Discriminant Analysis with Tensor Representation*

[1]Shuicheng Yan, [2]Dong Xu, [3]Qiang Yang, [4]Lei Zhang, [1,4]Xiaoou Tang, [4]Hong-Jiang Zhang

[1]Department of Information Engineering, the Chinese University of Hong Kong, Shatin, Hong Kong
[2] MOE-Microsoft Key Laboratory of Multimedia Computing and Communication & Department of EEIS, University of Science and Technology of China, Hefei, Anhui, P. R. China
[3]Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong
[4]Microsoft Research Asia, Beijing, P.R. China
Contact: scyan@ie.cuhk.edu.hk

Abstract

*In this paper, we present a novel approach to solving the supervised dimensionality reduction problem by encoding an image object as a general tensor of 2nd or higher order. First, we propose a Discriminant Tensor Criterion (DTC), whereby multiple interrelated lower-dimensional discriminative subspaces are derived for feature selection. Then, a novel approach called k-mode Cluster-based Discriminant Analysis is presented to iteratively learn these subspaces by unfolding the tensor along different tensor dimensions. We call this algorithm **D**iscriminant **A**nalysis with **Te**nsor **R**epresentation (DATER), which has the following characteristics: 1) multiple interrelated subspaces can collaborate to discriminate different classes; 2) for classification problems involving higher-order tensors, the DATER algorithm can avoid the curse of dimensionality dilemma and overcome 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 generalized eigenvalue decomposition. We provide extensive experiments by encoding face images as 2nd or 3rd order tensors to demonstrate that the proposed DATER algorithm based on higher order tensors has the potential to outperform the traditional subspace learning algorithms, especially in the small sample size cases.*

1. Introduction

Subspace learning [15] is an important topic in computer vision research. Most traditional algorithms, such as *Principal Component Analysis* (PCA) [10] and *Linear Discriminant Analysis* (LDA) [1], treat an input image object as a vector [2][6]. Some recent works, however, have started to consider an object as a two dimensional matrix for unsupervised learning [8][17]. Liu *et. al.* [5] proposed a special LDA to compute the intra-class and inter-class scatter matrices by replacing the feature vectors with matrix-formed features. 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.

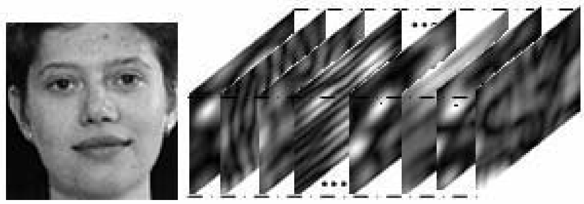
Our observation is as follows. In the real world, the extracted feature of an object often has some specialized structures and such structures are in the form of 2nd or even higher-order tensors. For example, this is the case when a captured image is a 2nd-order tensor, *i.e.* matrix, and when the sequential data such as a video sequence for event analysis, is in the form of 3rd-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 vector, which ignores the underlying data structure and often leads to the *curse of dimensionality* problem 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 2nd 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 avoid 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,* which maximizes the inter-class scatters and at the same time minimizes the intra-class scatters 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 an efficient procedure to iteratively learn these interrelated discriminant subspaces via a novel tensor analysis approach, called the *k-mode cluster-based discriminant analysis*. We explore the theoretical foundation of the *k-mode cluster-based discriminant analysis* to show that it unfolds the tensors into matrices along the *k*-th dimension. When the column vectors of the unfolded matrices are considered as the new objects to be analyzed, the cluster-based discriminant analysis is performed by clustering these samples according to their column indices of the unfolded matrices. This explanation, as we show later, provides an intuitive explanation for the superiority

*This work was performed at Microsoft Research Asia.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**



(a) 2nd-order Tensor (b) 3rd-order Gabor Tensor

Figure 1. Tensor representation examples: 2nd and 3rd order object representations.

of our proposed algorithm in comparison to other vector based approaches.

We summarize the advantages of our algorithm, *Discriminant Analysis with Tensor Representation* (DATER), as follows:

1) DATER is a general supervised dimensionality reduction framework. It can avoid the *curse of dimensionality dilemma* using higher order tensors and *k-mode cluster-based discriminant analysis,* because the latter is performed in a much lower-dimension feature space than the traditional vector-based methods, such as LDA, do.

2) DATER also helps overcome the *small sample problem*. As explained later, in the *k-mode cluster-based discriminant analysis*, the sample size is effectively multiplied by a large scale.

3) Much more feature dimensions are available in DATER than in LDA, because the available feature dimensions of LDA is theoretically limited by the number of classes in the data, whereas DATER is not.

4) The computational cost can be reduced to a large extent as the generalized eigenvalue decomposition in each step is performed on a feature space with smaller size.

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

2. Discriminant Analysis with Tensor Representation

Most previous approaches to subspace learning, such as the popular PCA and LDA, consider an object as a vector. The corresponding learning algorithms are performed on a very high dimensional feature space. As a result, these methods usually 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 a 2nd or higher order tensor. For example, the image matrix in Figure 1 (a) is a 2nd-order tensor and the filtered Gabor-image in Figure 1 (b) is a 3rd-order tensor. In this work, we study how to conduct discriminant analysis in the general case that objects are represented as tensors of 2nd or higher order.

2.1. 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 $\mathrm{a,b,c}$. Assume that the training sample set consists of the *n-th* order tensors $\{\mathbf{X}_i \in \mathbb{R}^{m_1 \times m_2 \times .. \times m_n},\ i=1,...,N\}$, where N is the number of tensor objects and $\mathbf{X}_i$ belongs to the class and is indexed as $c_i \in \{1,2,...,N_c\}$. Consequently, the sample set can be represented as an (*n*+1)-*th* order sample tensor $\tilde{\mathbf{X}} \in \mathbb{R}^{m_1 \times m_2 \times .. \times m_n \times N}$.

Before describing the *Discriminant Tensor Criterion*, we review the terminologies on tensor operations [3][4][11][12]. The inner product of tensors $\mathbf{A}$ and $\mathbf{B}$ with the same dimensions is $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i_1=1,...,i_n=1}^{m_1,...,m_n} \mathbf{A}_{i_1,...,i_n} \mathbf{B}_{i_1,...,i_n}$; 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 2nd-order tensor case, *i.e.* matrix-form, the norm is called Frobenius norm and is written as $\| A \|_F$. The *k-mode* product of tensor $\mathbf{A}$ and matrix $U \in \mathbb{R}^{m_k \times m_k'}$ is defined as $\mathbf{B} = \mathbf{A} \times_k U$, where $\mathbf{B}_{i_1,...,i_{k-1},j,i_{k+1},...,i_n} = \sum_{i=1}^{m_k} \mathbf{A}_{i_1,...,i_{k-1},i,i_{k+1},...,i_n} * U_{i,j},\ j=1,...,m_k'$ [12].

The *Discriminant Tensor Criterion* is designed to pursue multiple interrelated projection matrices, *i.e.* subspaces, which maximize the inter-class scatters and at the same time minimize the intra-class scatters as measured in the tensor metric as described above. That is,

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

where $\bar{\mathbf{X}}_c$ is the average tensor of the samples belonging to class *c*, $\bar{\mathbf{X}}$ is the total average tensor of all the samples, and n_c is the sample number of class *c*. Similar to the Fisher Criterion [1], the inter-class scatter is measured by the sum of the weighted distances between the class center tensors $\bar{\mathbf{X}}_c$ and total sample center tensor $\bar{\mathbf{X}}$; meanwhile, the intra-class scatter is measured by the sum of the distances between each sample to its corresponding center tensor. Despite the similarity, the data representation and metric are different between these two criterions.

Equation (1) is equivalent to a high-order nonlinear programming problem with a nonlinear constraint; thus normally there is no closed-form solution. Alternatively, we search for an iterative optimization approach to derive the interrelated discriminative subspaces.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

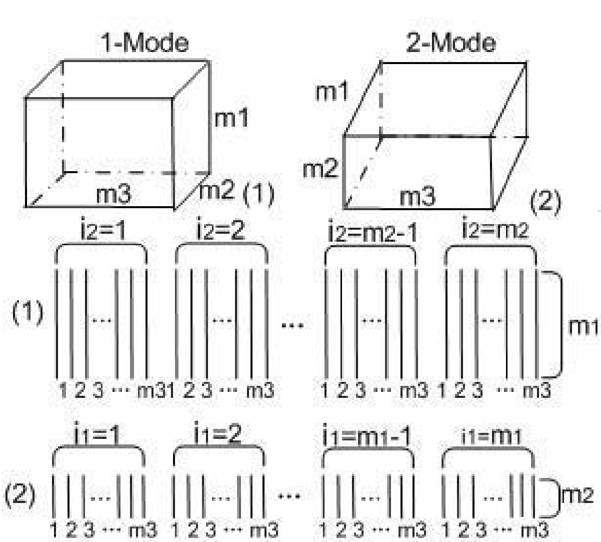


Figure 2. Illustration of the k-Mode unfolding and k-Mode Cluster-Based Discriminant Analysis for a 3rd -order Tensor.

2.2. *k*-mode Cluster-based Discriminant Analysis

Before describing the iterative approach to optimizing the *Discriminant Tensor Criterion*, we introduce the *k*-mode cluster-based Discriminant Analysis approach, which optimizes the objective function from the *k-th* dimension of the *n-th* order tensor. We also provide a sound theoretical foundation for the optimization of the criterion by introducing the cluster-based Discriminant Analysis algorithm.

Cluster-based Discriminant Analysis. In the face recognition problem, it is often the case that the difference between the images of the same person is larger than the difference between the images of different persons due to the pose, illumination and expression variations. For this problem the highly nonlinear properties can be observed in the classification hyper-plane. An intuitive solution for this nonlinearity problem is to partition the sample data into several clusters and conduct LDA within each cluster in a local or global manner. To this end, a *Cluster-based Discriminant Analysis* method can be applied. It derives the same optimal projections for different clusters by optimizing the global Fisher Criterion that sums the scatter matrices from different clusters.

In describing the *Cluster-based Discriminant Analysis*, we assume that the samples are represented as vectors $\{x_i \in \mathbb{R}^m,\ i=1,...,N\}$. The data are clustered into K clusters, and k_i is the corresponding cluster index for sample $\boldsymbol{x_i}$. The Cluster-based Discriminant Analysis optimizes the following objective function:

$$w^* = \arg\max_w \frac{Tr(w' S_B w)}{Tr(w' S_W w)}$$

$$S_B = \sum_{k=1}^{K} S_B^k,\ \ S_B^k = \sum_{c=1}^{N_c} n_c^k (\overline{x}_c^k - \overline{x}^k)(\overline{x}_c^k - \overline{x}^k)^{\mathrm{T}} \quad (2)$$

$$S_w = \sum_{k=1}^{K} S_W^k,\ \ S_W^k = \sum_{i=1}^{N} \delta_{k,k_i} (x_i - \overline{x}_{c_i}^k)(x_i - \overline{x}_{c_i}^k)^{\mathrm{T}}$$

where symbols with superscript k means that their values are computed within the cluster k; *e.g.* $\overline{x}_c^k$ is the average vector of the samples belonging to class c and cluster k. The operator $Tr(\cdot)$ is the trace of a matrix and $\delta_{k,k_i} = 1$ if $k = k_i$; *0*, otherwise. Cluster-based Discriminant Analysis is a dimensionality reduction algorithm that is linear locally while nonlinear globally; thus, it has the potential to be superior to the LDA algorithm in the ability to separate different classes.

k-mode Cluster-based Discriminant Analysis. We now discuss how to optimize the objective function from only one direction of the tensor, *i.e.*

$$(U_k^*) = \arg\max_{U_k} \frac{\sum_c n_c \|\bar{\mathbf{X}}_c \times_k U_k - \bar{\mathbf{X}} \times_k U_k\|^2}{\sum_i \|\mathbf{X}_i \times_k U_k - \bar{\mathbf{X}}_{c_i} \times_k U_k\|^2} \quad (3)$$

Before this analysis, we introduce the conception of *k-mode unfolding of a tensor*. Figure 2 demonstrates two ways to unfold a 3rd-order tensor. In the 1-mode version, a tensor is unfolded into a matrix along 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 *n*-th order tensor.

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

$$F^k \in \mathbb{R}^{m_k \times \prod_{i \neq k} m_i} \Leftarrow_k \mathbf{X} \in \mathbb{R}^{m_1 \times m_2 \times \ldots \times m_n} \quad \text{with}$$

$$F_{i_k, j}^k = \mathbf{X}_{i_1,\ldots,i_n},\ \ j = 1 + \sum_{l=1, l \neq k}^{n} (i_l - 1) \prod_{o=l+1, o \neq k}^{n} m_o \quad (4)$$

The problem in Eqn. (3) is actually a special, *k*-mode, cluster-based discriminant analysis problem. It can be understood in two steps: 1) the sample tensors are unfolded into matrices in *k*-mode; 2) the column vector of the unfolded matrices is considered as a new object with the same class label as the original sample tensor and then the Cluster-based Discriminant Analysis is conducted by partitioning these new objects into multiple clusters according to their column indices in the unfolded matrices. Theorem 1 below proves this property.

Theorem 1. The optimization problem in Eqn. (3) can be reformulated as a special Cluster-based Discriminant Analysis problem as follows

1063-6919/05 $20.00 © 2005 **IEEE**

Discriminant Analysis with Tensor Representation:

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

1. Initialize $U_1^0 = I_{m_1}, U_2^0 = I_{m_2}, ..., U_n^0 = I_{m_n}$;
2. For t =1, 2, … , T_{max} do

   a) For k =1, 2,…, n do

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

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

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

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

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

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

Figure 3. The procedure for Discriminant Analysis with Tensor Representation

$$U_k^* = arg \max_{U_k} \frac{Tr(U_k^{\mathrm{T}} S_B U_k)}{Tr(U_k^{\mathrm{T}} S_W U_k)}$$

$$S_B = \sum_{j=1}^{\prod_{o \neq k} m_o} S_B^j, S_B^j = \sum_{c=1}^{N_c} n_c (\bar{X}_c^{k,j} - \bar{X}^{k,j})(\bar{X}_c^{k,j} - \bar{X}^{k,j})^{\mathrm{T}} \qquad \mathbf{(5)}$$

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

where, for the ease of presentation, $X_i^{k,j}$ represents the j-th column vector of matrix X_i^k that is the k-mode unfolded matrix from sample tensor $\mathbf{X}_i$. $\bar{X}_c^{k,j}$ and $\bar{X}^{k,j}$ are defined in the same way as $X_i^{k,j}$ with respect to tensors $\bar{\mathbf{X}}_c$ and $\bar{\mathbf{X}}$. In this formulation, the objects to be analyzed are the column vectors of the k-mode unfolded matrices of the tensor samples and they are clustered according to their column indices in the unfolded matrices.

Proof. Here, we take S_W as 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 tensor $\mathbf{X}$; then, we have

$$\sum_i \| \mathbf{X}_i \times_k U_k - \bar{\mathbf{X}}_{c_i} \times_k U_k \|^2 = \sum_i \| X_i^{k\mathrm{T}} U_k - \bar{X}_{c_i}^{k\mathrm{T}} U_k \|_F^2$$
$$= \sum_i Tr[U_k^{\mathrm{T}} (X_i^k - \bar{X}_{c_i}^k)(X_i^k - \bar{X}_{c_i}^k)^{\mathrm{T}} U_k]$$
$$= Tr[U_k^{\mathrm{T}} (\sum_i (X_i^k - \bar{X}_{c_i}^k)(X_i^k - \bar{X}_{c_i}^k)^{\mathrm{T}}) U_k]$$
$$= Tr[U_k^{\mathrm{T}} (\sum_{j=1}^{\prod_{o \neq k} m_o} \sum_i (X_i^{k,j} - \bar{X}_{c_i}^{k,j})(X_i^{k,j} - \bar{X}_{c_i}^{k,j})^{\mathrm{T}}) U_k]$$
$$= Tr(U_k^{\mathrm{T}} S_W U_k)$$

Similarly, we can prove that

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

Therefore, the optimization problem in Eqn. (3) can be reformulated as a special formulation of Eqn. (2), in which the tensor object is first unfolded into a matrix and each column vector of the k-mode unfolded matrices is considered as a new object. Then the *Cluster-based Discriminant Analysis* is conducted by clustering these new objects according to their column indices in the unfolded matrices. The new formulation of Eqn. (4) integrates the *k*-mode unfolding method and the *Cluster-based Discriminant Analysis* algorithm; hence the procedure to optimize it is called *k-mode Cluster-based Discriminant Analysis* and it can be solved via the generalized eigenvalue decomposition method. ■

2.3. Discriminant Analysis with Tensor Representation

As aforementioned, Discriminant Tensor Criterion often has no closed-form solutions. In response to this problem, we present an iterative procedure to solve the problem. In each iteration, assuming that $U_1, ..., U_{k-1}, U_{k+1}, ..., U_n$ were known, the Discriminant Tensor Criterion can be revised to

$$U_k^* = arg \max_{U_k} \frac{\sum_c n_c \| \bar{\mathbf{X}}_c \times_1 U_1 .. \times_n U_n - \bar{\mathbf{X}} \times_1 U_1 .. \times_n U_n \|^2}{\sum_i \| \mathbf{X}_i \times_1 U_1 .. \times_n U_n - \bar{\mathbf{X}}_{c_i} \times_1 U_1 .. \times_n U_n \|^2} \qquad \mathbf{(6)}$$

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

$$U_k^* = arg \max_{U_k} \frac{\sum_c n_c \| \bar{\mathbf{Y}}_c \times_k U_k - \bar{\mathbf{Y}} \times_k U_k \|^2}{\sum_i \| \mathbf{Y}_i \times_k U_k - \bar{\mathbf{Y}}_{c_i} \times_k U_k \|^2} \qquad \mathbf{(7)}$$

Eqn (7) has the same appearance of Eqn. (3) by replacing $\mathbf{X}_i$ with $\mathbf{Y}_i$; thus, it can be solved using the above described *k-mode cluster-based discriminant analysis* algorithm. That is, we first conduct the *k*-mode unfolding for each tensor object $\mathbf{Y}_i$, then the *cluster-based discriminant analysis* is applied on the column vectors of the unfolded matrices by partitioning the new objects into multiple clusters according to their column indices in the unfolded matrices. Note that the tensor $\mathbf{Y}_i$ may have different dimensions from $\mathbf{X}_i$, and the *k*-mode unfolding of the tensor $\mathbf{Y}_i$ is defined according to its own tensor dimensions. The entire procedure to optimize the *Discriminant Tensor Criterion* is listed in Figure 3.

3. Algorithmic Analysis

In this section, we discuss the merits of the proposed procedure in terms of learnablity and computational complexity. As described later, LDA is intrinsically a special case of *Discriminant Analysis with Tensor Representation* (DATER), when the latter reduces to the first-order tensor.

We show that *DATER* with higher-order tensors is superior to LDA in many aspects.

Singularity and Curse of dimensionality. In LDA, the size of the scatter matrices are $\prod_{k=1}^{n} m_k \times \prod_{k=1}^{n} m_k$ 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 intra-class scatter matrix is singular and the accuracy and robustness of the solution are degraded. For most pattern recognition problems, $\prod_{k=1}^{n} m_k$ is very large, hence to train a credible classifier requires a huge number of training samples for the learnablity of LDA. In DATER, however, the step-wise intra-class scatter matrix is of size $m_k \times m_k$, which is much smaller than that of LDA. As described in Section 2, the objects to be analyzed in DATER are the column vectors of the unfolded matrices and the sample number is enlarged to $\prod_{i \neq k} m_i * N$. $\prod_{i \neq k} m_i * N > m_k$ can be satisfied in most cases; therefore, there is far less singularity problem in DATER when using 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.

Available Projection Directions. The most important factor limiting the application of LDA is that the available dimension for pattern recognition has the upper bound $N_c - 1$. Although many algorithms have been proposed to utilize the null space of the intra-class scatter matrix, the intrinsic dimension cannot be larger than $N_c - 1$. In the proposed DATER 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 DATER in each step.

Proof. As in the Equation (5), $S_B = \sum_{j=1}^{\prod_{o \neq k} m_o} S_B^j$, then $rank(S_B) \leq \sum_{j=1}^{\prod_{i \neq k} m_i} rank(S_B^j) \leq (N_c - 1)\prod_{i \neq k} m_i$; on the other hand, $rank(S_B) \leq m_k$ and the equality is satisfied when all the column vectors of matrix S_B is in full rank along the row direction. 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 DATER and DATER presents discriminating capability evaluation for most features.

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

Connections to LDA and 2DLDA. LDA and 2DLDA [5] both optimize the so called Fisher Criterion:

$$w^* = arg\,\max_{w} \frac{Tr(w^T S_B w)}{Tr(w^T S_W w)} \quad \textbf{(8)}$$

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

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

In 2DLDA, the sample data are matrices represented as $\{X_i \in \mathbb{R}^{m_1 \times m_2}, i=1,...,N\}$ and the scatter matrices are computed by replacing the vectors in (8) as matrices and:

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

In both cases, the averages are defined in the same way as the case with tensor representation. Actually, with simple algebraic computation, LDA and 2DLDA are both special formulations of our proposed DATER: LDA can be reformulated as a special case of DATER with *n*=1; while 2DLDA can be reformulated as a special case of DATER with *n*=2, $U_1 = I_{m_1}$ and using only one subspace instead of two in regular DATER with n=2.

4. Experiments

In this section, two benchmark face databases CMU PIE [9] and FERET [7] were used to evaluate the effectiveness of our proposed algorithm, *Discriminant Analysis with Tensor Representation*, in face recognition accuracy. Our proposed algorithm is referred to as *DATER/2-2* and *DATER/3-3* for problems with tensor of 2nd and 3rd 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 2DLDA algorithms. The 2DLDA algorithm has been proved to be special DATER using a single subspace, thus is referred to as DATER/2-1 in the experiment. In order to compare with Fisherface fairly, we also report the best result on different feature dimensions, 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-dimension tensors or vectors via the learned subspaces, and the nearest neighbor was used as the classifier 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 algorithms are also evaluated with different number of training samples to demonstrate their robustness for small sample-size problems.

4.1. PIE database—DATER/3-3 and DATER/2-2

The CMU PIE (Pose, Illumination, and Expression) 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 sub-databases were used to evaluate our proposed algorithms.

In the first sub-database, referred to as PIE-1, 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. Histogram equilibrium was applied in the preprocessing step.

The data set was randomly partitioned into gallery and probe sets; and two samples per person was used for training. We extracted 40 Gabor features with five different scales and eight different directions in the down-sampled positions and each image is encoded as a 3rd order tensor of size 16*16*40. Table 2 shows the detailed face recognition accuracies. The results clearly demonstrate that DATER/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.

Table 2. Recognition accuracy (%) comparison of Eigenface, Fisherface/O and DATER 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
DATER/2-1 (Grey)	72.8
DATER/2-2 (Grey)	80.2
DATER/3-3 (Gabor)	82.9

Another sub-database PIE-2 consists of the same five poses as in PIE-1, but the illumination indexed as 10 and 13 were also used. Therefore, the PIE-2 database is more difficult for classification. We conducted three sets of experiments on this sub-database. Table 3 lists all the comparative experimental results of the DATER/2-2, Eigenface, Fisherface/O and DATER/2-1. The reconstruction based Eigenface performs very poor in all the three cases; Fisherface is better than Eigenface, yet it also fails in the cases with only two training images for each person. In all the three experiments, DATER/2-2 performs the best.

Table 3. Recognition accuracy (%) comparison of DATER/2-2, Eigenface, Fisherface/*O* and DATER/2-1 on the PIE-2 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
DATER/2-1	74.3	71.9	63.5
DATER/2-2	82.3	80.7	66.7

4.2. FERET database—DATER/3-3

In this experiment, seventy people of the FERET database were used and each person has six different images, 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 3rd order tensor of size 28*23*40 for DATER/3-3.

We compared all the above mentioned algorithms on the FERET database. Table 4 demonstrates the comparative face recognition accuracies. Similar to the results in the PIE-1 sub-database, it shows that the Gabor features significantly improve the performance and DATER/3-3 consistently outperforms all the other algorithms.

Table 4. Recognition accuracy (%) comparison of DATER/3-3, Eigenface, Fisherface/*O*, DATER/2-1 and DATER/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
DATER/2-1 (Grey)	73.5
DATER/2-2 (Grey)	80.4
DATER/3-3 (Gabor)	83.6

4.3. Discussions

From the above experimental results, we can observe that:

1) DATER/3-3 consistently outperforms the other algorithms in all the cases. And in all cases, DATER/2-2 presents the best performance on the gray-level images.
2) DATER/3-3 and DATER/2-2 are very robust in the cases with a small number of samples. In these cases, Eigenface almost fails and no satisfying results are obtained. Fisherface/*O* is slightly better than Eigenface, but is still much worse than DATER/2-2 and DATER/3-3.

3) DATER/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 DATER/2-2 and DATER/3-3 in face recognition accuracy in all the cases.
4) As discussed in [1] that LDA is not always superior to PCA, especially in the cases when the training set can not represent the data distribution well. There are some cases in which Eigenface outperforms Fisherface, such as in the PIE-1 database.
5) 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 [13], dual-space LDA [14] and generalized Singular Value Decomposition [18], can also be applied on DATER with higher order tensors.

5. Conclusions

In this paper, a novel algorithm, *Discriminant Analysis with Tensor Representation (DATER)*, has been proposed for supervised dimensionality reduction with a general tensor representation. In *DATER*, image objects were encoded as *n-th* order tensors. An approach called *k-mode cluster-based Discriminant Analysis* was proposed to iteratively learn the multiple interrelated discriminative subspaces for the dimensionality reduction of the higher order tensor. Compared with traditional algorithms, such as PCA and LDA, the proposed algorithm effectively avoids the curse of dimensionality dilemma and overcomes the small-sample-size problem. Due to the low requirement on samples and the high performance in classification problem, *DATER* can be an alternative of the LDA algorithm for solving problems where objects are encoded as tensors. An interesting future application of our proposed DATER algorithm is to apply DATER/4-4 for video-based face recognition [16] and we are planning to explore this application in future work.

Acknowledgment

This work was partially supported by the Nature Science Fund of Anhui Province, P. R. China, under grant no. 03042307, a Hong Kong RGC Grant HKUST-6180/02E and a Hong Kong RGC/NSFC joint project N_CUHK409/03.

References

[1] P. Belhumeur, J. Hespanha and D. Kriegman. “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, No. 7, 1997, pp. 711-720.

[2] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. "Face Recognition using Laplacianfaces”, IEEE Transactions on PAMI, Vol. 27, No. 3, Mar. 2005.

[3] T. Kolda. “Orthogonal Tensor Decompositions”, SIAM Journal on Matrix Analysis and Applications, Volume 23, Number 1, 2001, pp. 243-255.

[4] L. Lathauwer, B. Moor and J. Vandewalle. “A Multilinear Singular Value Decomposition”, SIAM Journal on Matrix Analysis and Applications, Volume 21, Number 4, 2000, pp. 1253-1278.

[5] K. Liu, Y. Cheng and J. Yang. “Algebraic feature extraction for image recognition based on an optimal discriminant criterion”, Pattern Recognition (26), No. 6, June 1993, pp. 903-911.

[6] B. Moghaddam and A. Pentland. “Probabilistic Visual Learning for Object Representation”, *IEEE Trans. On PAMI*, vol. 19, pp. 696-710, 1997.

[7] I. Philips, H. Wechsler, J. Huang, and P. Rauss. “The FERET database and evaluation procedure for face recognition algorithms”, Image and Vision Computing, Vol. 16, PP.295-306, 1998.

[8] A. Shashua and A. Levin. “Linear Image Coding for Regression and Classification using the Tensor-rank Principle”, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Dec. 2001, Hawaii.

[9] T. Sim, S. Baker, and M. Bsat. “The CMU Pose, Illumination, and Expression (PIE) Database”, *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition,* May, 2002.

[10] M. Turk and A. Pentland. “Face Recognition Using Eigenfaces”, *IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, 1991.

[11] M. Vasilescu and D. Terzopoulos, "Multilinear Subspace Analysis for Image Ensembles", Proc. Computer Vision and Pattern Recognition Conf. (CVPR '03), Vol.2, Madison, WI, June, 2003, 93-99.

[12] M. Vasilescu and D. Terzopoulos. "TensorTextures: Multilinear Image-Based Rendering", Proc. ACM SIGGRAPH 2004, Conference Los Angeles, CA, August, 2004.

[13] X. Wang and X. Tang, "Random sampling LDA for face recognition", Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), June 2004.

[14] X. Wang and X. Tang, “Dual-space linear discriminant analysis for face recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2004.

[15] X. Wang and X. Tang, “A unified framework for subspace face recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1222-1228, 2004.

[16] X. Tang and Z. Li, "Frame synchronization and multi-level subspace analysis for video based face recognition," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), June 2004.

[17] J. Yang, D. Zhang, A. Frangi and J. Yang. “Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition,” IEEE Trans. On PAMI, Vol. 26, No. 1, pp. 131-137, Jan. 2004.

[18] J. Ye, R. Janardan, C. Park, and H. Park. “An optimization criterion for generalized discriminant analysis on undersampled problems”, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 26, N. 8, pp. 982-994, 2004.