Multilinear Discriminant Analysis for Face Recognition

Shuicheng Yan, *Member, IEEE*, Dong Xu, Qiang Yang, *Senior Member, IEEE*, Lei Zhang, *Member, IEEE*, Xiaoou Tang, *Senior Member, IEEE*, and Hong-Jiang Zhang, *Fellow, IEEE*

***Abstract*—There is a growing interest in subspace learning techniques for face recognition; however, the excessive dimension of the data space often brings the algorithms into the curse of dimensionality dilemma. In this paper, we present a novel approach to solve the supervised dimensionality reduction problem by encoding an image object as a general tensor of second or even higher order. First, we propose a discriminant tensor criterion, whereby multiple interrelated lower dimensional discriminative subspaces are derived for feature extraction. Then, a novel approach, called k-mode optimization, is presented to iteratively learn these subspaces by unfolding the tensor along different tensor directions. We call this algorithm multilinear discriminant analysis (MDA), which has the following characteristics: 1) multiple interrelated subspaces can collaborate to discriminate different classes, 2) for classification problems involving higher order tensors, the MDA algorithm can avoid the curse of dimensionality dilemma and alleviate the small sample size problem, and 3) the computational cost in the learning stage is reduced to a large extent owing to the reduced data dimensions in k-mode optimization. We provide extensive experiments on ORL, CMU PIE, and FERET databases by encoding face images as second- or third-order tensors to demonstrate that the proposed MDA algorithm based on higher order tensors has the potential to outperform the traditional vector-based subspace learning algorithms, especially in the cases with small sample sizes.**

***Index Terms*—2-D LDA, 2-D PCA, linear discriminant analysis (LDA), multilinear algebra, principal component analysis (PCA), subspace learning.**

I. Introduction

SUBSPACE learning is an important direction in computer vision research [3], [9], [12], [13], [29]. Most traditional algorithms, such as the *principal component analysis* (PCA) [17], [20], [21], [24] and *linear discriminant analysis* (LDA) [1], [2], [11], [15], [22], [26], input an image object as a 1-D vector [4], [5]. Some recent works, however, have started to consider an object as a 2-D matrix for subspace learning. Yang *et al.* [25] and Liu *et al.* [10] proposed two methods, named 2-D PCA and 2-D LDA[1] to conduct PCA and LDA respectively by simply replacing the image vector with image matrix in computing the corresponding variance matrices. However, the intuitive meaning of learned eigenvectors is still unclear as stated in [25]. Kong *et al.* [7] proposed an algorithm called 2-D FDA, with the similar idea as 2-D LDA, for dimensionality reduction and showed its advantages in small sample size problem. Ye *et al.* [27], [28] presented more general forms of 2-D PCA and 2-D LDA by simultaneously computing two subspaces. Shashua *et al.* [18] also considered the image as a matrix and searched for the best tensor-rank approximation to the original third-order image tensor. Similar to PCA, it is an unsupervised learning algorithm, thus, not always optimal for classification tasks. These recent approaches beg the question of whether it is possible to gain even more in supervised or unsupervised learning by taking into account the representation of higher order tensors. In this paper, we give a positive answer to this question.

Our observation is as follows. In the real world, the extracted feature of an object often has some specialized structures and such structures are in the form of second or even higher order tensors [23]. For example, this is the case when a captured image is a second-order tensor, *i.e.,* matrix, and when the sequence data such as a video for event analysis, is in the form of third-order tensor. It would be desirable to uncover the underlying structures in these problems for data analysis. However, most previous work on dimensionality reduction and classification would first transform the input image data into a 1-D vector, which ignores the underlying data structure and often leads to the *curse of dimensionality* dilemma and the *small sample size problem*. In this paper, we investigate how to conduct discriminant analysis by encoding an object as a general tensor of second or higher order. Also, we explore the characteristics of the higher order tensor-based discriminant analysis in theory. We will demonstrate that this analysis allows us to alleviate the above two problems when using the vector representation.

More specifically, our contributions are as follows. First, we propose a novel criterion for dimensionality reduction, called *discriminant tensor criterion* (DTC) which maximizes the interclass scatter and at the same time minimizes the intraclass

Manuscript received February 16, 2005; revised May 7, 2006. This work was performed when S. Yan was a visiting student at Microsoft Research Asia. This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region and Q. Yang was supported by a Grant from Hong Kong RGC (RGC Project HKUST 6187/04E). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhigang (Zeke) Fan.

S. Yan and X. Tang are with Microsoft Research Asia, Beijing 100080, China, and also with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: scyan@ie.cuhk.edu.hk; xtang@ie.cuhk.edu.hk).

D. Xu is with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: dongxu@ee.columbia.edu).

Q. Yang is with the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong (e-mail: qyang@cse.ust.hk).

L. Zhang and H.-J. Zhang are with Microsoft Research Asia, Beijing 100080, China (e-mail: leizhang@microsoft.com; leizhang@microsoft.com).

Color versions of Figs. 6–9 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2006.884929

[1]Note that the algorithm in [10] is not called 2-D LDA; we call it 2-D LDA here because it shares the same framework as 2-D PCA.

scatter both measured in the tensor-based metric. Different from the traditional subspace learning criterion which derives only *one* subspace, in our approach *multiple* interrelated subspaces are obtained through the optimization of the criterion where the number of the subspaces is determined by the order of the feature tensor used.

Second, we present a procedure to iteratively learn these interrelated discriminant subspaces via a novel tensor analysis approach, called *k-mode optimization* approach. We explore the foundation of the *k-mode optimization* approach to show that it unfolds the tensors into matrices along the kth direction. When the column vectors of the unfolded matrices are considered as the new objects to be analyzed, a special discriminant analysis is performed by computing the scatter as the sums of the scatter computed from the new samples with the same column indices. This explanation provides an intuitive explanation for the superiority of our proposed algorithm in comparison with other vector-based approaches.

We summarize the advantages of our algorithm, *multilinear discriminant analysis* (MDA), as follows.

1) MDA is a general supervised dimensionality reduction framework. It can avoid the *curse of dimensionality dilemma* by using higher order tensors and *k-mode optimization* approach, because the latter is performed in a much lower-dimension feature space than the traditional vector-based methods, such as LDA, do.
2) MDA also helps alleviate the *small sample size problem*. As explained later, in the *k-mode optimization*, the sample size is effectively multiplied by a large scale.
3) Many more feature dimensions are available in MDA than in LDA because the available feature dimension of LDA is theoretically limited by the number of classes in the data, whereas the MDA is not.
4) The computational cost can be reduced to a large extent as the k-mode optimization in each step is performed on a feature space of smaller size.
5) The extension from vector to tensor for data representation and feature extraction is general, and it can also be applied in SVM and many other algorithms to improve algorithmic learnability and effectiveness.

As a result of all the above characteristics, we expect MDA to be a natural alternative to LDA algorithm and a more general algorithm for the pattern classification problems in image analysis where an object can be encoded in tensor representation.

The rest of the paper is organized as follows. In Section II, we introduce the *DTC* and its iterative solution procedure. In Section III, we justify the procedure and analyze the characteristics of the proposed algorithm. Then, in Section IV, we present the extensive face recognition experiments by encoding the image objects as second- or third-order tensors and compare the results to traditional subspace learning algorithms. Finally, in Section V, we conclude the paper with future work discussions.

II. Multilinear Discriminant Analysis

Most previous approaches to subspace learning, such as the popular PCA and LDA, consider an object as a 1-D vector. The corresponding learning algorithms are performed on a very high dimension feature space. As a result, these methods often suffer from the problem of *curse of dimensionality*. On a close examination, however, we have found that most objects in computer vision are more naturally represented as second- or higher order tensors. For example, the image matrix in Fig. 1(a) is a second-order tensor and the filtered Gabor image in Fig. 1(b) is a third-order tensor. In this work, we study how to conduct discriminant analysis in the general case that the objects are represented as tensors of second or higher order.

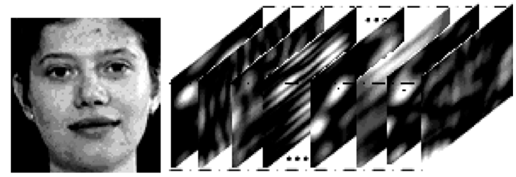


Fig. 1. Tensor representation examples: Second- and third-order object representations.

A. Discriminant Tensor Criterion

In this paper, the bold uppercase symbols represent tensor objects, such as $\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{Y}$; the normal uppercase symbols represent matrices, such as U, S; the italic lowercase symbols represent vectors, such as x, y; and the normal lowercase symbols represent scale numbers, such as $\mathrm{a}, \mathrm{b}, \mathrm{c}$. Assume that the training samples are represented as the nth-order tensors $\{\mathbf{X}_i \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_n}, i = 1, \ldots, N\}$, and $\mathbf{X}_i$ belongs to the class indexed as $c_i \in \{1, 2, \ldots, N_c\}$. Consequently, the sample set can be represented as an $(n+1)$th-order sample tensor $\tilde{\mathbf{X}} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_n \times N}$.

Before describing the *DTC*, we review the terminologies on tensor operations [6], [8], [23]. The inner product of two tensors $\mathbf{A}$ and $\mathbf{B}$ of the same dimensions is defined as $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i_1=1,\ldots,i_n=1}^{m_1,\ldots,m_n} \mathbf{A}_{i_1,\ldots,i_n} \mathbf{B}_{i_1,\ldots,i_n}$; the norm of a tensor $\mathbf{A}$ is defined as $\|\mathbf{A}\| = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}$, and the distance between tensors $\mathbf{A}$ and $\mathbf{B}$ are defined as $D(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|$. In the second-order tensor case, *i.e.,* matrix-form, the norm is called Frobenius norm and is written as $\|A\|_F$. The *k-mode* product of a tensor $\mathbf{A}$ and a matrix $U \in \mathbb{R}^{m_k \times m'_k}$ is defined as $\mathbf{B} = \mathbf{A} \times_k U$, where

$$\mathbf{B}_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_n} = \sum_{i=1}^{m_k} \mathbf{A}_{i_1,\ldots,i_{k-1},i,i_{k+1},\ldots,i_n} * U_{i,j}, j = 1, \ldots, m'_k.$$

The *DTC* is designed to pursue multiple interrelated projection matrices, *i.e.,* subspaces, which maximize the interclass scatter and at the same time minimize the intraclass scatter measured in tensor metric described above. That is

$$(U_k^*|_{k=1}^n) = \arg\max_{U_k|_{k=1}^n} \frac{\sum_c n_c \|\overline{\mathbf{X}}_c \times_1 U_1 \cdots \times_n U_n - \overline{\mathbf{X}} \times_1 U_1 \cdots \times_n U_n\|^2}{\sum_i \|\mathbf{X}_i \times_1 U_1 \cdots \times_n U_n - \overline{\mathbf{X}}_{c_i} \times_1 U_1 \cdots \times_n U_n\|^2} \tag{1}$$

where $\overline{\mathbf{X}}_c$ is the average tensor of the samples belonging to class c, $\overline{\mathbf{X}}$ is the total average tensor of all the samples, and n_c is

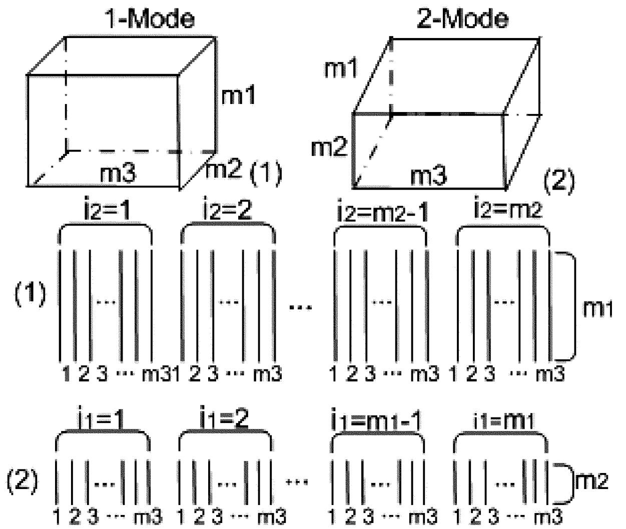


Fig. 2. Illustration of the k-mode unfolding of a third -order tensor.

sample number of class c. Similar to the Fisher criterion, the interclass scatter is measured by the sum of the weighted distances between the class center tensors $\overline{\mathbf{X}}_c$ and total sample center tensor $\overline{\mathbf{X}}$; meanwhile, the intraclass scatter is measured by the sum of the distances between each sample to its corresponding class center tensor. Despite the similarity, the data representation and metric are different between these two criterions.

Equation (1) is equivalent to a higher order nonlinear optimization problem with a higher order nonlinear constraint; thus, it is difficult to find a closed-form solution. Alternatively, we search for an iterative optimization approach to derive the interrelated discriminative subspaces.

B. k-Mode Optimization

We now discuss how to optimize the objective function from only one direction of the tensor, *i.e.,*

$$(U_k^*) = \arg\max_{U_k} \frac{\sum_c n_c \|\overline{\mathbf{X}}_c \times_k U_k - \overline{\mathbf{X}} \times_k U_k\|^2}{\sum_i \|\mathbf{X}_i \times_k U_k - \overline{\mathbf{X}}_{c_i} \times_k U_k\|^2}. \quad (2)$$

Before this analysis, we introduce the conception of *k-mode unfolding of a tensor*. Fig. 2 demonstrates two ways to unfold a third-order tensor. In the 1-mode version, a tensor is unfolded into a matrix along the i_1 axis, and the matrix width direction is indexed by searching i_2 index and i_3 index iteratively. For the 2-mode version, the tensor is unfolded along the i_2 axis. This process can be extended to the general nth-order tensor.

Formally, the k-mode unfolding of a tensor into a matrix is defined as

$$F^k \in \mathbb{R}^{m_k \times \prod_{i \neq k} m_i} \Leftarrow_k \mathbf{X} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_n} \text{ with}$$
$$F^k_{i_k, j} = \mathbf{X}_{i_1,\ldots,i_n}, j = 1 + \sum_{l=1, l\neq k}^{n} (i_l - 1) \prod_{o=l+1, o\neq k}^{n} m_o. \quad (3)$$

The problem in (2) is a special discriminant analysis problem. It can be understood in two steps: 1) the sample tensors are unfolded into matrices in the k-mode; 2) the column vector of the unfolded matrices is considered as the new object with the same class label as the original sample tensor and then the scatter matter for discriminant analysis is the sum of the scatter computed from the new samples with the same column index in the unfolded matrices. Theorem 1 below gives the details.

Theorem-1: The optimization problem in (2) can be reformulated as a special discriminant analysis problem as follows:

$$U_k^* = \arg\max_{U_k} \frac{Tr(U_k^{\mathrm{T}} S_B U_k)}{Tr(U_k^{\mathrm{T}} S_W U_k)}$$
$$S_B = \sum_{j=1}^{\prod_{o\neq k} m_o} S_B^j, S_B^j = \sum_{c=1}^{N_c} n_c (\overline{X}_c^{k,j} - \overline{X}^{k,j})(\overline{X}_c^{k,j} - \overline{X}^{k,j})^{\mathrm{T}}$$
$$S_W = \sum_{j=1}^{\prod_{o\neq k} m_o} S_W^j, S_W^j = \sum_{i=1}^{N} (X_i^{k,j} - \overline{X}_{c_i}^{k,j})(X_i^{k,j} - \overline{X}_{c_i}^{k,j})^{\mathrm{T}} \quad (4)$$

where, for the ease of presentation, $X_i^{k,j}$ represents the jth column vector of matrix X_i^k which is the k-mode unfolded matrix from sample tensor $\mathbf{X}_i$. $\overline{X}_c^{k,j}$ and $\overline{X}^{k,j}$ are defined in the same way as $X_i^{k,j}$ with respect to tensors $\overline{\mathbf{X}}_c$ and $\overline{\mathbf{X}}$.

Proof: Here, we take S_W as an example to prove the theorem. With simple algebraic computation, we can obtain $\|\mathbf{X} \times_k U\| = \|X^{k\mathrm{T}} U\|_F$, where X^k is the k-mode unfolding of the tensor $\mathbf{X}$; then, we have

$$\begin{aligned}
&\sum_i \|\mathbf{X}_i \times_k U_k - \overline{\mathbf{X}}_{c_i} \times_k U_k\|^2 \\
&= \sum_i \|X_i^{k\mathrm{T}} U_k - \overline{X_{c_i}^k}^{\mathrm{T}} U_k\|_F^2 \\
&= \sum_i Tr[U_k^{\mathrm{T}} (X_i^k - \overline{X}_{c_i}^k)(X_i^k - \overline{X}_{c_i}^k)^{\mathrm{T}} U_k] \\
&= Tr\left[U_k^{\mathrm{T}} \left(\sum_i (X_i^k - \overline{X}_{c_i}^k)(X_i^k - \overline{X}_{c_i}^k)^{\mathrm{T}}\right) U_k\right] \\
&= Tr\left[U_k^{\mathrm{T}} \left(\sum_{j=1}^{\prod_{o\neq k} m_o} \sum_i (X_i^{k,j} - \overline{X}_{c_i}^{k,j})(X_i^{k,j} - \overline{X}_{c_i}^{k,j})^{\mathrm{T}}\right) U_k\right] \\
&= Tr(U_k^{\mathrm{T}} S_W U_k).
\end{aligned}$$

Similarly, $\sum_c n_c \|\overline{\mathbf{X}}_c \times_k U_k - \overline{\mathbf{X}} \times_k U_k\|^2 = Tr(U_k^{\mathrm{T}} S_B^k U_k)$.

Therefore, the optimization problem in (2) can be reformulated as a special discriminant analysis problem, and it can be solved in the same way for the traditional LDA algorithm [1], [3].

C. Multilinear Discriminant Analysis

As described above, the DTC has no closed-form solution. In response to this issue, we present an iterative procedure to solve the problem. In each iteration, $U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_n$ are assumed known, then the DTC is changed to

$$U_k^* = \arg\max_{U_k} \frac{\sum_c n_c \|\overline{\mathbf{X}}_c \times_1 U_1 \cdots \times_n U_n - \overline{\mathbf{X}} \times_1 U_1 \cdots \times_n U_n\|^2}{\sum_i \|\mathbf{X}_i \times_1 U_1 \cdots \times_n U_n - \overline{\mathbf{X}}_{c_i} \times_1 U_1 \cdots \times_n U_n\|^2}. \quad (5)$$

Multilinear Discriminant Analysis:

Given the sample set $\tilde{\mathbf{X}} \in \mathbb{R}^{m_1 \times m_2 \times \ldots \times m_n \times N}$, their class labels $c_i \in \{1, 2, \ldots, N_c\}$, and the final lower dimensions $m'_1 \times m'_2 \times \ldots \times m'_n$.

1. Initialize $U_1^0 = I_{m_1}, U_2^0 = I_{m_2}, \ldots, U_n^0 = I_{m_n}$;
2. For $t = 1, 2, \ldots, T_{max}$ do
   a) For $k = 1, 2, \ldots, n$ do

   $\mathbf{Y}_i = \mathbf{X}_i \times_1 U_1^t \ldots \times_{k-1} U_{k-1}^t \times_{k+1} U_{k+1}^{t-1} \ldots \times_n U_n^{t-1}$

   $Y_i^k \Leftarrow_k \mathbf{Y}_i$

   $S_B = \sum_{j=1}^{\prod_{o \neq k} m_o} S_B^j, S_B^j = \sum_{c=1}^{N_c} n_c (\bar{Y}_c^{k,j} - \bar{Y}^{k,j})(\bar{Y}_c^{k,j} - \bar{Y}^{k,j})^{\mathrm{T}}$

   $S_W = \sum_{j=1}^{\prod_{o \neq k} m_o} S_W^j, S_W^j = \sum_{i=1}^{N} (Y_i^{k,j} - \bar{Y}_{c_i}^{k,j})(Y_i^{k,j} - \bar{Y}_{c_i}^{k,j})^{\mathrm{T}}$

   $S_B U_k^t = S_W U_k^t \Lambda_k, \quad U_k^t \in \mathbb{R}^{m_k \times m'_k}$

   b) If $t > 2$ and $\| U_k^t - U_k^{t-1} \| < m'_k m_k \varepsilon, k = 1, \ldots, n$, break;
3. Output the projections $U_k = U_k^t \in \mathbb{R}^{m_k \times m'_k}, k = 1, \ldots, n$.

Fig. 3. Procedure for MDA.

Denote $\mathbf{Y}_i = \mathbf{X}_i \times_1 U_1 \cdots \times_{k-1} U_{k-1} \times_{k+1} U_{k+1} \cdots \times_n U_n$, then

$$U_k^* = \arg\max_{U_k} \frac{\sum_c n_c \|\overline{\mathbf{Y}}_c \times_k U_k - \overline{\mathbf{Y}} \times_k U_k\|^2}{\sum_i \|\mathbf{Y}_i \times_k U_k - \overline{\mathbf{Y}}_{c_i} \times_k U_k\|^2}. \tag{6}$$

It has the same formulation as (2) by replacing $\mathbf{X}_i$ with $\mathbf{Y}_i$; thus, it can be solved using the above described *k-mode optimization* approach. Therefore, the projection matrices can be iteratively optimized, and the entire procedure to optimize the *DTC* is listed in Fig. 3.

D. Classification With Multilinear Discriminant Analysis

With the learned projection matrices $(U_k^*|_{k=1}^n)$, the low-dimensional representation of the training sample $\mathbf{X}_i$, $i = 1, \ldots, N$, can be computed as $\mathbf{Z}_i = \mathbf{X}_i \times_1 U_1 \times_2 U_2 \cdots \times_n U_n$. When a new data $\mathbf{X}$ comes, we first compute its low-dimensional representation as

$$\mathbf{Z} = \mathbf{X} \times_1 U_1 \times_2 U_2 \cdots \times_n U_n.$$

Then its class label is predicted to be that of the sample whose low-dimensional representation is nearest to $\mathbf{Z}$, that is

$$i^* = \arg\min_i \|\mathbf{Z}_i - \mathbf{Z}\|$$

and then the sample $\mathbf{X}$ is classified to the class c_{i^*}. In this paper, we use this method for final classification in all the experiments due to its simplicity in computation.

III. Algorithmic Analysis and Discusses

In this section, we introduce the merits of our proposed procedure MDA in terms of learnability and time complexity and also discuss its relationship with LDA, 2-D LDA [10], Tensorface [23], and Shasua's work [18].

Algorithmic Analysis

1) Singularity and Curse of Dimensionality: In LDA, the size of the scatter matrix is $\prod_{k=1}^n m_k \times \prod_{k=1}^n m_k$ if a tensor is transformed into a vector. It is often the case that $N - N_c < \prod_{k=1}^n m_k$ for a moderate data set. Thus, in many cases, the intraclass scatter matrix is singular; thus, the accuracy and robustness of the solution are often degraded. For most pattern recognition problems, $\prod_{k=1}^n m_k$ is very large, and, hence, to train a credible classifier requires a huge number of training samples for the learnability of LDA. In MDA, however, the step-wise intraclass scatter matrix is in size of $m_k \times m_k$, which is much smaller than that of LDA. As described in Section II, the objects to be analyzed in MDA are the column vectors of the unfolded matrices and the sample number can be considered to be enlarged to $\prod_{i \neq k} m_i * N$. In most cases, $\prod_{i \neq k} m_i * N > m_k$ can be satisfied; therefore, there is far less singularity problem in MDA when based on higher order tensors. Moreover, the number m_k is much smaller than $\prod_{k=1}^n m_k$, so the *curse of dimensionality* dilemma is reduced to a large extent.

2) Available Projection Directions: The most important factor limiting the application of LDA is that the available dimension for pattern recognition has an upper bound $N_c - 1$. Although many approaches have been proposed to utilize the null space of the intraclass scatter matrix, the intrinsic dimension cannot be larger than $N_c - 1$. In our proposed MDA algorithm, the largest number of the available dimensions for each subspace can be obtained through the following theorem.

Theorem 2: The largest number of the available dimension is $\min\{m_k, (N_c - 1) \prod_{i \neq k} m_i\}$ for MDA in each step.

Proof: As in (4), $S_B = \sum_{j=1}^{\prod_{o \neq k} m_o} S_B^j$, then $rank(S_B) \leq \sum_{j=1}^{\prod_{i \neq k} m_i} rank(S_B^j) \leq (N_c - 1) \prod_{i \neq k} m_i$; and, at the same time, $rank(S_B) \leq m_k$ with the equality satisfied when all the S_B is in full rank. So, the largest number of the available dimension is $\min\{m_k, (N_c - 1) \prod_{i \neq k} m_i\}$. ■

Moreover, there are n projection matrices; thus, there are far more projection directions for dimensionality reduction in MDA, and it provides discriminating capability for much more features.

3) Computational Cost: For ease of understanding, let us assume that the sample tensor has uniform dimension numbers for all directions, i.e., $m_i = m, i = 1, \ldots, n$. Therefore, the complexity of LDA is $\mathrm{O(m^{3n})}$, while in MDA, the complexity to compute the scatter matrices is $\mathrm{O(n * m^{n+1})}$ and complexity for k-mode optimization is $\mathrm{O(n{*}m^3)}$ for each loop, which is much lower than that of LDA. Although MDA has no closed-form solution and many loops are required to achieve convergence, it is still much faster than LDA owing to its simplicity in each iteration.

A. Discusses

1) Connections to LDA and 2-D LDA: LDA and 2-D LDA [25] both optimize the so-called Fisher criterion

$$w^* = \arg\max_w \frac{Tr(w^{\mathrm{T}} S_B w)}{Tr(w^{\mathrm{T}} S_W w)}. \tag{7}$$

In LDA, the sample data are represented as vectors $\{x_i \in \mathbb{R}^m, i = 1, \ldots, N\}$ and the scatter matrices are

$$S_B = \sum_{c=1}^{N_c} n_c(\overline{x}_c - \overline{x})(\overline{x}_c - \overline{x})^{\mathrm{T}}, S_w = \sum_{i=1}^{N} (x_i - \overline{x}_{c_i})(x_i - \overline{x}_{c_i})^{\mathrm{T}}.$$

In 2-D LDA, the sample data are matrices represented as $\{X_i \in \mathbb{R}^{m_1 \times m_2}, i = 1, \ldots, N\}$ and the scatter matrices are computed by replacing the vectors in (7) as matrices and

$$S_B = \sum_{c=1}^{N_c} n_c(\overline{X}_c - \overline{X})^{\mathrm{T}}(\overline{X}_c - \overline{X})$$
$$S_w = \sum_{i=1}^{N} (X_i - \overline{X}_{c_i})^{\mathrm{T}}(X_i - \overline{X}_{c_i}).$$

In both cases, the averages are defined in the same way as the case with tensor representation. Actually, with simple algebraic computation, LDA and 2-D LDA are both special formulations of our proposed MDA: LDA can be reformulated as a special case of MDA with $n = 1$, while 2-D LDA can be reformulated as a special case of MDA with $n = 2$ and using only one subspace instead of two in regular MDA with $n = 2$.

2) Relationship With Tensorface: Tensorface [23] is a recently proposed algorithm for dimensionality reduction. It also utilizes the higher order statistics for data analysis as in MDA. However, in Tensorface, an image is still treated as a vector and the image ensembles of all persons under all predefined poses, illuminations and expressions are treated as a high-order tensor; while in MDA, an image object is directly treated as a matrix and the whole data set is encoded as a third-order tensor. The difference in tensor composition makes these two algorithms different in many aspects: 1) the semantics of the learned subspaces are different. In Tensorface, the learned subspaces characterize the variations from external factors including pose, illumination, and expression; whereas in MDA, these subspaces characterize the discriminating information from internal factors such as row and column directions. 2) In Tensorface, the image object is treated as a vector and usually the dimension is extremely high; thus, it may suffer from the curse of dimensionality dilemma. However, as described beforehand, MDA can effectively overcome this issue. 3) They are superior to each other in different aspects: Tensorface is good at classifying faces with different pose, illumination or expression; while MDA works well in the small sample size problem. As the above differences between Tensorface and MDA algorithm, we do not further compare them in the experiment section.

3) Relationship With Shasua's Work: The general idea of Shasua's work [18] is to consider the collection of sample matrices as a third-order tensor and search for an approximation of its tensor-rank. It is different from the MDA algorithm in the following aspects: 1) Shasua's work aims to approximate the sample matrices with a set of rank-one matrices; while MDA does not target at data reconstruction. 2) In Shasua's work, the rank-one matrices are learned one by one; while in MDA, the projection matrices are optimized iteratively. 3) Shasua's work is unsupervised and, hence, not always optimal for classification task; while MDA is supervised, and the data structure along with the label information are both effectively utilized.

Fig. 4. Ten samples in the ORL face database.

IV. Experiments

In this section, three benchmark face databases, ORL [14], FERET [16], and CMU PIE [19] were used to evaluate the effectiveness of our proposed algorithm, *MDA*, in face recognition accuracy. Our proposed algorithm is referred to as *MDA/2-2* and *MDA/3-3* for problems with tensor of second and third order, respectively, where the first number (the first 2 in 2-2) refers to the tensor order and the second number means the number of subspaces used.

These algorithms were compared with the popular Eigenface, Fisherface, and the 2-D LDA algorithms. The 2-D LDA algorithm has been proved to be special MDA using a single subspace, thus, is referred to as MDA/2-1 in this work. In order to compare with the Fisherface fairly, we also report the best result on different feature dimensions in the LDA step, which is referred to as the symbol *O* after Fisherface in all results.

In all the experiments, the gallery and probe data were both transformed into lower dimensional tensors or vectors via the learned subspaces, and the nearest neighbor classifier was used for final classification. The experiments were conducted by encoding the face images in different ways, *i.e.,* vector, matrix, and the filtered Gabor tensor. Moreover, the performances on the cases with different number of training samples were also evaluated to demonstrate their robustness in the small sample size problems.

A. ORL Database — MDA/2-2

The ORL database contains 400 images of 40 individuals. These images were captured at different times and have different variations including expression (open or closed eyes, smiling or nonsmiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20°. All images were in grayscale and normalized to the resolution of $112 * 92$ pixels and histogram equilibrium was applied in the preprocessing step. Ten sample images of one person in the ORL database after the scale normalization are displayed in Fig. 4.

Four sets of experiments were conducted to compare the performance of MDA/2-2 with Eigenface, Fisherface/*O*, and MDA/2-1. In each experiment, the image set was partitioned into the gallery and probe set with different numbers. For ease of representation, the experiments are named as Gm/Pn which means that m images per person are *randomly* selected for training and the remaining n images for testing.

Table I shows the best face recognition accuracies of all the algorithms in our experiments with different gallery and probe set partitions. The comparative results show that MDA/2-2 outperforms Eigenface, Fisherface/*O*, and MDA/2-1 on all four sets

TABLE I
RECOGNITION ACCURACY (%) COMPARISON OF MDA/2-2, EIGENFACE, FISHERFACE/*O*, AND MDA/2-1ON ORL DATABASE

	G5/P5	G4/P6	G3/P7	G2/P8
Eigenface	97.5	91.3	87.7	81.6
Fisherface/*O*	92.0/ 92.0	85.4/ 85.8	87.5/ 87.5	74.7/ 79.7
MDA/2-1	98.5	96.3	94.3	88.1
MDA/2-2	99.0	98.3	95.0	90.1

Fig. 5. Ten images of one person in PIE1 database.

of experiments, especially in the cases with a small number of training samples.

B. PIE Database — MDA/3-3 and MDA/2-2

The CMU Pose, Illumination, and Expression (PIE) database contains more than 40 000 facial images of 68 people. The images were acquired over different poses, under variable illumination conditions and with different facial expressions. In this experiment, two subdatabases were used to evaluate our proposed algorithms.

In the first subdatabase, referred to as PIE1, five near frontal poses (C27, C05, C29, C09, and C07) and illumination indexed as 08 and 11 were used. Each person has ten images and all the images were aligned by fixing the locations of two eyes, and normalized to $64 * 64$ pixels. Similar to the previous experiments, histogram equilibrium was applied in the preprocessing step. Fig. 5 shows ten examples of one person.

The data set was randomly partitioned into gallery and probe sets; and two samples each person was used for training. We extracted the 40 Gabor features with five different scales and eight different directions in the down-sampled positions and each image is encoded as a third-order tensor of size $16{*}16{*}40$. Table II shows the detailed face recognition accuracies. The results clearly demonstrate that MDA/3-3 is superior to all other algorithms. Moreover, it shows that the Gabor feature can help improve the face recognition accuracy in both Eigenface and Fisherface/*O*. For a detailed illustration on the face recognition rates with different feature numbers, Figs. 6–8 plot the recognition rates of Eigenface, 2-D LDA, and MDA/2-2 when using different number of low-dimensional features in the G4/P6 experiment of the PIE1 subset. Meanwhile, we test the stability of MDA with respect to two factors: one is the number of iterations and another is to initiate $U_1 = I_{m_1}$ first or $U_2 = I_{m_2}$ first. The results from G3/P7 experiment on PIE1 subset as plotted in Fig. 9 show that the face recognition rate is stable with respect to different number of iterations; and the recognition rates are also stable to initiate $U_1 = I_{m_1}$ first or $U_2 = I_{m_2}$ first.

Another subdatabase PIE2 consists of the same five poses as in PIE1, but the illumination indexed as 10 and 13 were also used. Therefore, the PIE2 database is more difficult for classification. We conducted three sets of experiments on this subdatabase. Table III lists all the comparative experimental results

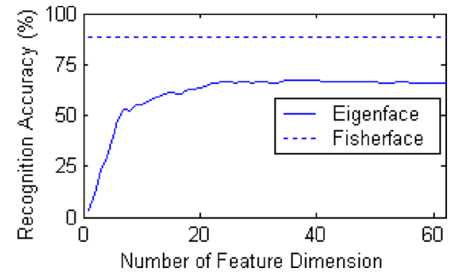


Fig. 6. Recognition accuracies (%) of Eigenface and Fisherface versus number of features on PIE1 database.

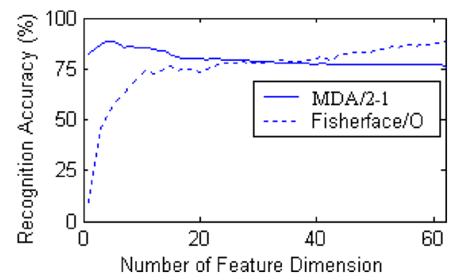


Fig. 7. Recognition accuracies (%) of MDA/2-1 and Fisherface/*O* versus number of features on PIE1 database.

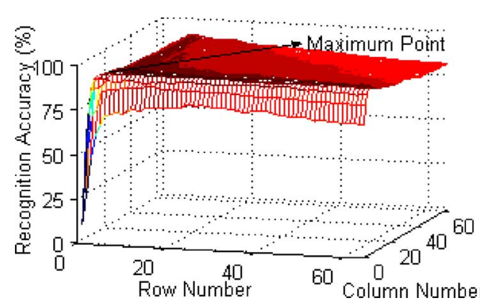


Fig. 8. Recognition accuracies (%) of MDA/2-2 versus numbers of features along the row and column directions, respectively, on PIE1 database.

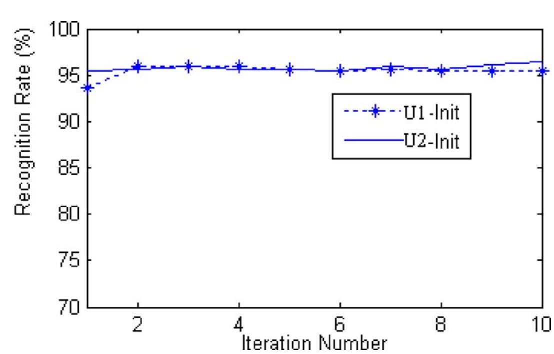


Fig. 9. Face recognition rates of MDA/2-2 versus Iteration numbers on experiment G3/P7 of the PIE1 subset. Note that the legend *U1-Init* means that we initiate $U_1 = I_{m_1}$ first in the MDA algorithm, and *U2-Init* means that we initiate $U_2 = I_{m_2}$ first in the MDA algorithm.

of the MDA/2-2, Eigenface, Fisherface/*O*, and MDA/2-1. The reconstruction-based Eigenface performs very poor in all the three cases; Fisherface is better than Eigenface, yet it also fails

TABLE II
RECOGNITION ACCURACY (%) COMPARISON OF EIGENFACE, FISHERFACE/*O*, AND MDA WITH TENSORS OF DIFFERENT ORDERS ON PIE-1 DATABASE

Algorithm	Accuracy
Eigenface (Grey)	56.9
Eigenface (Gabor)	70.6
Fisherface/*O* (Grey)	53.8/66.1
Fisherface/*O* (Gabor)	71.6/79.8
MDA/2-1 (Grey)	72.8
MDA/2-2 (Grey)	80.2
MDA/3-3 (Gabor)	82.9

TABLE III
RECOGNITION ACCURACY (%) COMPARISON OF MDA/2-2, EIGENFACE, FISHERFACE/*O*, AND MDA/2-1 ON THE PIE2 DATABASE

	G4/P6	G3/P7	G2/P8
Eigenface	38.9	28.3	26.6
Fisherface/*O*	79.9/80.2	65.3/65.8	38.1/47.6
MDA/2-1	74.3	71.9	63.5
MDA/2-2	82.3	80.7	66.7

TABLE IV
RECOGNITION ACCURACY (%) COMPARISON OF MDA/3-3, EIGENFACE, FISHERFACE/*O*, MDA/2-1 AND MDA/2-2 ON THE FERET DATABASE

Algorithm	Accuracies
Eigenface (Grey)	65.7
Eigenface (Gabor)	75.7
Fisherface/*O* (Grey)	69.3/74.3
Fisherface/*O* (Gabor)	73.9/76.1
MDA/2-1 (Grey)	73.5
MDA/2-2 (Grey)	80.4
MDA/3-3 (Gabor)	83.6

in the cases with only two training images for each person. In all the three experiments, MDA/2-2 performs the best.

C. Feret Database—MDA/3-3 and MDA/2-2

Two types of experiments were conducted on the FERET database. One is conducted on 70 people of the FERET database with six different images for each person; two of them were applied as gallery set and the other four for probe set. We extracted 40 Gabor features with five different scales and eight different directions in the down-sampled positions and each image was encoded as a third-order tensor of size $28*23*40$ for MDA/3-3.

We compared all the above mentioned algorithms on the FERET database. Table IV demonstrates the comparative face recognition accuracies. Similar to the results in the PIE1 subdatabase; it shows that the Gabor features can significantly improve the performance and MDA/3-3 consistently outperforms all the other algorithms.

Other types of experiments were conducted on the large-scale case, where the training CD [16] of the FERET databases was used for the training of the different algorithms, and then the *fa* and *fb* images of 1195 persons were used as the gallery and probe set respectively. The images are aligned by fixing the locations of the two eyes and resized to $54 * 48$ pixels. When the training image number is big enough, the PCA dimension $N - N_c$ is not always the best for Fisherface algorithm as reported in [5]; thus, we also report the results of the PCA dimensions by preserving 98% and 95% energies with the similar idea as in [5]. The experimental results are reported in Table V, and it shows that MDA/2-1 is worse than Fisherface in accuracy while MDA/2-2 still outperforms all the other algorithms, even the PCA dimension of Fisherface algorithm is tuned by different energy.

D. Discussions

From the experimental results listed in Tables I–IV and Figs. 6–9, we can have a lot of observations.

1) For all the cases directly based on original gray-level features, MDA/2-2 shows best among all the evaluated algorithms, which validates the effectiveness of the matrix, namely second-order tensor, representation in improving algorithmic learnability compared with vector representation.
2) MDA/3-3 consistently outperforms all other algorithms in all cases. The superiority of MDA/3-3 stems from two aspects: on the one hand, the third-order tensor representation is obtained from the Garbor features which are more robust compared with the original gray-level features; one the other hand, the curse of dimensionality and small sample size problem are greatly alleviated in MDA/3-3 as discussed in Section III-A, and, hence, the derived subspace potentially have greater discrimination capability compared with other vector-based algorithms.
3) MDA/3-3 and MDA/2-2 are very robust in the cases only with a small number of training samples, which again validates the advantage of MDA algorithms in alleviating the small sample size problem. In these cases, Eigenface almost fails to present acceptable results. Fisherface/*O* is a little better than Eigenface, but still much worse than MDA/2-2 and MDA/3-3.
4) MDA/2-1 is also robust in the cases with a small number of training samples and outperforms Fisherface/*O* in most cases. However, it is worse than MDA/2-2 and MDA/3-3 in face recognition accuracy in all the cases. It demonstrates that the collaboration of multiple subspaces can greatly enhance the classification capability.
5) As discussed in [1], LDA is not always superior to PCA, especially in the cases when the training set cannot well represent the data distribution. There are some cases in which Eigenface outperforms Fisherface, such as in the ORL database and the PIE1 database.
6) Many methods have been proposed to improve the performance of Fisherface. In this work, we only tested one way that explores the performances on all feature dimensions. We did not further evaluate the other methods because those methods, such as random subspace [22], can also be applied on MDA with tensors of higher order tensors.

V. CONCLUSION

In this paper, a novel algorithm, *MDA*, has been proposed for supervised dimensionality reduction with general tensor representation. In *MDA*, the image objects were encoded as an nth-order tensor. An approach called k*-mode optimization* was proposed to iteratively learn multiple interrelated discriminative subspaces for dimensionality reduction of the higher order tensor. Compared with traditional algorithms, such as PCA and LDA, our proposed algorithm effectively avoids the curse of dimensionality dilemma

TABLE V
RECOGNITION ACCURACY (%) COMPARISON OF EIGENFACE, FISHERFACE, MDA/2-1, AND MDA/2-2 ON FERET DATABASE FOR THE LARGE-SCALE CASE. NOTE THAT THE NUMBER IN THE BRACKET IS THE FEATURE DIMENSION WITH THE HIGHEST RECOGNITION RATE FOR EACH ALGORITHM

Eigenface	Fisherface	Fisherface(95% energy)	Fisherface(98% energy)	MDA/2-1	MDA/2-2
81.1 (590)	87.8	94.0	88.6	86.8(11*48)	94.8(11*12)

and alleviates the small sample size problem. Due to the low requirement on samples and the high performance in classification problem, *MDA* should be a general alternative of LDA algorithm for problems encoding objects as tensors. An interesting application of our proposed MDA algorithm is to apply MDA/4-4 for video-based face recognition and we are planning to explore this application in our future researches.

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] C. Fraley and A. Raftery, "Model-based clustering, discriminant analysis and density estimation," *J. Amer. Statist. Assoc.*, vol. 97, pp. 611–631, Jun. 2002.

[3] K. Fukunaga, *Statistical Pattern Recognition*. New York: Academic, 1990.

[4] X. He, S. Yan, Y. Hu, and H. Zhang, "Learning a locality preserving subspace for visual recognition," in *Proc. 9th IEEE Int. Conf. Computer Vision*, Oct. 13–16, 2003, pp. 385–392.

[5] X. He, S. Yan, Y. Hu, H. Liu, and H. Zhang, "Spectral analysis for face recognition," presented at the 6th Asian Conf. Computer Vision, Korea, Jan. 28–30, 2004.

[6] T. Kolda, "Orthogonal tensor decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 243–255.

[7] H. Kong, E. Teoh, J. Wang, and R. Venkateswarlu, "Two dimensional fisher discriminant analysis: Forget about small sample size problem," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, Mar. 19–23, 2005, pp. 761–764.

[8] L. Lathauwer, B. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278.

[9] C. Liu, H. Shum, and C. Zhang, "A two-step approach to hallucinating faces global parametric model and local nonparametric model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp. 192–198.

[10] K. Liu, Y. Cheng, and J. Yang, "Algebraic feature extraction for image recognition based on an optimal discriminant criterion," *Pattern Recognit.*, vol. 26, no. 6, pp. 903–911, Jun. 1993.

[11] Q. Liu, H. Lu, and S. Ma, "Improving kernel fisher discriminant analysis for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 42–49, Jan. 2004.

[12] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognit.*, vol. 33, no. 11, pp. 1771–1782, Nov. 2000.

[13] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, Jul. 1997.

[14] The Olivetti & Oracle Research Laboratory Face Database of Faces. [Online]. Available: http://www.cam-orl.co.uk/facedatabase.html

[15] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," presented at the IEEE Conf. Computer Vision and Pattern Recognition Seattle, WA, Jun. 1994.

[16] I. Philips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput.*, vol. 16, pp. 295–306, 1998.

[17] T. Shakunaga and K. Shigenari, "Decomposed eigenface for face recognition under various lighting conditions," presented at the IEEE Int. Conf. Computer Vision and Pattern Recognition, Dec. 11–13, 2001.

[18] A. Shashua and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," presented at the IEEE Conf. Computer Vision and Pattern Recognition, Dec. 2001.

[19] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," presented at the IEEE Int. Conf. Automatic Face and Gesture Recognition, May 2002.

[20] J. Tenenbaum and W. Freeman, "Separating style and content with bilinear models," *Neural Comput.*, vol. 12, no. 6, pp. 1247–1283, 2000.

[21] M. Turk and Pentland, "Face recognition using eigenfaces," presented at the IEEE Conf. Computer Vision and Pattern Recognition, Maui, HI, 1991.

[22] X. Wang and X. Tang, "Random sampling LDA for face recognition," presented at the IEEE Conf. Computer Vision and Pattern Recognition, Jun. 2004.

[23] M. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis for image ensembles," in *Proc. Computer Vision and Pattern Recognition*, Madison, WI, Jun. 2003, vol. 2, pp. 93–99.

[24] M. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using kernel methods," presented at the 5fth Int. Conf. Automatic Face and Gesture Recognition, Washington, DC, May 2002.

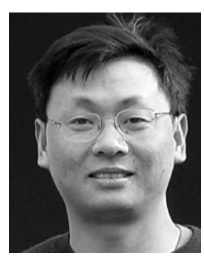
[25] J. Yang, D. Zhang, A. Frangi, and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[26] J. Yang, Y. Yu, and W. Kunz, "An efficient LDA algorithm for face recognition," presented at the 6th Int. Conf. Control, Automation, Robotics, and Vision, Singapore, 2000.

[27] J. Ye, "Generalized low rank approximations of matrices," presented at the ICML, 2004.

[28] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," *Adv. Neural Inf. Process. Syst.*, vol. 17, pp. 1569–1576, 2005.

[29] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical performance analysis of linear discriminant classifiers," in *Proc. IEEE. Conf. Computer Vision Pattern Recognition*, Santa Barbara, CA, Jun. 1998, pp. 164–169.

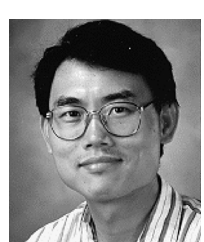


Shuicheng Yan (M'06) received the B.S. and Ph.D. degrees from the Applied Mathematics Department, School of Mathematical Sciences, Peking University, China, in 1999 and 2004, respectively.

His research interests include computer vision and machine learning.

Dong Xu received the B.S. and Ph.D. degrees from the Electronic Engineering and Information Science Department, University of Science and Technology of China, in 2001 and 2005, respectively.

His research interests include pattern recognition, computer vision, and machine learning.



Qiang Yang (M'02–SM'04) received the Ph.D. degree from the University of Maryland, College Park, in 1989.

He is a faculty member at the Department of Computer Science, Hong Kong University of Science and Technology. His research interests are AI planning, machine learning, case-based reasoning, and data mining.

Dr. Yang is an Associate Editor for the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING and *IEEE Intelligent Systems*.

Lei Zhang (M'03) received the B.S., M.S., and Ph.D. degrees in computer science from Tsinghua University, China, in 1993, 1995, and 2001, respectively.

He is a Researcher and a Project Leader with Microsoft Research Asia. In 2001, he joined the Media Computing Group at Microsoft Research Asia and was working on content-based image analysis, machine learning, and computer vision. In 2004, he joined the Web Search and Mining Group, where he started working on web search and information retrieval. His current research interests are search relevance ranking, web-scale image retrieval, and photo management and sharing.

Dr. Zhang is a member of ACM.

Xiaoou Tang (S'93–M'96–SM'02) received the B.S. degree from the University of Science and Technology of China, Hefei, in 1990, the M.S. degree from the University of Rochester, Rochester, NY, in 1991, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

He is a Professor and Director of the Multimedia Laboratory in the Department of Information Engineering, The Chinese University of Hong Kong. He is also the Group Manager of the Visual Computing Group, Microsoft Research Asia, Beijing.

Dr. Tang is a Local Chair of the IEEE International Conference on Computer Vision (ICCV) 2005, an Area Chair of ICCV'07, a Program Chair of ICCV'09, and a General Chair of the ICCV International Workshop on Analysis and Modeling of Faces and Gestures 2005. He is a Guest Editor of the Special Issue on Underwater Image and Video Processing for the IEEE JOURNAL OF OCEANIC ENGINEERING and the Special Issue on Image- and Video-Based Biometrics for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He is an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. His research interests include computer vision, pattern recognition, and video processing.

Hong-Jiang Zhang (F'04) received the Ph.D. degree from the Technical University of Denmark and the B.S. degree from Zhengzhou University, China, both in electrical engineering, in 1982 and 1991, respectively.

From 1992 to 1995, he was with the Institute of Systems Science, National University of Singapore, where he led several projects in video and image content analysis and retrieval and computer vision. From 1995 to 1999, he was a Research Manager at Hewlett-Packard Labs, responsible for research and technology transfers in the areas of multimedia management, computer vision, and intelligent image processing. In 1999, he joined Microsoft Research and is currently the Managing Director of the Advanced Technology Center. He has authored three books, more than 300 referred papers, eight special issues of international journals on image and video processing, content-based media retrieval, and computer vision, as well as more than 50 patents or pending applications.

Dr. Zhang currently serves on the editorial boards of five IEEE/ACM journals and a dozen committees of international conferences.