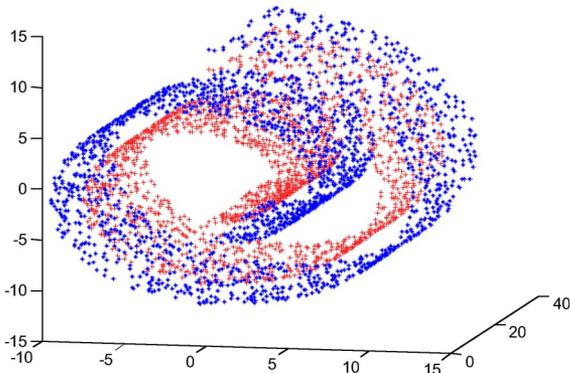


Fig. 8. Points sampled from the two-layer Swiss roll.

3) The linear reconstruction coefficients can exactly reflect the geometrical configuration of the associated samples, including the neighborhood information. Therefore, there is no need to consider the neighborhood information, that is, select the neighbors when capturing the intraclass geometry. To test this point, we studied the influence of k' , which is the number of selected neighbors having the same label, on the recognition rates based on the UMIST database. In this experiment, nine samples for each class were selected for training, and the reduced dimension was fixed at 13. By varying k' from 0 to $n_i - 1$ (8), the recognition rate curve with respect to k' can be obtained. As can be seen in part b of Fig. 10, there is no appreciable change in the recognition rate.

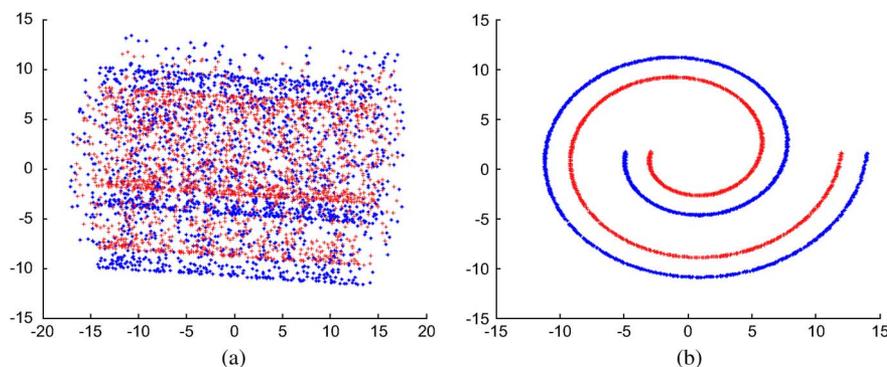


Fig. 9. Experimental results on the two-layer Swiss roll. (a) Result of ONPP. (b) Result of DONPP.

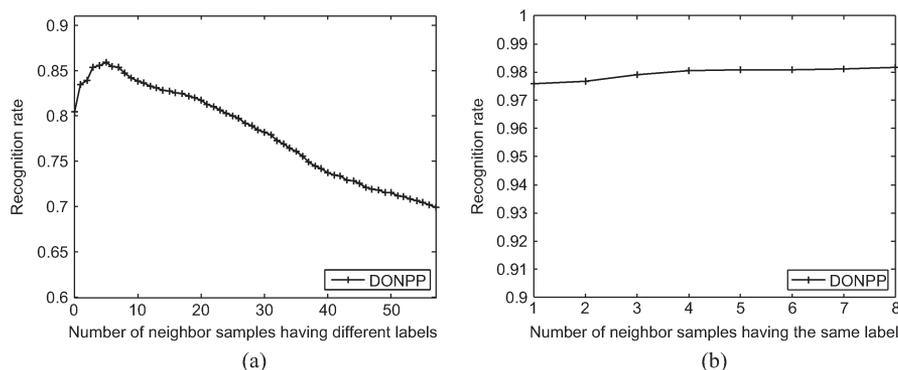


Fig. 10. Building a patch. (a) Recognition rate versus k . (b) Recognition rate versus k' .

- 4) In Fig. 5 of the experimental results on YALE, we have an interesting observation that the recognition rates start dropping significantly with the increase of dimension after an optimal value. The effective dimension of the projected subspace of DONPP is determined by the rank of XLX^T in (16), which intrinsically depends on the number of training samples (N). In the experiment on YALE, the number of training samples is small, so the number of effective dimension is small, and there is a drop for classification accuracy after an optimal value.

V. CONCLUSION

According to the patch alignment scheme, we have proposed DONPP, which is a new dimensionality reduction algorithm. DONPP is an enhanced variant of ONPP for classification problems. The advantages of DONPP are as follows: DONPP performs better than ONPP in classification, it preserves discriminative information over the built patches, and its projection matrix is orthogonal. Experiments have demonstrated the effectiveness of the proposed algorithm compared with representative dimensionality reduction algorithms, e.g., LDA, LPP, NPE, MFA, and ONPP. In addition, we have also developed SDONPP, which considers both labeled and unlabeled samples in the construction of the projection matrix. Experimental results have shown that unlabeled samples are helpful in further improving the classification accuracy of DONPP.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments on the first version of this paper.

REFERENCES

- [1] P. Belhumeur, J. Hespanha, 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.
- [2] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing System*, vol. 14. Cambridge, MA: MIT Press, 2002, pp. 585–591.
- [3] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, 2006.
- [4] R. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ: Princeton Univ. Press, 1961.
- [5] Y. Bengio, J. Paiement, P. Vincent, O. Dellalaeu, L. Roux, and M. Quimet, "Out-of sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering," in *Advances in Neural Information Processing System*, vol. 16. Cambridge, MA: MIT Press, 2004.
- [6] D. Cai, X. He, and J. Han, "Using graph model for face analysis," Univ. Illinois Urbana-Champaign, Urbana, IL, Department of Computer Science Tech. Rep. No. 2636, Sep. 2005.
- [7] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–7.
- [8] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ: Wiley, 2000.
- [9] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.
- [10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
- [11] Y. Fu and T. S. Huang, "Image classification using correlation tensor analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 226–234, Feb. 2008.

- [12] D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," in *Face Recognition: From Theory to Applications*, ser. NATO ASI Series F, Computer and Systems Science, vol. 163, H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. S. Huang, Eds. New York: Springer-Verlag, 1998, pp. 446–456.
- [13] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing System*, vol. 16. Cambridge, MA: MIT Press, 2004.
- [14] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1208–1213.
- [15] X. He, D. Cai, and J. Han, "Learning a maximum margin subspace for image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 189–201, Feb. 2008.
- [16] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, 1933.
- [17] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002.
- [18] 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.
- [19] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high order tensor data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 342–352, Apr. 2008.
- [20] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (COIL-20)," Columbia Univ., New York, Tech. Rep. CUCS-005-96, 1996.
- [21] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [22] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifold," *J. Mach. Learn. Res.*, vol. 4, pp. 119–155, 2003.
- [23] L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee, "Spectral methods for dimensionality reduction," in *Semisupervised Learning*, O. Chapelle, B. Schoelkopf, and A. Zien, Eds. Cambridge, MA: MIT Press, 2006.
- [24] 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.
- [25] 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.
- [26] J. Tenenbaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
- [27] J. Sun, D. Tao, S. Papadimitriou, P. Yu, and C. Faloutsos, "Incremental tensor analysis: Theory and applications," *ACM Trans. Knowl. Discov. Data*, vol. 2, no. 3, art. 11, pp. 1–37, Oct. 2008.
- [28] D. Xu, S. Yan, D. Tao, S. Lin, and H.-J. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content based image retrieval," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2811–2821, Nov. 2007.
- [29] S. Si, D. Tao, and B. Geng, "Bregman divergence based regularization for transfer subspace learning," *IEEE Trans. Knowl. Data Eng.*, [Online]. Available: <http://www2.computer.org/portal/web/csdl/doi/10.1109/TKDE.2009.126>
- [30] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2005.
- [31] T. Zhang, D. Tao, X. Li, and J. Yang, "A unifying framework for spectral analysis based dimensionality reduction," in *Proc. Int. Joint Conf. Neural Netw.*, 2008, pp. 1670–1677.
- [32] T. Zhang, X. Li, D. Tao, and J. Yang, "Multimodal biometrics using geometry preserving projections," *Pattern Recognit.*, vol. 41, no. 3, pp. 805–813, Mar. 2008.
- [33] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 725–738.
- [34] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, [Online]. Available: <http://www2.computer.org/portal/web/csdl/doi/10.1109/TKDE.2008.212>
- [35] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 912–919.



Tianhao Zhang (M'09) received the Ph.D. degree in pattern recognition and intelligence system from Shanghai Jiao Tong University, Shanghai, China, in 2008.

From 2007 to 2008, he was a Research Assistant (Visiting Ph.D.) with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. He is currently a Postdoctoral Researcher with the Section for Biomedical Image Analysis, University of Pennsylvania, Philadelphia. He is also a Reviewer for *Neurocomputing*, *Optical Engineering*, and the

International Journal of Computer Mathematics. His research interests include machine learning, computer vision, and medical image analysis.

Dr. Zhang is a Reviewer for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B (TSMC-B). He has extensively published in the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the TSMC-B, *Pattern Recognition*, the European Conference on Computer Vision, International Joint Conference on Neural Networks (IJCNN), etc. He has been a program committee member of the Pacific-Rim Symposium on Image and Video Technology in 2007, the International Conference on Image Analysis and Recognition in 2008, the International Conference on Image and Graphics in 2009, and the Pacific-Asia Conference on Knowledge Discovery and Data Mining in 2009, and a workshop in the International Conference on Data Mining in 2008. He also reviews for the IJCNN in 2008, the Medical Image Computing and Computer Assisted Intervention in 2009, and the International Conference on Acoustics, Speech, and Signal Processing in 2009.



Kaiqi Huang (M'07–SM'09) received the M.S. degree in electrical engineering from Nanjing University of Science and Technology, Nanjing, China, and the Ph.D. degree in signal and information processing from Southeast University, Nanjing.

He is currently an Associate Professor with the National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, China. His research interests include visual surveillance, image and video analysis, human vision and cognition, and computer vision.

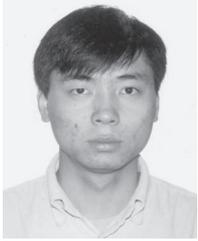
Dr. Huang is a member of the IEEE SMC Technical Committee on Cognitive Computing. He has published about 50 papers on the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS; the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY; *Pattern Recognition*; *Computer Vision and Image Understanding*; the International Conference on Image Processing; and the IEEE Conference on Computer Vision and Pattern Recognition. He was the Deputy Secretary-General of the IEEE Beijing Section from 2006 to 2008. He has been the Chair or a program committee member of more than 30 international conferences and workshops.

Xuelong Li (M'02–SM'07) is currently a Researcher (Full Professor) with the State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.



Jie Yang was born in Shanghai, China, in August 1964. He received the B.S. degree in automatic control and the M.S. degree in pattern recognition and intelligent system from Shanghai Jiao Tong University, Shanghai, in 1985 and 1988, respectively, and the Ph.D. degree from the University of Hamburg, Hamburg, Germany, in 1994.

He is currently a Professor and the Director of Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. He is the Principal Investigator of more than 30 nation and ministry scientific research projects in computer vision, pattern recognition, data mining, and artificial intelligence, including two 973 projects, three national 863 projects, three national nature fund projects, and three international cooperative projects with France, Korea, and Japan. He has published more than three hundred articles in national or international academic journals and conferences. Up to now, he has guided two postdoctoral, 26 doctors and 34 masters, awarded three researcher achievement prizes from the Ministry of Education, China and Shanghai municipality.



Dacheng Tao (M'07) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Phil. degree from the Chinese University of Hong Kong, Hong Kong, and the Ph.D. degree from the University of London, London, U.K.

He is currently a Nanyang Assistant Professor with the School of Computer Engineering, Nanyang Technological University, Singapore; a Visiting Research Fellow with the University of London; a Visiting Professor with Xidian University, Xi'an, China;

and a Guest Professor with Wuhan University, Wuhan, China. He is an Associate Professor for *Neurocomputing* (Elsevier) and the *Official Journal of the International Association for Statistical Computing-Computational Statistics and Data Analysis* (Elsevier). He has authored or edited six books and eight journal special issues. His research interests include the application of statistics and mathematics to data analysis problems in computer vision, data mining, machine learning, and visual surveillance.

Prof. Tao is an Associate Editor for the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING. He has published more than

100 scientific papers in journals and conference proceedings, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (TKDE), the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), the European Conference on Computer Vision, the International Conference on Data Mining (ICDM), the *ACM Transactions on Knowledge Discovery from Data, Multimedia, Data Mining and Knowledge Discovery in Databases* (KDD), etc. He has cochaired special sessions, invited sessions, workshops, panels, and conferences. He has served on more than 80 major international conferences, including CVPR, the IEEE International Conference on Computer Vision, the European Conference on Computer Vision, ICDM, KDD, and *Multimedia*, and on more than 30 prestigious international journals, including TPAMI, TKDE, the *ACM Transactions on Information Systems*, and TIP. He was the recipient of best paper runner-up awards and finalist awards, several Meritorious Awards from the International Interdisciplinary Contest in Modeling (which is the highest level mathematical modeling contest in the world and is organized by COMAP), and the K. C. WONG Education Foundation Award. He has been interviewed by ScienceWatch.com (Thomson Scientific) for one of his papers in the TPAMI.