Image Classification Using Correlation Tensor Analysis

Yun Fu, *Student Member, IEEE*, and Thomas S. Huang, *Life Fellow, IEEE*

***Abstract*—Images, as high-dimensional data, usually embody large variabilities. To classify images for versatile applications, an effective algorithm is necessarily designed by systematically considering the data structure, similarity metric, discriminant subspace, and classifier. In this paper, we provide evidence that, besides the Fisher criterion, graph embedding, and tensorization used in many existing methods, the correlation-based similarity metric embodied in supervised multilinear discriminant subspace learning can additionally improve the classification performance. In particular, a novel discriminant subspace learning algorithm, called correlation tensor analysis (CTA), is designed to incorporate both graph-embedded correlational mapping and discriminant analysis in a Fisher type of learning manner. The correlation metric can estimate intrinsic angles and distances for the locally isometric embedding, which can deal with the case when Euclidean metric is incapable of capturing the intrinsic similarities between data points. CTA learns multiple interrelated subspaces to obtain a low-dimensional data representation reflecting both class label information and intrinsic geometric structure of the data distribution. Extensive comparisons with most popular subspace learning methods on face recognition evaluation demonstrate the effectiveness and superiority of CTA. Parameter analysis also reveals its robustness.**

***Index Terms*—Correlation tensor analysis (CTA), discriminant analysis, face recognition, image classification, subspace learning.**

I. Introduction

IMAGE classification, especially face image classification, attracts much attention recently as a result of the increasing demand for developing real-world vision systems. Face recognition using appearance-based learning algorithms, a particular open topic in such areas, brings out a great deal of research and discussion due to its promising applications. However, improving the face recognition performance is still difficult because of the large variability in the facial appearance. In general, there are four main factors that mostly affect the recognition accuracy: *discriminant subspace* (dimensionality reduction), *similarity metric*, *data structure*, and *classifier*.

Classical methods, such as principal component analysis (PCA) [24] and linear discriminant analysis (LDA) [14], represent the input data as vectors and use Euclidean distance as the basic metric for nearest neighbor (NN) classification. These methods seek the global projection via an optimization formulation assuming the Gaussian distribution in the original data space. For some specific classification tasks, PCA is not very effective for extracting the most discriminant features. LDA may perform poorly for problems with small sample sizes.

When focusing more on the subspace learning aspect, some recent studies, such as locality preserving projection (LPP) [5] and locally embedded analysis (LEA) [8], reveal that local features and intrinsic geometric structures in the input space take on more discriminating power for classification. Such techniques assume that the high-dimensional data can be considered as a set of geometrically related points lying on a smooth low-dimensional manifold, e.g., a face manifold. Moreover, the subspace learning procedure incorporating both Fisher criterion and graph embedding is proven to be able to further boost the classifier's discriminating power [1], [9], [10].

Using the similarity metric for manifold structure representation and discriminant feature extraction is also reported in several studies [22], [29]. It is often the case that the Euclidean metric is incapable of capturing the intrinsic similarities between image data points. The semidefinite embedding (SDE) [23] estimates local angles and distances for the locally isometric embedding. The basic idea is to support the existence of a rotation, reflection and/or translation for the local mapping between data points and their neighbors. Conformal embedding analysis (CEA) [7] boosts its discriminating power by projecting the high-dimensional data onto the unit hypersphere [21], [22] and preserves intrinsic angle-based neighbor affinity with local graph modeling.

Besides the existing work on subspace learning aspect or manifold-based discriminant feature extraction, some recent studies also focus on the data structure aspect. These methods, such as tensor PCA [11], tensor LDA [12], tensor subspace analysis (TSA) [6], tensor LDE [15], and tensorface [13], treat original data as second- or higher-order tensors [25], [26]. For supervised classification involving high-dimensionality, the tensorization can lead to structured dimensionality reduction by learning multiple interrelated subspaces. Discriminant analysis using tensor representation can avoid the curse of dimensionality dilemma and overcome the small sample size problem. The computational cost of these methods is very low because the data dimensions in the generalized eigenvalue decomposition are significantly reduced.

Instead of focusing on a single factor, we are interested in comprehensively considering the factors in a systematic way and design a new discriminant subspace learning method, called correlation tensor analysis (CTA). Incorporating both correla-

Manuscript received June 19, 2007; revised November 2, 2007. This work was supported in part by the Beckman Graduate Fellowship, in part by the U.S. Government VACE program, and in part by the National Science Foundation under Grant CCF 04-26627. The views and conclusions are those of the authors, not of the U.S. Government or its Agencies. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dan Schonfeld.

The authors are with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign (UIUC), Urbana, IL 61801 USA (e-mail: yunfu2@ifp.uiuc.edu; huang@ifp.uiuc.edu).

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/TIP.2007.914203

tional mapping and discriminant analysis, CTA conducts the graph embedding to keep the data relationships measured by correlation for the same class; at the same time, it keeps apart data points of different classes in a Fisher analysis manner. As for the data structure aspect, CTA is designed to capture higher-order structures in the data patterns by encoding each object as a second- or higher-order tensor. Therefore, CTA learns multiple interrelated subspaces to obtain a low-dimensional data representation reflecting both class label information and intrinsic manifold structure. To be consistent in the classification, we use the maximal correlation affinity classifier to apply the similarity matching characteristic to CTA.

CTA has several properties that are worthy of highlighting. 1) The correlation metric used in CTA framework can estimate intrinsic angles and distances for the locally isometric embedding to support the existence of gray-level variations, rotations, and/or translations for the local mapping between image data points and their neighbors. This characteristic can deal with the case when Euclidean metric is incapable of capturing the intrinsic similarities between image data points. 2) CTA is designed to handle tensor representations for data structure. It, therefore, inherits the advantages of tensorization to avoid the curse of dimensionality dilemma, overcome the small sample size problem and significantly reduce computational cost. 3) CTA integrates the Fisher criterion and manifold embedding for supervised discriminant modeling. No data distribution assumption is made since locality affinity graphs are adopted to represent the intrinsic geometrical structure. 4) CTA embedding is derived from flexible designs. Different modes for setting correlation affinity weights can fit different learning scenarios and experimental protocols.

To demonstrate the proposed method, we focus on the second-order tensor analysis and perform appearance-based face recognition experiments on the CMU PIE [20] and Yale-B [18] + Extended Yale-B [19] databases. We extensively compare the classification performance of CTA with most popular subspace learning methods focusing on different aspects to boost the discriminating power. Experimental results demonstrate that CTA benefits from its encouraging properties and achieves competitive face recognition performance with a low computational complexity. Parameter analysis also suggests some empirical ways to better tune the parameters to improve the performance.

Before getting into details of our proposed method, in Section II, we first provide the objective and theoretical foundation of CTA. The formulation of CTA algorithm is presented in detail in Section III. We demonstrate the effectiveness of CTA in Section IV with experiments and conclude the paper at last.

II. Theoretical Foundation and Objective

Before introducing the CTA algorithm, we first present the basic theoretical foundation and objective in this section.

A. Similarity by Correlation

"*Correlation*" indicates the strength and direction of a linear relationship between two random variables. In the broad sense, measuring the degree of correlation is apt to measure the particular intrinsic distance or similarity between a variable pair. Suppose two image samples are considered as two random variables, X_1 and X_2. Then each pixel in the image denotes a measurement. Typically, if the two series of measurements are represented by two vectors, respectively, the correlation coefficient between them can be viewed as the normalized inner product or the cosine angle between them. The Pearson correlation coefficient is one possible definition which only works with centered data that are shifted by the mean of measurements so as to have an average of zero. The uncentered Pearson correlation coefficient is another possible way to define such metric [28]. If the two series of (centered or uncentered) measurements are represented by two mth-order tensors, $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}$, respectively, we extend the correlation concept to the following definition:

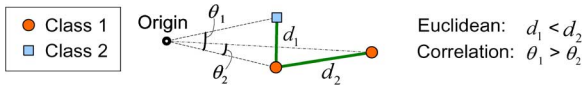

Fig. 1. Euclidean measure versus correlation measure.

$$\mathrm{Corr}(X_1, X_2) = \frac{\langle \mathbf{X}_1, \mathbf{X}_2 \rangle}{\sqrt{\langle \mathbf{X}_1, \mathbf{X}_1 \rangle}\sqrt{\langle \mathbf{X}_2, \mathbf{X}_2 \rangle}} \tag{1}$$

where, in a general definition, the inner product of two tensors $\mathbf{X}_1$ and $\mathbf{X}_2$ with the same dimension is $\langle \mathbf{X}_1, \mathbf{X}_2 \rangle = \sum_{i_1=1,i_2=1,\ldots,i_m=1}^{D_1,D_2,\ldots,D_m} \mathbf{X}_{1;i_1,i_2,\ldots,i_m} \mathbf{X}_{2;i_1,i_2,\ldots,i_m}$. Note that $\mathrm{Corr}(X_1, X_2) \in [-1, 1]$ and large correlation value indicates small distance between the two samples. Hence, the metric between X_1 and X_2 can be defined as $1 - \mathrm{Corr}(X_1, X_2)$.

In the first-order tensor case, for example, we can see that the correlation metric on sample vectors considers the angle distance while discarding the scaling on the magnitude. Such a metric has the advantage of stability to noise and is insensitive to global scaling of the vector magnitude, implying a good generalization ability. For example, in Fig. 1, the two classes are correctly classified with a correlation measure, while the Euclidean measure fails in this case. This extreme case of example is not typical of the classification performance of correlation measure, though it really often happens in image processing when illumination dominate the variations. We utilize it to clearly justify the intuition of out proposed algorithm for image classification.

B. Correlation Graph Embedding

Suppose $\{\mathbf{X}_i | \mathbf{X}_i \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}\}_{i=1}^n$ denotes a set of mth-order tensors representing the measurements of n images $\{X_i\}_{i=1}^n$. If $\{X_i\}_{i=1}^n$ constructs a nonlinear submanifold embedded in $\mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}$, we can learn a faithful embedding to reveal the intrinsic structure of the manifold by keeping the correlation affinity between data points. CTA borrows the Graph Embedding [1] framework for designing the objective. The essential trick is to design a new similarity metric to form the graph and learn the embedding. In the graph, the connectivity can be defined by k-maximal correlational neighbors—the case when X_i is in one of the k neighbors of X_j corresponding to

the k largest correlation values and vice versa. We, therefore, formulate the correlation graph embedding energy function as

$$\sum_{i,j=1}^{n} (1 - \mathrm{Corr}(Y_i, Y_j)) \cdot w_{ij} \tag{2}$$

where low-dimensional representation $\{Y_i\}_{i=1}^n$ is one-to-one corresponded to $\{X_i\}_{i=1}^n$ and $\{w_{ij}\}_{i,j=1}^n$ is the weight set associated with $\{X_i\}_{i=1}^n$ in the graph. In the CTA framework, the measurements of $\{Y_i\}_{i=1}^n$ are represented by the set of mth-order tensors $\{\mathbf{Y}_i | \mathbf{Y}_i \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_m}\}_{i=1}^n$, where $d_i \leqslant D_i$ for $i = 1, 2, \ldots, m$. The purpose of this objective function is to describe the pairwise correlation of the global data manifold. A larger correlation value between two data points is weighted by a larger weight since the data pair has shorter distance between them. The embedding is usually calculated by minimizing (2) subject to some constraints.

C. Weight Modes

Suppose the original image samples are already normalized to have unit norm.[1] In general, the correlation distance between two images can be defined by $\mathrm{dist}(X_i, X_j) = 1 - \langle \mathbf{X}_i, \mathbf{X}_j \rangle$. Hence, we have $\mathrm{dist}(X_i, X_j) \in [0, 2]$. We define three optional modes of w_{ij} configuration to form the weight matrices $\mathbf{W}_\mathrm{s}$ and $\mathbf{W}_\mathrm{d}$ [7] which are used for solving (2).

1) *Balanced Soft Weights*. If node i and j are connected, the weight of the edge between X_i and X_j is set by $w_{ij} = \exp(-\mathrm{dist}(X_i, X_j)/t) = \exp((\langle \mathbf{X}_i, \mathbf{X}_j \rangle - 1)/t)$. Otherwise, $w_{ij} = 0$ if node i and j are not connected.
2) *Unbalanced Soft Weights*. If node i and j are connected, for $\mathbf{W}_\mathrm{s}$, the weight is set by $w_{ij}^{(\mathrm{s})} = \exp((\langle \mathbf{X}_i, \mathbf{X}_j \rangle - 1)/t_\mathrm{s})$, while for $\mathbf{W}_\mathrm{d}$, the weight is set by $w_{ij}^{(\mathrm{d})} = \exp((\langle \mathbf{X}_i, \mathbf{X}_j \rangle - 1)/t_\mathrm{d})$, where t_s and t_d can be different. Otherwise, $w_{ij} = 0$ if node i and j are not connected.
3) *Balanced Rigid Weights*. If node i and j are connected, the weight of the edge between them is set by $w_{ij} = 1$. Otherwise, $w_{ij} = 0$ if they are not connected.

The $\mathbf{W}_\mathrm{s}$ and $\mathbf{W}_\mathrm{d}$ are symmetric and nonnegative matrices which are defined over all data points to model the local correlation affinity structure of the manifold. The basic criteria in choosing the weights is to penalize the distance measure between data pair via the weights for calculating the embedding. The three modes here are related and defined to adapt to different type of applications, such as manifold visualization or pattern classification. More possible weight modes can be retrieved from [27].

III. Correlation Tensor Analysis

We are interested in the multilinear subspace analysis for classification. Therefore, the formulations for CTA are all derived from the tensor representations.

[1]In the first-order tensor representation (vector), the normalization process projects the original data onto a unit hypersphere. The correlation between two samples is equivalent to the cosine of the angle between them [7].

A. Discriminant Correlation Embedding With Tensor Representation

Assume that the given high-dimensional data samples are denoted as a set of n mth-order tensors $\{\mathbf{X}_i | \mathbf{X}_i \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}\}_{i=1}^n$ and the corresponding class labels $\{l_i | l_i \in \{1, \ldots, n_c\}\}_{i=1}^n$, with each datum X_i belonging to a class l_i. $\{X_i\}_{i=1}^n$ also constructs a nonlinear submanifold embedded in the original space $\mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m}$, e.g., the face submanifold. To obtain an embedding for discriminative subspace learning, the basic objective of CTA is to find a low-dimensional tensor representation $\{\mathbf{Y}_i\}_{i=1}^n$ of the original data with following three criterions through a transformation function $\mathcal{F}(\cdot)$, where $\mathbf{Y}_i = \mathcal{F}(\mathbf{X}_i)$:

- preserve the same-class pairwise correlation affinity while keeping away the different-class pairwise correlation affinity after the embedding;
- if the two original high-dimensional data samples have large correlation affinity, then the correlation affinity between the corresponded low-dimensional points are large as well;
- the embedded low-dimensional representation can better reflect the class relations with respect to the labeling information for classification.

In this paper, the transformation function $\mathcal{F}(\cdot)$ is specified by a tensor formulation. Let $\mathbf{Y}_i = \mathbf{X}_i \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_m \mathbf{U}_m$, where $\mathbf{U}_j \in \mathbb{R}^{D_j \times d_j}$ for $j = 1, \ldots, m$. So, our goal is to calculate all the optimal $\mathbf{U}_j$.

B. Objective Function and Optimization

In the embedding design, we define two correlation affinity graphs $\mathcal{G}_\mathrm{s}$ and $\mathcal{G}_\mathrm{d}$, both with n nodes. The ith node corresponds to the datum X_i. For $\mathcal{G}_\mathrm{s}$, we only consider each pair of data, X_i and X_j, that are from the *same* class with $l_i = l_j$. An edge is constructed between nodes i and j if X_j is among k_s-maximal correlational neighbors of X_i and vice versa. For $\mathcal{G}_\mathrm{d}$, we only consider each pair of data, X_i and X_j, that are from the *different* class with $l_i \neq l_j$. An edge is constructed between nodes i and j if X_j is among k_d-maximal correlational neighbors of X_i and vice versa. Note that k_s and k_d can be different and chosen with empirical values. Then the objective function of CTA is defined as follows:

$$\arg\max_{\mathbf{U}} \sum_{i,j=1}^{n} (1 - \mathrm{Corr}(Y_i, Y_j)) \cdot \left(w_{ij}^{(\mathrm{d})} - w_{ij}^{(\mathrm{s})} \right). \tag{3}$$

The above transformation from $\{X_i\}_{i=1}^n$ to $\{Y_i\}_{i=1}^n$ is originally derived from the basic idea of CEA [7], which conducts discriminant subspace learning by projecting the original data points from a high-dimensional unit hypersphere to a low-dimensional unit hypersphere in terms of a Fisher type of learning criterion. According to the definition in (1) and (3) can be rewritten as

$$\arg\max_{\mathbf{U}} \sum_{i,j=1}^{n} \left(1 - \frac{\langle \mathbf{Y}_i, \mathbf{Y}_j \rangle}{\sqrt{\langle \mathbf{Y}_i, \mathbf{Y}_i \rangle \langle \mathbf{Y}_j, \mathbf{Y}_j \rangle}} \right) \cdot w_{ij}^{(\mathrm{d-s})} \tag{4}$$

where $\mathbf{Y}_i = \mathbf{X}_i \times_q \mathbf{U}_q|_{q=1}^m$, $\mathbf{Y}_j = \mathbf{X}_j \times_q \mathbf{U}_q|_{q=1}^m$, and $w_{ij}^{(\mathrm{d-s})} = w_{ij}^{(\mathrm{d})} - w_{ij}^{(\mathrm{s})}$. This is a nonlinear objective function. Define the k-mode unfolding [2] of tensor $\mathbf{X}$ into a matrix as

$$\mathbf{X} \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_m} \Rightarrow_k \mathbf{Z}^{(k)} \in \mathbb{R}^{D_k \times \prod_{i \neq k} D_i}, \text{ with}$$
$$\mathbf{Z}_{i_k,j}^{(k)} = \mathbf{X}_{i_1,\ldots,i_m}, j = 1 + \sum_{l=1, l\neq k}^{m} (i_l - 1) \prod_{r=l+1, r\neq k}^{m} D_r. \quad (5)$$

Suppose $\mathbf{Z}_i = \mathbf{X}_i \times_1 \mathbf{U}_1 \ldots \times_{k-1} \mathbf{U}_{k-1} \times_{k+1} \mathbf{U}_{k+1} \ldots \times_m \mathbf{U}_m$, we have $\|\mathbf{Z}_i \times_k \mathbf{U}_k\| = \|\mathbf{Z}_i^{(k)T}\mathbf{U}_k\|$ and $\|\mathbf{Z}_j \times_k \mathbf{U}_k\| = \|\mathbf{Z}_j^{(k)T}\mathbf{U}_k\|$. Then (4) can be rewritten as (6), shown at the bottom of the page.

There is no closed-form solution for the optimization problem in (6). An iterative procedure can be used to find the approximate solutions. The basic idea is to first initialize $\mathbf{U}_1, \ldots, \mathbf{U}_m$ with a valid initial solution; then assume $\mathbf{U}_1, \ldots, \mathbf{U}_{k-1}, \mathbf{U}_{k+1}, \ldots, \mathbf{U}_m$ are known so that $\mathbf{U}_k$ is solved by fixing the others. To solve $\mathbf{U}_k$, we define the gradient descent rule for optimization by differentiating ε with respect to matrix $\mathbf{U}_k$. This can be optimized by steepest descent, conjugate gradients or other fast gradient descent methods. Then we can simply derive (7)

$$\frac{\partial \varepsilon(\mathbf{U}_k)}{\partial \mathbf{U}_k} = \sum_{i,j=1}^{n} \left[\frac{f_{ij} \left(\mathbf{Z}_i^{(k)} \mathbf{Z}_i^{(k)T} \mathbf{U}_k \right)_{:}^{T}}{f_i^3 f_j} + \frac{f_{ij} \left(\mathbf{Z}_j^{(k)} \mathbf{Z}_j^{(k)T} \mathbf{U}_k \right)_{:}^{T}}{f_j^3 f_i} - \frac{\left(\left(\mathbf{Z}_i^{(k)} \mathbf{Z}_j^{(k)T} + \mathbf{Z}_j^{(k)} \mathbf{Z}_i^{(k)T} \right) \mathbf{U}_k \right)_{:}^{T}}{f_i f_j} \right] \cdot w_{ij}^{(\mathrm{d-s})} \quad (7)$$

where

$$f_i = \sqrt{\mathrm{Tr}\left\{ \mathbf{U}_k^T \mathbf{Z}_i^{(k)} \mathbf{Z}_i^{(k)T} \mathbf{U}_k \right\}}$$
$$f_j = \sqrt{\mathrm{Tr}\left\{ \mathbf{U}_k^T \mathbf{Z}_j^{(k)} \mathbf{Z}_j^{(k)T} \mathbf{U}_k \right\}}$$
$$f_{ij} = \mathrm{Tr}\left\{ \mathbf{U}_k^T \mathbf{Z}_i^{(k)} \mathbf{Z}_j^{(k)T} \mathbf{U}_k \right\} \quad (8)$$

and $()_{:}$ denotes the long column vector formed by concatenating the columns of matrix in ().

C. Initial Solution

Since the objective function in (6) is not convex, we can only calculate a local maxima. Especially, when the dimension of the data space is too large, the gradient descent may not be deep enough to converge to a good solution. In order to get a sufficient optimum, a good initial solution before iterative gradient descent is very crucial. We, therefore, need to find an approximate formulation to calculate the initial point $\mathbf{U}^{(0)}$. Since $1 - \mathrm{Corr}(Y_i, Y_j)$ is highly correlated with $\langle \mathbf{Y}_i, \mathbf{Y}_i - \mathbf{Y}_j \rangle$, we define the objective formulation for initial point as (9). Without losing generality, we consider the following common form with $p = \mathrm{d}$ or s:

$$\begin{aligned} \varepsilon_p\left(\mathbf{U}_k^{(0)}\right) &= \sum_{i,j=1}^{n} \langle \mathbf{Y}_i, \mathbf{Y}_i - \mathbf{Y}_j \rangle \cdot w_{ij}^{(p)} \\ &= \sum_{i,j=1}^{n} \left\langle \mathbf{X}_i \times_q \mathbf{U}_q|_{q=1}^m, \mathbf{X}_i \times_q \mathbf{U}_q|_{q=1}^m \right. \\ &\qquad \left. - \mathbf{X}_j \times_q \mathbf{U}_q|_{q=1}^m \right\rangle \cdot w_{ij}^{(p)}. \end{aligned} \quad (9)$$

Then (9) can be rewritten as (10)

$$\begin{aligned} \varepsilon_p\left(\mathbf{U}_k^{(0)}\right) &= \sum_{i,j=1}^{n} \left(\left\langle \mathbf{Z}_i^{(k)T}\mathbf{U}_k^{(0)}, \mathbf{Z}_i^{(k)T}\mathbf{U}_k^{(0)} \right\rangle \right. \\ &\qquad \left. - \left\langle \mathbf{Z}_i^{(k)T}\mathbf{U}_k^{(0)}, \mathbf{Z}_j^{(k)T}\mathbf{U}_k^{(0)} \right\rangle \right) \cdot w_{ij}^{(p)} \\ &= \sum_{i,j=1}^{n} \mathrm{Tr}\left\{ \mathbf{U}_k^{(0)T} \mathbf{Z}_i^{(k)} \mathbf{Z}_i^{(k)T} \mathbf{U}_k^{(0)} \right. \\ &\qquad \left. - \mathbf{U}_k^{(0)T} \mathbf{Z}_i^{(k)} \mathbf{Z}_j^{(k)T} \mathbf{U}_k^{(0)} \right\} \cdot w_{ij}^{(p)} \\ &= \mathrm{Tr}\left\{ \mathbf{U}_k^{(0)T} \left(\sum_{i,j=1}^{n} \mathbf{Z}_i^{(k)} w_{ij}^{(p)} \mathbf{Z}_i^{(k)T} \right. \right. \\ &\qquad \left. \left. - \sum_{i,j=1}^{n} \mathbf{Z}_i^{(k)} w_{ij}^{(p)} \mathbf{Z}_j^{(k)T} \right) \mathbf{U}_k^{(0)} \right\} \\ &= \mathrm{Tr}\left\{ \mathbf{U}_k^{(0)T} \mathbf{Z}^{(k)} (\mathbf{D}_p - \mathbf{W}_p) \mathbf{Z}^{(k)T} \mathbf{U}_k^{(0)} \right\} \end{aligned} \quad (10)$$

$$\begin{aligned} \varepsilon(\mathbf{U}_k) &= \sum_{i,j=1}^{n} \left(1 - \frac{\left\langle \mathbf{Z}_i^{(k)T}\mathbf{U}_k, \mathbf{Z}_j^{(k)T}\mathbf{U}_k \right\rangle}{\sqrt{\left\langle \mathbf{Z}_i^{(k)T}\mathbf{U}_k, \mathbf{Z}_i^{(k)T}\mathbf{U}_k \right\rangle \left\langle \mathbf{Z}_j^{(k)T}\mathbf{U}_k, \mathbf{Z}_j^{(k)T}\mathbf{U}_k \right\rangle}} \right) \cdot w_{ij}^{(\mathrm{d-s})} \\ &= \sum_{i,j=1}^{n} \left(1 - \frac{\mathrm{Tr}\left\{ \mathbf{U}_k^T \mathbf{Z}_i^{(k)} \mathbf{Z}_j^{(k)T} \mathbf{U}_k \right\}}{\sqrt{\mathrm{Tr}\left\{ \mathbf{U}_k^T \mathbf{Z}_i^{(k)} \mathbf{Z}_i^{(k)T} \mathbf{U}_k \right\} \mathrm{Tr}\left\{ \mathbf{U}_k^T \mathbf{Z}_j^{(k)} \mathbf{Z}_j^{(k)T} \mathbf{U}_k \right\}}} \right) \cdot w_{ij}^{(\mathrm{d-s})} \end{aligned} \quad (6)$$

(a) (b)

Fig. 2. Face image samples. (a) Two subjects in the PIE database. (b) One subject in the Yale-B database.

TABLE I
FACE RECOGNITION PERFORMANCE COMPARISON ON THE PIE DATABASE

Method	5 Train. Sample		10 Train. Sample		15 Train. Sample		20 Train. Sample	
	Error	Dim.	Error	Dim.	Error	Dim.	Error	Dim.
Baseline	75.66%	400	63.54%	400	54.80%	400	47.16%	400
PCA	75.66%	200	63.54%	245	54.80%	285	47.16%	282
LDA	48.73%	70	33.58%	51	26.78%	46	21.11%	45
LPP	62.68%	196	36.15%	67	23.76%	59	17.65%	66
LSDA	53.40%	67	30.51%	68	21.83%	63	15.13%	55
TLDA	46.50%	9,9	32.97%	9,9	26.78%	10,10	22.58%	12,12
TLPP	47.57%	9,9	34.25%	10,10	27.01%	10,10	22.27%	12,12
TLDE	62.53%	20,20	47.12%	20,20	38.24%	20,20	30.25%	19,19
N-TLDA	42.14%	9,9	26.65%	9,9	21.21%	9,9	17.12%	9,9
N-TLPP	48.38%	13,13	33.64%	18,18	22.14%	10,10	17.12%	9,9
N-TLDE	48.38%	20,20	32.60%	20,20	24.54%	20,20	19.12%	20,20
CTA	**41.78%**	16,16	**24.39%**	16,16	**17.26%**	14,14	**12.82%**	14,14

where $\mathbf{D}_p$ is diagonal and $\mathbf{D}_p[i,i] = \sum_j w_{ij}^{(p)}$ for $p = \mathrm{d}$ or s. We further write the objective function of initial solution as follows:

$$\arg\max_{\mathbf{U}_k^{(0)}} \frac{\varepsilon_{\mathrm{d}}\left(\mathbf{U}_k^{(0)}\right)}{\varepsilon_{\mathrm{s}}\left(\mathbf{U}_k^{(0)}\right)} = \arg\max_{\mathbf{U}_k^{(0)}} \frac{\mathrm{Tr}\left\{\mathbf{U}_k^{(0)T}\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{d}}-\mathbf{W}_{\mathrm{d}})\mathbf{Z}^{(k)T}\mathbf{U}_k^{(0)}\right\}}{\mathrm{Tr}\left\{\mathbf{U}_k^{(0)T}\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{s}}-\mathbf{W}_{\mathrm{s}})\mathbf{Z}^{(k)T}\mathbf{U}_k^{(0)}\right\}}. \quad (11)$$

The optimization problem of the initial point here is still not a closed-form solution; therefore, we still need iterative calculations. Again, we consider the idea to first arbitrarily initialize $\mathbf{U}_1^{(0)},\ldots,\mathbf{U}_m^{(0)}$; then assume $\mathbf{U}_1^{(0)},\ldots,\mathbf{U}_{k-1}^{(0)},\mathbf{U}_{k+1}^{(0)},\ldots,\mathbf{U}_m^{(0)}$ are known so that $\mathbf{U}_k^{(0)}$ is solved by fixing the others. One possible way to compute $\mathbf{U}_k^{(0)}$ is to solve the generalized eigenvalue decomposition problem in (12), which is equivalent to the optimization problem in (11)

$$\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{d}}-\mathbf{W}_{\mathrm{d}})\mathbf{Z}^{(k)T}\mathbf{U}_k^{(0)} = \lambda\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{s}}-\mathbf{W}_{\mathrm{s}})\mathbf{Z}^{(k)T}\mathbf{U}_k^{(0)}. \quad (12)$$

An alternative way to compute $\mathbf{U}_k^{(0)}$ is to transform the trace ratio optimization problem in (11) into a trace difference optimization problem defined as follows:

$$\arg\max_{\mathbf{U}_k^{(0)}} \mathrm{Tr}\left\{\mathbf{U}_k^{(0)T}\left[\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{d}}-\mathbf{W}_{\mathrm{d}})\mathbf{Z}^{(k)T} - \mu\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{s}}-\mathbf{W}_{\mathrm{s}})\mathbf{Z}^{(k)T}\right]\mathbf{U}_k^{(0)}\right\} \quad (13)$$

where μ is the ratio value in (11) computed in the previous iteration. In this case, the generalized eigenvalue decomposition problem is changed to

$$\left[\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{d}}-\mathbf{W}_{\mathrm{d}})\mathbf{Z}^{(k)T} - \mu\mathbf{Z}^{(k)}(\mathbf{D}_{\mathrm{s}}-\mathbf{W}_{\mathrm{s}})\mathbf{Z}^{(k)T}\right]\mathbf{U}_k^{(0)} = \lambda\mathbf{U}_k^{(0)}. \quad (14)$$

It has been demonstrated in [3] that the second method can monotonously increase the value of the objective function with guaranteed convergency, which may leads to superior performance over the first.

D. Maximal Correlation Affinity Classification

In addition to subspace learning, we also use the correlation affinity for classification to keep the consistency of the similarity measure. For a testing data point tensor $\mathbf{X}_t \in \mathbb{R}^{D_1\times D_2\times\cdots\times D_m}$, we can project it into the learned tensor subspace by $\mathbf{Y}_t = \mathcal{F}(\mathbf{X}_t) \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_m}$. The classification problem is defined to predict $l_t := l_{i^*}$, satisfying

$$\mathbf{Y}_{i^*} = \mathcal{F}(\mathbf{X}_{i^*}) = \arg\max_{\mathbf{Y}_i} \frac{\langle\mathbf{Y}_t,\mathbf{Y}_i\rangle}{\sqrt{\langle\mathbf{Y}_t,\mathbf{Y}_t\rangle}\sqrt{\langle\mathbf{Y}_i,\mathbf{Y}_i\rangle}}. \quad (15)$$

IV. EXPERIMENTS

We evaluate the performance of proposed CTA algorithm for face recognition on the two popular large-size face image databases: CMU PIE [20] and Yale-B [18] + Extended Yale-B [19]. Fig. 2 shows some face image samples in the two databases. These two databases both contain large illumination variations in face images, which are apt to be tackled by most

TABLE II
FACE RECOGNITION PERFORMANCE COMPARISON ON THE YALE-B DATABASE

Method	5 Train. Sample		10 Train. Sample		20 Train. Sample		30 Train. Sample	
	Error	Dim.	Error	Dim.	Error	Dim.	Error	Dim.
Baseline	54.73 %	1024	36.06%	1024	31.22%	1024	27.71%	1024
PCA	54.73 %	185	36.06%	367	31.22%	321	27.71%	377
LDA	37.56%	48	18.91%	37	16.87%	45	14.94%	39
LPP	34.08%	87	18.03%	75	30.26%	211	20.20%	122
LSDA	35.46%	49	19.49%	35	26.26%	38	29.33%	96
TLDA	45.41%	32,32	30.51%	32,32	28.53%	32,32	21.22%	32,32
TLPP	36.98%	27,27	21.49%	22,22	15.55%	32,32	10.68%	32,32
TLDE	45.45%	32,32	26.22%	32,32	20.75%	32,32	14.16%	32,32
N-TLDA	34.70%	32,32	19.20%	32,32	16.63%	32,32	13.16%	32,32
N-TLPP	29.53%	32,32	9.70%	31,31	6.58%	31,31	4.64%	30,30
N-TLDE	38.85%	32,32	19.20%	32,32	12.35%	31,31	11.92%	32,32
CTA	**16.99%**	26,26	**7.60%**	23,23	**4.96%**	23,23	**2.94%**	23,23

discriminant subspace learning methods for classification. We compare CTA with the most popular linear subspace learning algorithms for face recognition, e.g., PCA [24], LDA [14], LPP [5], LSDA [4], Tensor LDA (TLDA) [12], Tensor LPP (TLPP) [6], Tensor LDE (TLDE) [10], Normalization+TLDA (N-TLDA), Normalization+TLPP (N-TLPP), and Normalization+TLDE (N-TLDE). The normalization preprocessing in N-TLDA, N-TLPP, and N-TLDE is the same as that in CTA when we calculate the Pearson correlation coefficient. Here, PCA is unsupervised while the others are all trained in a supervised manner. Those subspace learning methods are combined with Nearest Neighbor (NN) classifier for classification. The baseline recognition is performed in the original image space without any dimensionality reduction. All the results for those algorithms in comparison are from the best tuning of their parameters. The feature structures for baseline, PCA, LDA, LPP, and LSDA are all vectors, while for TLDA, TLPP, TLDE, N-TLDA, N-TLPP, N-TLDE, and CTA are all matrices. For all the tensorization methods, without losing generality, we only test the dimension reduction case of $d_1 = d_2$ for simplification.

A. Experimental Results on PIE Database

The CMU PIE (Pose, Illumination, and Expression) database contains in total 41368 images of 68 subjects with 500+ images for each. The face images were captured by 13 synchronized cameras and 21 flashes, under varying pose, illumination, and expression. For each subject, we manually select 168 images from five near frontal poses (C05, C07, C09, C27, C29) and all the images under different illuminations and expressions. Face images are manually aligned, cropped out from the selected images and resized to be 20×20, with 256 gray levels per pixel. The feature of each image is represented by a 400-dimensional column vector or a 20×20 matrix (2-D tensor).

We randomly choose 34 images per individual to form a subdatabase. There are 2312 images in total. A random database partition is performed with 5, 10, 15, and 20 images per individual for training, and the rest of the database for testing. Different subspace learning methods are applied to represent the facial features for NN classification. Table I shows the recognition results for each method with both error rates and dimensions for the optimal face subspaces. We can see that PCA performs the worst; LDA, LPP, and LSDA, are generally comparable; TLDA, TLPP, and TLDE perform comparable to each other. The normalization preprocessing in N-TLDA, N-TLPP, and N-TLDE can significantly improve the performance of TLDA, TLPP, and TLDE. As can be seen, CTA consistently outperforms the other 11 methods in all four cases of database partitions with the lowest error rates of 41.78%, 24.39%, 17.26%, and 12.82%, respectively. The dimensions of the 4 CTA subspaces corresponding to the best results are 16×16, 16×16, 14×14, and 14×14, respectively. The improvement of the recognition accuracy by CTA is significant.

B. Experimental Results on Yale-B Database

The Yale Face Database B contains 5760 single light source images of 10 subjects, each under 576 viewing conditions (nine poses $\times$ 64 illumination conditions). The extended Yale Face Database B contains 16128 images of 28 human subjects under nine poses and 64 illumination conditions. The data format of the two databases are the same. We combine the two databases and form a new one including 38 subjects' face images. For this database, the images are cropped and resized to 32×32, with 256 gray levels per pixel. The feature of each image is represented by a 1024-dimensional column vector or a 32×32 matrix (2-D tensor).

We select around 64 near frontal images under different illuminations per individual for our experiment. A random database partition is performed with 5, 10, 20, and 30 images per individual for training, and the rest of the database for testing. Again, different subspace learning methods are applied to represent the facial features for NN classification. Table II shows the recognition results for each method with both error rates and dimensions for the optimal face subspaces. We still see that PCA performs the worst; the other six methods are comparable

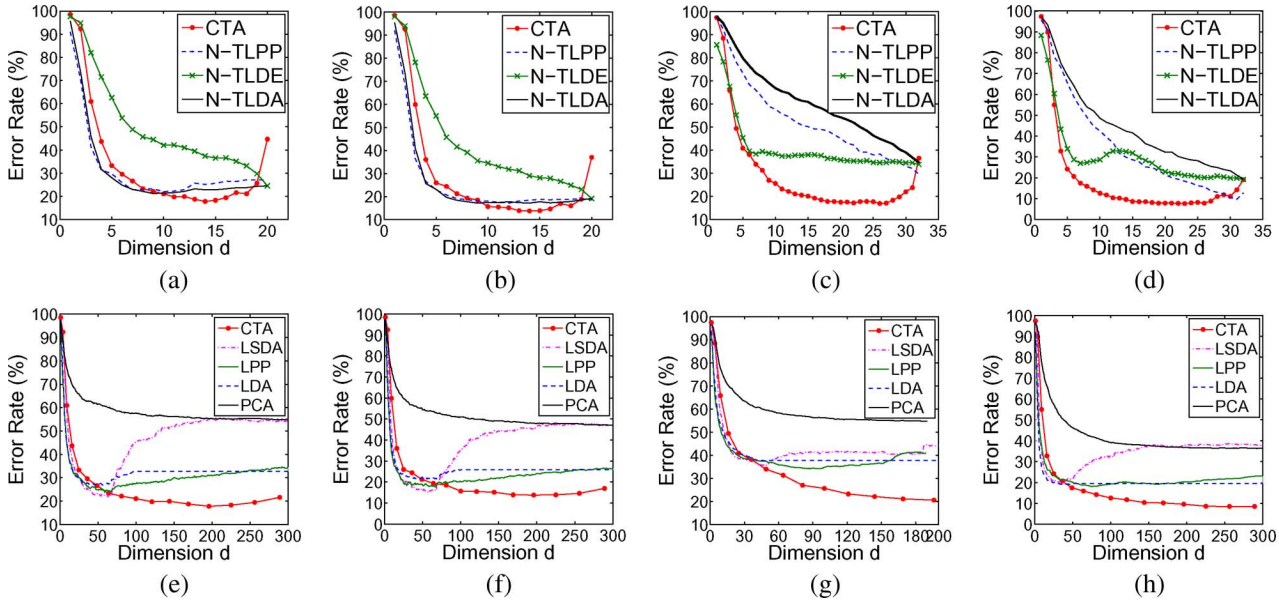

Fig. 3. Comparison of face recognition error rates versus dimensions of the learned subspaces on PIE and Yale-B. The dimensions of the learned subspaces for the tensorization methods are represented by $d = d_1 \times d_2$. (a) PIE: 15 train. samples; (b) PIE: 20 train. samples; (c) Yale-B: 5 train. samples; (d) Yale-B: 10 train. samples; (e) PIE: 15 train. samples; (f) PIE: 20 train. samples; (g) Yale-B: 5 train. samples; (h) Yale-B: 10 train. samples.

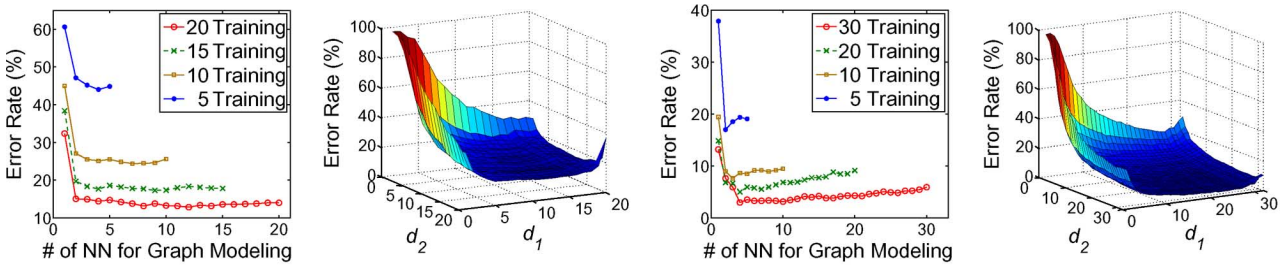

Fig. 4. Parameter analysis for face recognition experiments. (a), (c) Error rate of CTA face recognition versus number of NN for graph embedding on PIE and Yale-B in four cases of database partitions. (b), (d) Error rate of CTA face recognition versus number of dimension (d_1, d_2) for the learned tensor subspaces with 20 and 30 images per subject for training on PIE and Yale-B, respectively.

to each other. The normalization preprocessing in N-TLDA, N-TLPP, and N-TLDE can significantly improve the performance of TLDA, TLPP, and TLDE. CTA still consistently outperforms the other 11 methods in all four cases of database partitions with the lowest error rates of 16.99%, 7.60%, 4.96%, and 2.94%, respectively. The dimensions of the four CTA subspaces corresponding to the best results are 26×26, 23×23, 23×23, and 23×23, respectively. The improvement of the recognition accuracy by CTA is significant.

C. Discussion and Parameter Analysis

Fig. 3 shows the comparison plots of face recognition error rates versus dimensions of the learned subspaces on PIE and Yale-B for different database partition cases. The dimensions of the learned subspaces for the tensorization methods are represented by $d = d_1 \times d_2$. One possible reason of the superior performance of CTA in the test lies on the large illumination variation of PIE database. CTA benefits from its robust similarity measure which has the property of tolerating lighting and noise variations [7], [22]. It also shows that normalization, as a general preprocessing, can improve the recognition accuracy. However, since CTA significantly outperforms N-TLDA, N-TLPP, and N-TLDE, shown in Fig. 3(a)–(d), the advantages of CTA benefit not only from the data normalization, but also from its intrinsic discriminant properties, correlation metric, and maximal correlation affinity classification. In addition, CTA can run much faster than the other methods using vector-based data representation. Note that the sharp rise of error rate for CTA in Fig. 3(a) and (b), when the dimension is 20, is just a spike point, which can be ignored since 20 is the upper bound of the tensor space. The same phenomenon happens in Fig. 3(c) and (d) when 32 is the upper bound of the tensor space.

In the experiments, the number of NN for graph modeling (k_s, k_d) and the dimension of the tensor subspace (d_1, d_2) are tuned in a brute-force manner. To simplify the analysis, we consider the cases when $k_s = k_d = k$ and $d_1 = d_2 = d$. Fig. 4(a) and (c) shows the parameter analysis for the face recognition experiments. For each case of database partition, they illustrate the error rate of CTA face recognition versus k on PIE and Yale-B, respectively. We observe that the four error

rate curves in each figure approximately follow a similar trend. The variation of recognition performance becomes small when $k \geq 3$ or 4. So, CTA is not sensitive to the parameter k when it is relatively large. Fig. 4(b) and (d) illustrates the error rate of CTA face recognition versus number of dimension (d_1, d_2) for the learned tensor subspaces with 20 and 30 images per subject for training on PIE and Yale-B, respectively. These results are consistent with that of Fig. 4(a) and (c). The 3-D surfaces indicate the low recognition errors when the numbers of dimension (d_1, d_2) are all relatively large.

As discussed before, setting correlation affinity weights with different modes may introduce more parameters. For example, balanced soft weights may introduce one parameter t; unbalanced soft weights may introduce two parameters t_{s} and t_{d}. The more-parameter case may gain better performance, but may also lead to tedious tuning work at the same time.

V. Conclusion

We proposed to use correlation tensor analysis for appearance-based discriminant subspace learning. The design of CTA is focused not only on subspace learning and data structure aspects, but also on the similarity metric aspect to significantly boost the discriminating power of feature extraction. Incorporating both correlation measure and tensor representation, CTA conducts the graph embedding to differentiate image classes in a Fisher type of learning manner. Real-world face recognition experiments have demonstrated that the correlation-based similarity measure incorporated with supervised multilinear subspace learning can additionally improve the classification performance, especially for classifying illumination dominated imaging data. In particular, CTA benefits from both graph embedding and tensor analysis to embody nice properties, such as efficient dimensionality reduction (especially for small size training data), effective discriminative embedding, and low computational cost.

References

[1] S. C. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extension: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[2] S. C. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Discriminant analysis with tensor representation," in *Proc. IEEE Conf. CVPR*, 2005, vol. 1, pp. 526–532.

[3] H. Wang, S. C. Yan, T. S. Huang, and X. O. Tang, "A convergent solution to tensor subspace learning," in *Proc. IJCAI*, 2007, pp. 629–634.

[4] D. Cai, X. F. He, K. Zhou, J. W. Han, and H. J. Bao, "Locality sensitive discriminant analysis," in *Proc. IJCAI*, 2007, pp. 708–713.

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

[6] X. F. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," presented at the NIPS, 2005.

[7] Y. Fu, M. Liu, and T. S. Huang, "Conformal embedding analysis with local graph modeling on the unit hypersphere," presented at the IEEE CVPR Workshop on Component Analysis in Computer Vision, 2007.

[8] Y. Fu and T. S. Huang, "Locally linear embedded eigenspace analysis," IFP-TR, UIUC. [Online]. Available: www.ifp.uiuc.edu/~yunfu2/papers/LEA-Yun05.pdf, Jan. 1, 2005

[9] Y. Fu, S. Yan, and T. S. Huang, "Classification and feature extraction by simplexization," *IEEE Trans. Inf. Forensics Security*, to be published.

[10] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Conf. CVPR*, 2005, pp. 846–853.

[11] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. 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.

[12] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," presented at the NIPS, 2004.

[13] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," presented at the IEEE Conf. CVPR, 2003.

[14] P. N. Belhumeur, J. P. Hepanha, and D. J. 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.

[15] J. Xia, D.-Y. Yeung, and G. Dai, "Local discriminant embedding with tensor representation," in *Proc. IEEE Int. Conf. Image Processing*, 2006, pp. 929–932.

[16] M.-H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using Kernel methods," in *Proc. IEEE Conf. FG*, 2002, pp. 215–220.

[17] J. Ham, D. D. Lee, S. Mika, and B. Schölkopf, "A Kernel view of the dimensionality reduction of manifolds," presented at the ICML, 2004.

[18] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[19] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[20] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illuminlation, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[21] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the unit hypersphere using Von Mises-Fisher distributions," *J. Mach. Learn. Res.*, vol. 6, pp. 1345–1382, 2005.

[22] O. Hamsici and A. M. Martinez, "Spherical-homoscedastic distributions: The equivalency of spherical and normal distributions in classification," *J. Mach. Learn. Res.*, vol. 8, pp. 1583–1623, 2007.

[23] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," in *Proc. IEEE Conf. CVPR*, 2004, vol. 2, pp. 988–995.

[24] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. CVPR*, 1991, pp. 586–591.

[25] 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.

[26] D. Tao, X. Li, W. Hu, S. J. Maybank, and X. Wu, "Supervised tensor learning," in *Knowledge and Information Systems*. New York: Springer, 2007.

[27] X. J. Zhu, "Semi-supervised learning with graphs," Ph.D. dissertation, CMU-LTI-05-192, 2005.

[28] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," in *Proc. Nat. Acad. Sci. USA*, 1998, vol. 95, pp. 14863–14868.

[29] T.-K. Kim, J. Kittler, and R. Cippola, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1005–1018, Jun. 2007.

Yun Fu (S'07) received the B.E. degree in information engineering from the School of Electronic and Information Engineering, Xi'an Jiaotong University, China, in 2001, the M.S. degree in pattern recognition and intelligence systems from the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, in 2004, and the M.S. degree in statistics from the Department of Statistics, University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2007. He is currently pursuing the Ph.D. degree in the Electrical and Computer Engineering Department, UIUC.

His research interests include machine learning, human computer interaction, and image processing.

Mr. Fu is the recipient of a 2002 Rockwell Automation Master of Science Award; two Edison Cups of the 2002 GE Fund "*Edison Cup*" Technology Innovation Competition; the 2003 HP Silver Medal and Science Scholarship for Excellent Chinese Student; the 2007 Beckman Graduate Fellowship; and the 2007 DoCoMo USA Labs Innovative Paper Award (IEEE ICIP'07 Best Paper Award). He is a student member of Institute of Mathematical Statistics (IMS) and a 2007–2008 Beckman Graduate Fellow.

Thomas S. Huang (S'61–M'63–SM'76–F'79–LF'01) received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, R.O.C., and the M.S. and D.Sc. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge.

He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University, West Lafayette, IN, from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, Urbana, where he is now the William L. Everitt Distinguished Professor of Electlrical and Computer Engineering, a Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology and Co-Chair of the Institute's major research theme Human Computer Intelligent Interaction. He has published 20 books, and over 500 papers in network theory, digital filtering, image processing, and computer vision. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals.

Dr. Huang is a Member of the National Academy of Engineering, a Foreign Member of the Chinese Academies of Engineering and Sciences, and a Fellow of the International Association of Pattern Recognition and the Optical Society of America, and has received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. In 2000, he received the Honda Lifetime Achievement Award for "contributions to motion analysis." In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize from the International Association of Pattern Recognition and the Pan Wen-Yuan Outstanding Research Award. In 2005, he received the Okawa Prize. In 2006, he was named by IS&T and SPIE as the Electronic Imaging Scientist of the Year. He is a Founding Editor of the *International Journal Computer Vision, Graphics, and Image Processing* and Editor of the Springer Series in Information Sciences, published by Springer Verlag.