Uncorrelated Multilinear Principal Component Analysis for Unsupervised Multilinear Subspace Learning

Haiping Lu, *Member, IEEE*, Konstantinos N. (Kostas) Plataniotis, *Senior Member, IEEE*, and Anastasios N. Venetsanopoulos, *Fellow, IEEE*

***Abstract*—This paper proposes an uncorrelated multilinear principal component analysis (UMPCA) algorithm for unsupervised subspace learning of tensorial data. It should be viewed as a multilinear extension of the classical principal component analysis (PCA) framework. Through successive variance maximization, UMPCA seeks a tensor-to-vector projection (TVP) that captures most of the variation in the original tensorial input while producing uncorrelated features. The solution consists of sequential iterative steps based on the alternating projection method. In addition to deriving the UMPCA framework, this work offers a way to systematically determine the maximum number of uncorrelated multilinear features that can be extracted by the method. UMPCA is compared against the baseline PCA solution and its five state-of-the-art multilinear extensions, namely two-dimensional PCA (2DPCA), concurrent subspaces analysis (CSA), tensor rank-one decomposition (TROD), generalized PCA (GPCA), and multilinear PCA (MPCA), on the tasks of unsupervised face and gait recognition. Experimental results included in this paper suggest that UMPCA is particularly effective in determining the low-dimensional projection space needed in such recognition tasks.**

***Index Terms*—Dimensionality reduction, face recognition, feature extraction, gait recognition, multilinear principal component analysis (MPCA), tensor objects, uncorrelated features.**

I. Introduction

INPUT data sets in many practical pattern recognition problems are multidimensional in nature and they can be formally represented using tensors. Several indices are needed to address the elements of a tensor and the number of indices used defines the "order" of a tensor, with each index defining one "mode" [1]. There are many real-world tensor objects [2]–[5], such as gray-level images as second-order tensors consisting of the column and row modes [6]–[8], and gray-scale video sequences as third-order tensors consisting of the column, row, and time modes [9], [10]. In addition, many streaming and mining data are frequently organized as third-order tensors [11]–[14]. For instance, data in social network analysis are usually organized in three modes, namely time, author, and keywords [11]. A typical real-world tensor object is often specified in a high-dimensional tensor space and recognition methods operating directly on this space suffer from the so-called curse of dimensionality [15]. Nonetheless, a class of tensor objects in most applications are highly constrained to a subspace, a manifold of intrinsically low dimension [15]. Subspace learning, a popular dimensionality reduction method, is frequently employed to transform a high-dimensional data set into a low-dimensional space of equivalent representation while retaining most of the underlying structure [16]. The focus of this paper is on unsupervised subspace learning of tensorial data. For convenience of discussion, Table I lists the acronyms used in this paper.

TABLE I
List of Acronyms

Acronym	Description
2DPCA	Two-dimensional Principal Component Analysis [17]
CSA	Concurrent Subspaces Analysis [18]
EMP	Elementary Multilinear Projection [6], [19]
MPCA	Multilinear Principal Component Analysis [9]
PCA	Principal Component Analysis [20], [21]
TROD	Tensor Rank-One Decomposition [22]
TTP	Tensor-to-Tensor Projection [19]
TVP	Tensor-to-Vector Projection [6], [19]
UMPCA	Uncorrelated Multilinear Principal Component Analysis
UMLDA	Uncorrelated Multilinear Discriminant Analysis [6]

For unsupervised subspace learning of tensorial data, an obvious first choice is to utilize existing linear solutions such as the celebrated principal component analysis (PCA). PCA is a classical linear method for unsupervised subspace learning that transforms a data set consisting of a large number of interrelated variables to a new set of uncorrelated variables, while retaining most of the input data variations [20]. However, PCA on tensor objects requires the reshaping (vectorization) of tensors

Manuscript received October 31, 2008; revised July 15, 2009; accepted July 30, 2009. First published September 29, 2009; current version published November 04, 2009. This work was supported in part by the Ontario Centres of Excellence under the Communications and Information Technology Ontario Partnership Program and the Bell University Labs—at the University of Toronto.

H. Lu was with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering University of Toronto, Toronto, ON M5S 3G4 Canada. He is now with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore (e-mail: hplu@ieee.org).

K. N. Plataniotis is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering University of Toronto, Toronto, ON M5S 3G4 Canada (e-mail: kostas@comm.toronto.edu).

A. N. Venetsanopoulos is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering University of Toronto, Toronto, ON M5S 3G4 Canada and also with Ryerson University, Toronto, ON M5B 2K3 Canada (e-mail: anv@comm.toronto.edu).

Digital Object Identifier 10.1109/TNN.2009.2031144

into vectors in a very high-dimensional space. This not only increases the computational complexity and memory demands but most importantly destroys the structural correlation of the original data [7]–[9], [23]–[25]. It is commonly believed that potentially more compact or useful low-dimensional representations can be obtained directly from the original tensorial representation. Recently, there have been several proposed PCA extensions operating directly on tensor objects rather than their vectorized versions [8], [9], [26].

The tensor rank-one decomposition (TROD) algorithm introduced in [22] extracts features for a set of images based on variance maximization and the solution involves (greedy) successive residue calculation. The two-dimensional PCA (2DPCA) proposed in [17] constructs an image covariance matrix using image matrices as inputs. However, in 2DPCA, a linear transformation is applied only to the right-hand side of image matrices so the image data is projected in one mode only, resulting in poor dimensionality reduction [7]. In comparison, the generalized low rank approximation of matrices (GLRAM) algorithm in [7] applies linear transformations to the input image matrices from both the left- and right-hand sides and it is shown to outperform 2DPCA. While GLRAM targets approximation and reconstruction in its formulation, the generalized PCA (GPCA) proposed in [8] aims to maximize the captured variation, as a two-dimensional extension of PCA. In addition, the two-dimensional singular value decomposition (2DSVD) [27] provides a near-optimal solution for GLRAM and GPCA. For higher order extensions, the concurrent subspaces analysis (CSA) formulated in [26] targets at optimal reconstruction of general tensor objects, which can be considered as a further generalization of GLRAM, and the multilinear PCA (MPCA) introduced in [9] targets at variation maximization for general tensor objects, which can be considered as a further generalization of GPCA.

Nevertheless, none of the existing multilinear extensions of PCA mentioned above takes an important property of PCA into account, i.e., the fact that PCA derives uncorrelated features. Instead, these multilinear extensions of PCA produce orthogonal bases in each mode. Although uncorrelated features imply orthogonal projection bases in PCA [20], this is not necessarily the case for its multilinear extension. Uncorrelated features are highly desirable in many recognition tasks since they contain minimum redundancy and ensure linear independence among features [28]. In practical recognition tasks, uncorrelated features can greatly simplify the subsequent classification task. Thus, this paper investigates multilinear extension of PCA which can produce uncorrelated features. A novel uncorrelated multilinear PCA (UMPCA) is proposed for unsupervised tensor object subspace learning (dimensionality reduction). UMPCA utilizes the tensor-to-vector projection (TVP) principle introduced during the development of the uncorrelated multilinear discriminant analysis (UMLDA) framework presented in [6], and it parallelizes the successive variance maximization approach seen in the classical PCA derivation [20]. In UMPCA, a number of elementary multilinear projections (EMPs) are solved to maximize the captured variance, subject to the zero-correlation constraint. The solution is iterative in nature, as many other multilinear algorithms [8], [22], [26].

This paper makes two main contributions.

1) The introduction of a novel algorithm (UMPCA) for extracting uncorrelated features directly from tensors. The derived solution captures the maximum variation of the input tensors. As a multilinear extension of PCA, UMPCA not only obtains features that maximize the variance captured, but also enforces a zero-correlation constraint, thus extracting uncorrelated features. UMPCA is the only multilinear extension of PCA, to the best of the authors' knowledge, that can produce uncorrelated features in a fashion similar to that of the classical PCA, in contrast to other multilinear PCA extensions, such as 2DPCA [17], CSA [26], TROD [22], GPCA [8], and MPCA [9]. It should be noted that unlike the works reported in [9] and [26], TVP rather than tensor-to-tensor projection (TTP) is used here, and that unlike the heuristic approach of [22], this work takes a systematic approach in deriving the solution under the TVP paradigm. Interested readers should refer to [6], [9], and [19] for a detailed discussion on the topic of tensor projections, including TTP and TVP.
   It should be noted that although the proposed UMPCA algorithm takes an approach similar to that used in the UMLDA algorithm [6] to derive uncorrelated features through TVP, the two methods utilize different objective functions. UMPCA is an unsupervised learning algorithm which does not require labeled training data, while UMLDA is a supervised learning algorithm which requires access to labeled training samples. It is therefore obvious that UMPCA can be used in many applications where UMLDA cannot be applied. Examples include typical unsupervised learning tasks such as clustering, and the one-training-sample (OTS) scenario recently studied in the face recognition literature, where many supervised algorithms cannot be applied properly due to the availability of only one training sample per class [29], [30].
2) A systematic method to determine the maximum number of uncorrelated multilinear features that can be extracted under the UMPCA framework. The pertinent corollary provides insight into the possible uses of UMPCA. It helps designers and practitioners to understand the possible limitations of UMPCA and provides guidance on where and how UMPCA should be used. In the linear case, the derived constraint on the maximum number of uncorrelated features reduces to a well-known constraint on the rank of the data matrix in PCA.

The rest of this paper is organized as follows. Section II reviews the basic multilinear notation and operations, including TVP. In Section III, the problem is stated and the UMPCA framework is formulated, with an algorithm derived as a sequential iterative process. This section also includes a systematic way to determine the maximum number of uncorrelated features that can be extracted under the UMPCA framework. In addition, issues such as initialization, projection order, termination, convergence, and computational complexity are discussed in detail. Section IV studies the properties of the proposed UMPCA algorithm using three synthetic data sets and evaluates the effectiveness of UMPCA in face and gait recognition tasks by comparing its performance against that of PCA, 2DPCA, CSA,

TABLE II
LIST OF NOTATION

Notation	Description
$\mathcal{X}_m$	the mth input tensor sample, $m = 1, ..., M$
$\mathbf{u}^{(n)}$	the n-mode projection vector, $n = 1, ..., N$
$p = 1, ..., P$	the index of the EMP
P	the dimensionality of projected space (the number of EMPs in TVP)
$\{\mathbf{u}_p^{(n)^T}, n = 1, ..., N\}$	the pth EMP
$\mathbf{y}_m$	the projection of $\mathcal{X}_m$ on the TVP $\{\mathbf{u}_p^{(n)^T}, n = 1, ..., N\}_{p=1}^{P}$
$\mathbf{y}_m(p) = \mathbf{g}_p(m)$	the projection of $\mathcal{X}_m$ on the pth EMP $\{\mathbf{u}_p^{(n)^T}, n = 1, ..., N\}$
$\mathbf{g}_p$	the pth coordinate vector
K	the maximum number of iterations in UMPCA

TROD, GPCA, and MPCA. Finally, Section V draws the conclusions. It should be noted that in order to take a systematic approach to the problem of interest, this work shares some similarity in presentation with [6].

II. MULTILINEAR FUNDAMENTALS

This section introduces the fundamental multilinear notation, operations, and projections necessary for the presentation of UMPCA. For more detailed understanding of the materials presented here, readers should refer to previously published works, such as those in [1], [6], [9], and [19]. The notation conventions used in this paper are listed in Table II.

A. Notation and Basic Multilinear Operations

Following the notational convention of [1], vectors are denoted by lowercase boldface letters, e.g., $\mathbf{x}$; matrices are denoted by uppercase boldface letters, e.g., $\mathbf{U}$; and tensors are denoted by calligraphic letters, e.g., $\mathcal{A}$. Their elements are denoted with indices in parentheses. Indices are denoted by lowercase letters and span the range from 1 to the uppercase letter of the index, e.g., $n = 1, 2, \ldots, N$.

An Nth-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is addressed by N indices $i_n, n = 1, \ldots, N$, and each i_n addresses the n-mode of $\mathcal{A}$. The n-mode product of a tensor $\mathcal{A}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is a tensor with entries

$$(\mathcal{A} \times_n \mathbf{U})(i_1, \ldots, i_{n-1}, j_n, i_{n+1}, \ldots, i_N) = \sum_{i_n} \mathcal{A}(i_1, i_2, \ldots, i_N) \cdot \mathbf{U}(j_n, i_n). \quad (1)$$

The scalar product of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \cdots \sum_{i_N} \mathcal{A}(i_1, \ldots, i_N) \cdot \mathcal{B}(i_1, \ldots, i_N). \quad (2)$$

A rank-one tensor $\mathcal{A}$ equals to the outer product of N vectors: $\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}$, which means that $\mathcal{A}(i_1, i_2, \ldots, i_N) = \mathbf{u}^{(1)}(i_1) \cdot \mathbf{u}^{(2)}(i_2) \cdots \mathbf{u}^{(N)}(i_N)$ for all values of indices.

B. Tensor-to-Vector Projection

In order to extract uncorrelated features from tensorial data directly, i.e., without vectorization, this work employs the TVP introduced in [6]. A brief review on TVP is given here and detailed introduction is available in [6].

TVP is a generalized version of the projection framework first introduced in [22]. It consists of multiple EMPs. An EMP is a multilinear projection $\{\mathbf{u}^{(1)^T}, \mathbf{u}^{(2)^T}, \ldots, \mathbf{u}^{(N)^T}\}$ composed of one unit projection vector per mode, i.e., $\|\mathbf{u}^{(n)}\| = 1$ for $n = 1, \ldots, N$, where $\|\cdot\|$ is used to indicate the Euclidean norm for vectors. An EMP projects a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ to a scalar y through the N unit projection vectors as

$$y = \mathcal{X} \times_{n=1}^{N} \left\{\mathbf{u}^{(1)^T}, \mathbf{u}^{(2)^T}, \ldots, \mathbf{u}^{(N)^T}\right\} = \mathcal{X} \times_1 \mathbf{u}^{(1)^T} \times_2 \mathbf{u}^{(2)^T} \cdots \times_N \mathbf{u}^{(N)^T} = \langle \mathcal{X}, \mathcal{U} \rangle$$

where $\mathcal{U} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}$.

The TVP of a tensor object $\mathcal{X}$ to a vector $\mathbf{y} \in \mathbb{R}^P$ consists of P EMPs $\{\mathbf{u}_p^{(1)^T}, \mathbf{u}_p^{(2)^T}, \ldots, \mathbf{u}_p^{(N)^T}\}$, $p = 1, \ldots, P$, which can be written concisely as $\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\}_{p=1}^{P}$

$$\mathbf{y} = \mathcal{X} \times_{n=1}^{N} \left\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\right\}_{p=1}^{P} \quad (3)$$

where the pth component of $\mathbf{y}$ is obtained from the pth EMP as: $\mathbf{y}(p) = \mathcal{X} \times_1 \mathbf{u}_p^{(1)^T} \times_2 \mathbf{u}_p^{(2)^T} \cdots \times_N \mathbf{u}_p^{(N)^T}$. The TROD [22] in fact seeks a TVP to maximize the captured variance. However, it takes a heuristic approach. Section III proposes a systematic, more principled formulation by taking consideration of the correlations among features.

III. UNCORRELATED MULTILINEAR PCA

This section introduces the UMPCA framework for unsupervised subspace learning of tensor objects. The UMPCA objective function is first formulated. Then, the successive variance maximization approach and alternating projection method are adopted to derive uncorrelated features through TVP. A methodology to systematically determine the maximum number of uncorrelated features that can be extracted is introduced. Practical issues regarding initialization, projection order, termination, and convergence are addressed and the computational aspects of UMPCA are discussed in detail.

In the presentation, for the convenience of discussion and without loss of generality, training samples are assumed to be zero-mean so that the constraint of uncorrelated features is the same as orthogonal features [20], [31].[1] When the training sample mean is not zero, it can be subtracted to make the training samples to be zero-mean. It should be noted that the orthogonality and zero correlation discussed here are referring to the concepts in linear algebra rather than statistics [32].

A. Formulation of the UMPCA Problem

Following the standard derivation of PCA provided in [20], the variance of the principal components is considered one at a time, starting from the first principal component that targets to

[1]Let $\mathbf{x}$ and $\mathbf{y}$ be vector observations of the variables x and y. Then, $\mathbf{x}$ and $\mathbf{y}$ are orthogonal iff $\mathbf{x}^T\mathbf{y} = 0$, and $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated iff $(\mathbf{x} - \bar{x})^T(\mathbf{y} - \bar{y}) = 0$, where $\bar{x}$ and $\bar{y}$ are the means of $\mathbf{x}$ and $\mathbf{y}$, respectively [32]. Thus, two zero-mean ($\bar{x} = \bar{y} = 0$) vectors are uncorrelated when they are orthogonal [31].

capture the most variance. In the setting of TVP, the pth principal components are $\{\mathbf{y}_m(p), m = 1, \ldots, M\}$, where M is the number of training samples and $\mathbf{y}_m(p)$ is the projection of the mth sample $\mathcal{X}_m$ by the pth EMP $\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\}$: $\mathbf{y}_m(p) = \mathcal{X}_m \times_{n=1}^{N} \{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\}$. Accordingly, the variance is measured by their total scatter $S_{T_p}^{\mathbf{y}}$, which is defined as

$$S_{T_p}^{\mathbf{y}} = \sum_{m=1}^{M} (\mathbf{y}_m(p) - \bar{y}_p)^2 \tag{4}$$

where $\bar{y}_p = (1/M) \sum_m \mathbf{y}_m(p)$. In addition, let $\mathbf{g}_p$ denote the pth coordinate vector, which is the representation of the training samples in the pth EMP space. Its mth component $\mathbf{g}_p(m) = \mathbf{y}_m(p)$. With these definitions, the formal definition of the unsupervised multilinear subspace learning problem in UMPCA is as follows.

A set of M tensor object samples $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ are available for training. Each tensor object $\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ assumes values in the tensor space $\mathbb{R}^{I_1} \bigotimes \mathbb{R}^{I_2} \ldots \bigotimes \mathbb{R}^{I_N}$, where I_n is the n-mode dimension of the tensor and $\bigotimes$ denotes the Kronecker product [33]. The objective of UMPCA is to determine a TVP, which consists of P EMPs $\{\mathbf{u}_p^{(n)} \in \mathbb{R}^{I_n \times 1}, n = 1, \ldots, N\}_{p=1}^{P}$, so that the original tensor space $\mathbb{R}^{I_1} \bigotimes \mathbb{R}^{I_2} \ldots \bigotimes \mathbb{R}^{I_N}$ can be mapped into a vector subspace $\mathbb{R}^P$ (with $P < \prod_{n=1}^{N} I_n$)

$$\mathbf{y}_m = \mathcal{X}_m \times_{n=1}^{N} \left\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\right\}_{p=1}^{P}, \quad m = 1, \ldots, M \tag{5}$$

while the variance of the projected samples, measured by $S_{T_p}^{\mathbf{y}}$, is maximized in each EMP, subject to the constraint that the P coordinate vectors $\{\mathbf{g}_p \in \mathbb{R}^M, p = 1, \ldots, P\}$ are uncorrelated.

In other words, the UMPCA objective is to determine a set of P EMPs $\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\}_{p=1}^{P}$ that maximize the variance captured while producing uncorrelated features. The objective function for determining the pth EMP can be expressed as

$$\left\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\right\} = \arg\max \sum_{m=1}^{M} (\mathbf{y}_m(p) - \overline{y}_p)^2$$
$$\text{subject to } \mathbf{u}_p^{(n)^T} \mathbf{u}_p^{(n)} = 1 \quad \text{and} \quad \frac{\mathbf{g}_p^T \mathbf{g}_q}{\|\mathbf{g}_p\| \|\mathbf{g}_q\|} = \delta_{pq}, \quad p, q = 1, \ldots, P \tag{6}$$

where δ_{pq} is the Kronecker delta defined as

$$\delta_{pq} = \begin{cases} 1, & \text{if } p = q \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The constraint $\mathbf{u}_p^{(n)^T} \mathbf{u}_p^{(n)} = 1$ is imposed since as in the linear case [20], the maximum variance cannot be achieved for finite $\mathbf{u}_p^{(n)}$.

Remark 1: It should be noted that due to the nature of TVP, the UMPCA algorithm is a feature extraction algorithm which produces feature vectors in a manner similar to those of linear solutions. Thus, for the tensor sample $\mathcal{X}$, the corresponding UMPCA feature vector $\mathbf{y}$ is given as

$$\mathbf{y} = \mathcal{X} \times_{n=1}^{N} \left\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\right\}_{p=1}^{P}. \tag{8}$$

B. Derivation of the UMPCA Solution

To solve this UMPCA problem in (6), the successive variance maximization approach, first utilized in the derivation of PCA [20], is taken. The P EMPs $\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\}_{p=1}^{P}$ are determined sequentially in P steps. This stepwise process proceeds as follows.

- Step 1: Determine the first EMP $\{\mathbf{u}_1^{(n)^T}, n = 1, \ldots, N\}$ by maximizing $S_{T_1}^{\mathbf{y}}$.
- Step 2: Determine the second EMP $\{\mathbf{u}_2^{(n)^T}, n = 1, \ldots, N\}$ by maximizing $S_{T_2}^{\mathbf{y}}$ subject to the constraint that $\mathbf{g}_2^T \mathbf{g}_1 = 0$.
- Step 3: Determine the third EMP $\{\mathbf{u}_3^{(n)^T}, n = 1, \ldots, N\}$ by maximizing $S_{T_3}^{\mathbf{y}}$ subject to the constraint that $\mathbf{g}_3^T \mathbf{g}_1 = 0$ and $\mathbf{g}_3^T \mathbf{g}_2 = 0$.
- Step p ($p = 4, \ldots, P$): Determine the pth EMP $\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\}$ by maximizing $S_{T_p}^{\mathbf{y}}$ subject to the constraint that $\mathbf{g}_p^T \mathbf{g}_q = 0$ for $q = 1, \ldots, p-1$.

Fig. 1 lists the pseudocode of the proposed UMPCA algorithm, where $\mathbf{\Psi}_p^{(n)}$ is a matrix to be defined in the formulated eigenvalue problem below, and $\tilde{\mathbf{S}}_{T_p}^{(n)}$ is a total scatter matrix to be defined below. In the figure, the stepwise process described above corresponds to the loop indexed by p. In the following, how to compute these EMPs is presented in detail.

In order to determine the pth EMP $\{\mathbf{u}_p^{(n)^T}, n = 1, \ldots, N\}$, there are N sets of parameters corresponding to the N projection vectors to be determined, $\mathbf{u}_p^{(1)}, \mathbf{u}_p^{(2)}, \ldots \mathbf{u}_p^{(N)}$, one in each mode. It will be desirable to determine these N sets of parameters (N projection vectors) in all modes simultaneously so that $S_{T_p}^{\mathbf{y}}$ is globally maximized, subject to the zero-correlation constraint. Unfortunately, as in other multilinear subspace learning algorithms [6], [9], [18], there is no closed-form solution for this problem, except when $N = 1$, which is the classical PCA where only one projection vector is to be solved. Therefore, following the heuristic approaches in [6], [9], and [18], the alternating projection method is used to solve the multilinear problem, as described below.

To determine the pth EMP, the N sets of parameters for the N projection vectors are estimated one mode (i.e., one set) at a time. For mode n^*, a linear subproblem in terms of $\mathbf{u}_p^{(n^*)}$ is solved by fixing $\{\mathbf{u}_p^{(n)}, n \neq n^*\}$, the projection vectors for the other modes. Thus, there are N such conditional subproblems, corresponding to the loop indexed by n in Fig. 1. This process iterates until a stopping criterion is met, which corresponds to the loop indexed by k in Fig. 1.

To solve for $\mathbf{u}_p^{(n^*)}$, the conditional subproblem in the n^*-mode, the tensor samples are projected in these $(N-1)$ modes $\{n \neq n^*\}$ first to obtain the vectors

$$\tilde{\mathbf{y}}_m^{(n^*)}(p) = \mathcal{X}_m \times_1 \mathbf{u}_p^{(1)^T} \cdots \times_{n^*-1} \mathbf{u}_p^{(n^*-1)^T} \times_{n^*+1} \mathbf{u}_p^{(n^*+1)^T} \cdots \times_N \mathbf{u}_p^{(N)^T} \tag{9}$$

where $\tilde{\mathbf{y}}_m^{(n^*)}(p) \in \mathbb{R}^{I_{n^*}}$, assuming that $\{\mathbf{u}_p^{(n)}, n \neq n^*\}$ is given. This conditional subproblem then becomes to determine $\mathbf{u}_p^{(n^*)}$ that projects the vector samples $\{\tilde{\mathbf{y}}_m^{(n^*)}(p), m = 1, \ldots, M\}$ onto a line so that the variance is maximized, subject to the zero-correlation constraint, which is a PCA problem with the

Input: A set of tensor samples $\{\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}, m = 1, ..., M\}$, the desired feature vector length P, the maximum number of iterations K, and a small number η for testing convergence.

Output: The P EMPs $\{\mathbf{u}_p^{(n)^T}, n = 1, ..., N\}_{p=1}^{P}$ that captures the most variance in the projected space.

Algorithm:

For $p = 1 : P$ (step p: determine the pth EMP)

If $p > 1$, calculate the coordinate vector $\mathbf{g}_{p-1}$: $\mathbf{g}_{p-1}(m) = \mathcal{X}_m \times_1 \mathbf{u}_{p-1}^{(1)^T} \times_2 \mathbf{u}_{p-1}^{(2)^T} \ldots \times_N \mathbf{u}_{p-1}^{(N)^T}$.

- For $n = 1, ..., N$, initialize $\mathbf{u}_{p_{(0)}}^{(n)} \in \mathbb{R}^{I_n}$.
- For $k = 1 : K$
  - For $n = 1 : N$
    - Calculate $\tilde{\mathbf{y}}_m^{(n)}(p) = \mathcal{X}_m \times_1 \mathbf{u}_{p_{(k)}}^{(1)^T} \ldots \times_{n-1} \mathbf{u}_{p_{(k)}}^{(n-1)^T} \times_{n+1} \mathbf{u}_{p_{(k-1)}}^{(n+1)^T} \ldots \times_N \mathbf{u}_{p_{(k-1)}}^{(N)^T}$, for $m = 1, ..., M$.
    - Calculate $\mathbf{\Psi}_p^{(n)}$ and $\tilde{\mathbf{S}}_{T_p}^{(n)}$. Set $\mathbf{u}_{p_{(k)}}^{(n)}$ to be the (unit) eigenvector of $\mathbf{\Psi}_p^{(n)} \tilde{\mathbf{S}}_{T_p}^{(n)}$ associated with the largest eigenvalue.
  - If $k = K$ or $(S_{T_{p_{(k)}}}^{\mathbf{y}} - S_{T_{p_{(k-1)}}}^{\mathbf{y}})/S_{T_{p_{(k-1)}}}^{\mathbf{y}} < \eta$, set $\mathbf{u}_p^{(n)} = \mathbf{u}_{p_k}^{(n)}$ for all n and break.
- **Output** $\{\mathbf{u}_p^{(n)}\}$. Go the step $p + 1$ if $p < P$. Stop if $p = P$.

Fig. 1. Pseudocode implementation of the UMPCA algorithm for unsupervised subspace learning of tensor objects.

input samples $\{\tilde{\mathbf{y}}_m^{(n^*)}(p), m = 1, \ldots, M\}$. The total scatter matrix $\tilde{\mathbf{S}}_{T_p}^{(n^*)}$ corresponding to $\{\tilde{\mathbf{y}}_m^{(n^*)}(p), m = 1, \ldots, M\}$ is then defined as

$$\tilde{\mathbf{S}}_{T_p}^{(n^*)} = \sum_{m=1}^{M} \left(\tilde{\mathbf{y}}_m^{(n^*)}(p) - \bar{\tilde{\mathbf{y}}}_p^{(n^*)}\right) \left(\tilde{\mathbf{y}}_m^{(n^*)}(p) - \bar{\tilde{\mathbf{y}}}_p^{(n^*)}\right)^T \quad (10)$$

where $\bar{\tilde{\mathbf{y}}}_p^{(n^*)} = (1/M) \sum_m \tilde{\mathbf{y}}_m^{(n^*)}(p)$. With (10), it is now ready to determine the projection vectors. For $p = 1$, the $\mathbf{u}_1^{(n)}$ that maximizes the total scatter $\mathbf{u}_1^{(n)^T} \tilde{\mathbf{S}}_{T_1}^{(n)} \mathbf{u}_1^{(n)}$ in the projected space is obtained as the unit eigenvector of $\tilde{\mathbf{S}}_{T_1}^{(n)}$ associated with the largest eigenvalue, for $n = 1, \ldots, N$. For $p = 2, \ldots, P$, given the first $(p-1)$ EMPs, the pth EMP aims to maximize the total scatter $S_{T_p}^{\mathbf{y}}$, subject to the constraint that features projected by the pth EMP are uncorrelated with those projected by the first $(p-1)$ EMPs. Let $\tilde{\mathbf{Y}}_p^{(n^*)} \in \mathbb{R}^{I_{n^*} \times M}$ be a matrix with $\tilde{\mathbf{y}}_m^{(n^*)}(p)$ as its mth column, i.e.,

$$\tilde{\mathbf{Y}}_p^{(n^*)} = \left[\tilde{\mathbf{y}}_1^{(n^*)}(p), \tilde{\mathbf{y}}_2^{(n^*)}(p), \ldots, \tilde{\mathbf{y}}_M^{(n^*)}(p)\right] \quad (11)$$

then the pth coordinate vector is $\mathbf{g}_p = \tilde{\mathbf{Y}}_p^{(n^*)^T} \mathbf{u}_p^{(n^*)}$. The constraint that $\mathbf{g}_p$ is uncorrelated with $\{\mathbf{g}_q, q = 1, \ldots, p-1\}$ can be written as

$$\mathbf{g}_p^T \mathbf{g}_q = \mathbf{u}_p^{(n^*)^T} \tilde{\mathbf{Y}}_p^{(n^*)} \mathbf{g}_q = 0, \qquad q = 1, \ldots, p-1. \quad (12)$$

Thus, $\mathbf{u}_p^{(n^*)}$ $(p > 1)$ can be determined by solving the following constrained optimization problem:

$$\mathbf{u}_p^{(n^*)} = \arg\max \mathbf{u}_p^{(n^*)^T} \tilde{\mathbf{S}}_{T_p}^{(n^*)} \mathbf{u}_p^{(n^*)}$$
$$\text{subject to } \mathbf{u}_p^{(n^*)^T} \mathbf{u}_p^{(n^*)} = 1 \quad \text{and} \quad \mathbf{u}_p^{(n^*)^T} \tilde{\mathbf{Y}}_p^{(n^*)} \mathbf{g}_q = 0, \quad q = 1, \ldots, p-1. \quad (13)$$

The solution is given by the following theorem.

Theorem 1: The solution to the problem in (13) is the (unit-length) eigenvector corresponding to the largest eigenvalue of the following eigenvalue problem:

$$\mathbf{\Psi}_p^{(n^*)} \tilde{\mathbf{S}}_{T_p}^{(n^*)} \mathbf{u} = \lambda \mathbf{u} \quad (14)$$

where

$$\mathbf{\Psi}_p^{(n^*)} = \mathbf{I}_{I_{n^*}} - \tilde{\mathbf{Y}}_p^{(n^*)} \mathbf{G}_{p-1} \mathbf{\Phi}_p^{-1} \mathbf{G}_{p-1}^T \tilde{\mathbf{Y}}_p^{(n^*)^T} \quad (15)$$

$$\mathbf{\Phi}_p = \mathbf{G}_{p-1}^T \tilde{\mathbf{Y}}_p^{(n^*)^T} \tilde{\mathbf{Y}}_p^{(n^*)} \mathbf{G}_{p-1} \quad (16)$$

$$\mathbf{G}_{p-1} = [\mathbf{g}_1 \quad \mathbf{g}_2 \quad \ldots \mathbf{g}_{p-1}] \in \mathbb{R}^{M \times (p-1)} \quad (17)$$

and $\mathbf{I}_{I_{n^*}}$ is an identity matrix of size $I_{n^*} \times I_{n^*}$.

Proof: The proof of Theorem 1 is given in Appendix I. ■

By setting $\mathbf{\Psi}_1^{(n^*)} = \mathbf{I}_{I_{n^*}}$ and from Theorem 1, a unified solution for UMPCA is obtained: for $p = 1, \ldots, P$, $\mathbf{u}_p^{(n^*)}$ is obtained as the unit eigenvector of $\mathbf{\Psi}_p^{(n^*)} \tilde{\mathbf{S}}_{T_p}^{(n^*)}$ associated with the largest eigenvalue. As pointed out earlier, this solution is an approximate, suboptimal solution to the original formulation in (6), due to the heuristics employed to tackle the problem.

C. Determination of the Maximum Number of Extracted Uncorrelated Features

It should be pointed out that compared to PCA and its existing multilinear extensions, UMPCA has a possible limitation in the number of (uncorrelated) features that can be extracted. This is because the projection to be solved in UMPCA is highly constrained, in both their correlation property and the simplicity of the projection. Compared with PCA, the projection in UMPCA is a TVP rather than a linear projection from vector to vector. Thus, the number of parameters to be estimated in UMPCA is usually much smaller than that in PCA, as discussed in [19]. Compared to 2DPCA, CSA, TROD, and MPCA, the correlations among features extracted by UMPCA have to be zero. The

maximum number of features that can be extracted by UMPCA is given by the following corollary.

Corollary 1: The number of uncorrelated features that can be extracted by UMPCA P is upper-bounded by $\min\{\min_n I_n, M\}$, i.e., $P \leq \min\{\min_n I_n, M\}$, provided that the elements of $\tilde{\mathbf{Y}}_p^{(n)}$ are not all zero, where I_n is the n-mode dimensionality and M is the number of training samples.

Proof: The proof of Corollary 1 is given in Appendix II. ■

Corollary 1 implies that UMPCA may be more appropriate for high-resolution tensor objects where the dimensionality in each mode is high enough to result in a larger $\min_n I_n$ and enable the extraction of sufficient number of (uncorrelated) features. It also indicates that UMPCA is more suitable for applications that need only a small number (e.g., $\leq \min\{\min_n I_n, M\}$) of features, such as clustering of a small number of classes.

In addition, it is interesting to examine Corollary 1 for the linear case, i.e., for $N = 1$. Since UMPCA follows the approach of successive variance maximization in the classical derivation of PCA [20], when $N = 1$, the samples are vectors $\{\mathbf{x}_m \in \mathbb{R}^{I_1}\}$ and UMPCA reduces to PCA. Accordingly, in each step p, there is only one projection vector $\mathbf{u}_p$ to be solved to maximize the captured variance, subject to the zero-correlation constraint. Corollary 1 indicates that the maximum number of extracted features does not exceed $\min\{I_1, M\}$ when $N = 1$. This is exactly the case for PCA, where the number of PCA features cannot exceed the rank of the data matrix, which is upper-bounded by the minimum of the dimension of the samples I_1 and the number of samples M.

D. Initialization, Projection Order, Termination, and Convergence

This section discusses UMPCA design issues, including the initialization procedure, the projection order, termination conditions, as well as issues related to the convergence of the solution.

Due to the multilinear nature of UMPCA, the determination of each EMP $\{\mathbf{u}_p^{(n)}, n = 1, \ldots, N\}$ is iterative in nature. Since solving the projection vector in one mode requires the projection vectors in all the other modes, initial estimations for the projection vectors $\{\mathbf{u}_p^{(n)}\}$ are necessary. However, as in [6], [24], and [34]–[36], determining the optimal initialization in UMPCA is still an open problem. This work empirically studies two commonly used initialization methods [6], [19], [22], [35], [36]: uniform initialization and random initialization. The uniform initialization initializes each n-mode projection vector to the all ones vector $\mathbf{1}$, with proper normalization to have unit length. The random initialization draws each element of the n-mode projection vectors randomly from a zero-mean uniform distribution between $[-0.5, 0.5]$, with normalization to have unit length as well. The experimental results reported in Section IV indicate that the results of UMPCA are affected by initialization, and the uniform initialization gives more stable results.

The mode ordering (the loop indexed by n in Fig. 1) in computing the projection vectors affects performance. The optimal way to determine the projection order is an open research problem. Simulation studies on the effects of the projection order indicate that different projection order results in different amount of captured variance. However, there is no guidance either from the problem, the data, or the algorithm on the best projection order. Hence, there is no preference on a particular projection order, and the projection vectors are solved sequentially (from 1-mode to N-mode), in a fashion similar to the one used in [6], [9], and [18].

As can be seen in Fig. 1, the iterative procedure terminates when $(S^{\mathbf{y}}_{T_{p_{(k)}}} - S^{\mathbf{y}}_{T_{p_{(k-1)}}})/S^{\mathbf{y}}_{T_{p_{(k-1)}}} < \eta$, where $S^{\mathbf{y}}_{T_{p_{(k)}}}$ is the total scatter captured by the pth EMP obtained in the kth iteration of UMPCA and η is a small number threshold. Alternatively, the convergence of the projection vectors can also be tested: $\mathrm{dist}(\mathbf{u}^{(n)}_{p_{(k)}}, \mathbf{u}^{(n)}_{p_{(k-1)}}) < \epsilon$, where

$$\mathrm{dist}\left(\mathbf{u}^{(n)}_{p_{(k)}}, \mathbf{u}^{(n)}_{p_{(k-1)}}\right) = \min\left(\left\|\mathbf{u}^{(n)}_{p_{(k)}} + \mathbf{u}^{(n)}_{p_{(k-1)}}\right\|, \left\|\mathbf{u}^{(n)}_{p_{(k)}} - \mathbf{u}^{(n)}_{p_{(k-1)}}\right\|\right) \quad (18)$$

and ϵ is a user-defined small number threshold (e.g., $\epsilon = 10^{-3}$). Section IV indicates that the variance captured by a particular EMP usually increases rapidly for the first few iterations and slowly afterwards. Therefore, the iterative procedures in UMPCA can be terminated by simply setting a maximum number of iterations K in practice for convenience, especially when the computational cost is a concern.

Regarding convergence, the derivation of Theorem 1 (the end of Appendix I) implies that per iteration, the scatter $S^{\mathbf{y}}_{T_p}$ is a nondecreasing function since each update of the projection vector $\mathbf{u}_p^{(n^*)}$ in a given mode n^* maximizes $S^{\mathbf{y}}_{T_p}$. On the other hand, $S^{\mathbf{y}}_{T_p}$ is upper-bounded by the variation in the original samples, following similar argument in [9]. Therefore, UMPCA is expected to convergence over iterations, following [23, Th. 1] and [24]. Empirical results presented in Section IV indicate that UMPCA converges within ten iterations for typical tensor objects in practice. In addition, when the largest eigenvalues in each mode are with multiplicity 1, the projection vectors $\{\mathbf{u}_p^{(n)}\}$, which maximize the objective function $S^{\mathbf{y}}_{T_p}$, are expected to converge as well, where the convergence is up to sign. Simulation studies show that the projection vectors $\{\mathbf{u}_p^{(n)}\}$ do converge over a number of iterations.

E. Computational Aspects of UMPCA

Finally, the computational aspects of UMPCA are considered here. Specifically, the computational complexity and memory requirements are analyzed, following the framework used in [9] for MPCA, and in [6] for UMLDA. It is assumed again that $I_1 = I_2 = \cdots = I_N = (\prod_{n=1}^{N} I_n)^{(1/n)} = I$ for simplicity.

The most computationally demanding steps in UMPCA are the calculations of the projection $\tilde{\mathbf{y}}_m^{(n)}(p)$, the computation of $\tilde{\mathbf{S}}^{(n)}_{T_p}$ and $\mathbf{\Psi}_p^{(n)}$, and the calculation of the leading eigenvector of $\mathbf{\Psi}_p^{(n)}\tilde{\mathbf{S}}^{(n)}_{T_p}$. The complexity of calculating $\tilde{\mathbf{y}}_m^{(n)}(p)$ for $m = 1, \ldots, M$ and $\tilde{\mathbf{S}}^{(n)}_{T_p}$ are in order of $O(M \cdot \sum_{n=2}^{N} I^n)$ and $O(M \cdot I^2)$, respectively. The computation of $\mathbf{\Psi}_p^{(n)}$ is in the order of

$$O\left(I \cdot M \cdot (p-1) + (p-1) \cdot I \cdot (p-1) + (p-1)^3 + I \cdot (p-1)^2 + I \cdot (p-1) \cdot I\right) = O\left((p-1) \cdot \left[I \cdot M + 2 \cdot I \cdot (p-1) + (p-1)^2 + I^2\right]\right). \quad (19)$$

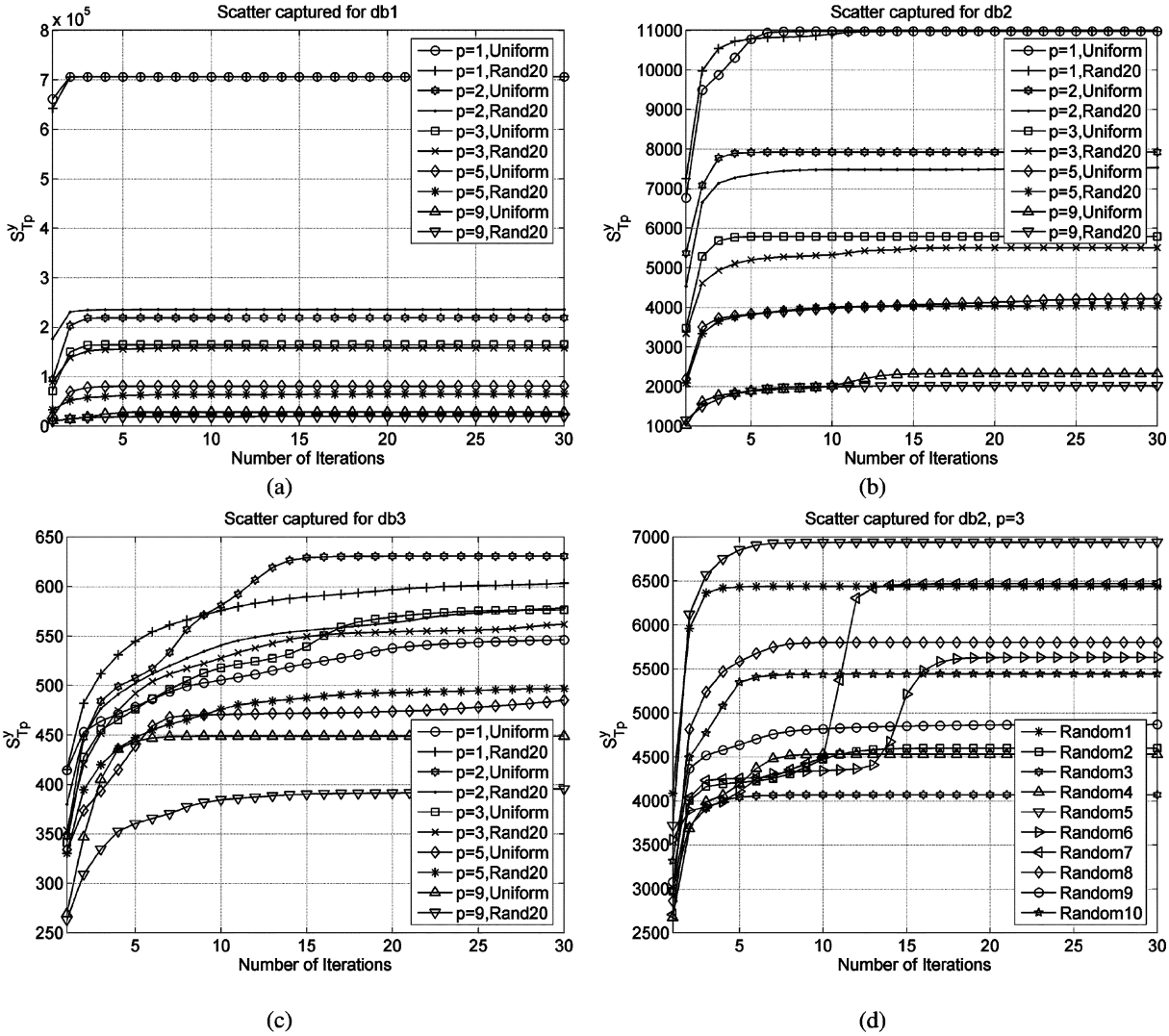


Fig. 2. Illustration of the effects of initialization on the scatter captured by UMPCA. Comparison of the captured $S^{\mathbf{y}}_{T_p}$ with uniform and random initialization (average of 20 repetitions) over 30 iterations for $p = 1, 2, 3, 5, 9$ on synthetic data set (a) db1, (b) db2, and (c) db3. (d) Illustration of the captured $S^{\mathbf{y}}_{T_p}$ of ten random initializations for $p = 3$ on db2.

The computation of $\mathbf{\Psi}_p^{(n)}\tilde{\mathbf{S}}_{T_p}^{(n)}$ and its eigendecomposition[2] are both of order $O(I^3)$. Therefore, the computational complexity per mode n for one iteration k of step p is

$$O\left(M\left[I^2+\sum_{n=2}^{N} I^n\right]+(p-1)\right.$$
$$\left.\times\left[I\cdot M+2I(p-1)+(p-1)^2+I^2\right]+2I^3\right). \quad (20)$$

Regarding the memory requirement, the respective computation can be done incrementally by reading $\mathcal{X}_m$ sequentially. Thus, except for $N = 1$, the memory needed for the UMPCA algorithm can be as low as $O(I^N)$, although sequential reading will lead to higher input/output (I/O) cost.

From Fig. 1 and the discussions above, as a sequential iterative solution, UMPCA may have a high computational and I/O cost. Nevertheless, solving the UMPCA projection is only in the training phase of the targeted pattern recognition tasks, so it can be done offline and the additional computational and I/O costs due to iterations and sequential processing are not considered a disadvantage. During testing, feature extraction from a test sample is an efficient linear operation, as in conventional linear subspace learning algorithms.

[2]Since only the largest eigenvalue and the corresponding eigenvector are needed in UMPCA, more efficient computational methods may be applied in practice.

IV. Experimental Studies

This section presents a number of experiments carried out in support of the following objectives.

1) Investigate the various properties of the UMPCA algorithm.
2) Demonstrate the utility of the UMPCA algorithm in typical learning applications by comparing the UMPCA recognition performance against that of the baseline PCA solution and its state-of-the-art multilinear extensions on two recognition problems involving tensorial data, namely, face and gait recognition.

A. Study of UMPCA Properties on Synthetic Data

The following properties of UMPCA are studied here: 1) the effects of the initialization procedure, 2) the effects of the projection order, and 3) the convergence of the algorithm. Similar to MPCA [9], UMPCA is an unsupervised algorithm derived under the variance maximization principle and eigendecomposition in each mode (Theorem 1) is an important step in computing the UMPCA projection. Thus, its properties are affected by the eigenvalue distribution of the input data. Three third-order tensorial data sets, namely, db1, db2, and db3, have been synthetically generated in [9], with eigenvalues in each mode spanning different magnitude ranges. In this work, experimental studies of the UMPCA properties are performed on these three synthetic data sets, with a brief description of the generation process included below.

For each set, M samples $\mathcal{A}_m \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ are generated according to $\mathcal{A}_m = \mathcal{B}_m \times_1 \mathbf{C}^{(1)} \times_2 \mathbf{C}^{(2)} \times_3 \mathbf{C}^{(3)} + \mathcal{D}_m$, where $\mathcal{B}_m \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, $\mathbf{C}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ $(n = 1,2,3)$ and $\mathcal{D}_m \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. All entries in $\mathcal{B}_m$ are drawn from a zero-mean unit-variance Gaussian distribution and multiplied by $(I_1 \cdot I_2 \cdot I_3 / i_1 \cdot i_2 \cdot i_3)^f$, where f controls the eigenvalue distributions. $\mathbf{C}^{(n)}$ $(n = 1,2,3)$ are orthogonal matrices obtained from applying singular value decomposition (SVD) on random matrices with entries drawn from zero-mean, unit-variance Gaussian distribution. All entries of $\mathcal{D}_m$ are from a zero-mean Gaussian distribution with variance 0.01. Three synthetic data sets, db1, db2, and db3, of size $30 \times 20 \times 10$ with $M = 100$ and $f = 1/2, 1/4$, and $1/16$, respectively, are generated, where a smaller f results in a narrower range of eigenvalue spread and vice versa. Practical data such as face and gait data share similar characteristics with the data in db1 [9].

1) The Effects of Initialization: The effects of initialization are studied first. The uniform initialization and random initialization discussed in Section III-D are tested up to 30 iterations, with the projection order fixed. Fig. 2 shows the simulation results on the three synthetic data sets. The results shown for random initialization in Fig. 2(a)–(c) are the average of 20 repeated trials. From Fig. 2(a) and (b), it can be seen that for $p = 1$, both the uniform and random initializations result in the same $S^{\mathbf{y}}_{T_p}$. For $p > 1$, two ways of initialization lead to different $S^{\mathbf{y}}_{T_p}$, with the uniform initialization performing better (i.e., results in larger $S^{\mathbf{y}}_{T_p}$) on db2. In addition, it should be noted that for db1 and db2, $S^{\mathbf{y}}_{T_p}$ decreases as p increases, which is expected since maximum variation should be captured in each EMP subject to the zero-correlation constraint. From the figures, the algorithm converges in around 5 and 15 iterations for db1 and db2, respectively. For db3 in Fig. 2(c), the uniform and random initializations do not result in the same $S^{\mathbf{y}}_{T_p}$ even for $p = 1$ and $S^{\mathbf{y}}_{T_p}$ does not always decrease as p increases, which indicates that some EMPs fail to capture the maximum variation. It may be partly explained by observing from Fig. 2(c) that the algorithm converges slowly and 30 iterations may not be sufficient to reach convergence (i.e., maximize the captured variation).

Fig. 2(d) further shows some typical results of the evolution of $S^{\mathbf{y}}_{T_p}$ for ten random initializations on db2 with $p = 3$. As seen from the figure, the results obtained from random initialization have high variance. Although when computational cost is not a concern, a number of random initializations can be tested to choose the one results in the best performance, i.e., the largest $S^{\mathbf{y}}_{T_p}$, the uniform initialization is a safe choice when testing several initializations is not desirable. Thus, the uniform initialization is used in all the following experiments for UMPCA.

2) The Effects of Projection Order: Next, the effects of the projection order are tested, with representative results shown in Fig. 3 for $p = 1,2$ on the three synthetic data sets. As shown in the figure, the projection order affects UMPCA as well, except for $p = 1$ on db1 and db2. Nonetheless, no one particular projection order consistently outperforms all the others. Thus, in the following experiments, the projection order is fixed to be sequential from 1 to N. As in initialization, if computational cost is not a concern, all possible projection orders could be tested and the one resulting in the largest $S^{\mathbf{y}}_{T_p}$ should be considered for each p.

3) Convergence Studies: Last, experimental studies are performed on the convergence of the total scatter captured in each EMP and the convergence of the corresponding projection vectors in each mode. Fig. 4 depicts the evolution of the captured total scatter and the two-mode projection vector difference over 50 iterations for $p = 1, \ldots, 10$. From Fig. 4(a), (c), and (e), it can be observed that the algorithm converges (in terms of the total scatter) on db1 and db2 in about 10 and 30 iterations, respectively, while on db3, the convergence speed is considerably lower, indicating again the difficulty of db3. Fig. 4(b), (d), and (f) demonstrates that the derived projection vectors converge too. It is also observed that the convergence speed of the projection vectors on db3 is much lower than the convergence speed on the other two data sets. This is likely due to the narrow range of eigenvalue spread in db3.

B. Evaluation on Face and Gait Recognition

The proposed UMPCA is evaluated on two unsupervised recognition tasks, namely, face recognition [37] and gait recognition [38], which can be considered as second-order and third-order tensor biometric classification problems, respectively. These two recognition tasks have practical importance in security-related applications such as biometric authentication and surveillance [39]. The evaluation is through performance comparison against the baseline PCA solution and its existing multilinear extensions.

1) The Face and Gait Data: The facial recognition technology (FERET) database [40], a standard testing database for face recognition performance evaluation, includes 14 126 images from 1199 individuals covering a wide range of variations in viewpoint, illumination, facial expression, races, and ages. The FERET subset selected in this experimental evaluation consists of those subjects that have at least eight images in the database with at most 15° of pose variation. Thus, 721 face images from 70 FERET subjects are considered. Since the focus here is on recognition rather than detection, all face images are manually cropped, aligned (with manually annotated coordinate information of eyes), and normalized to 80×80 pixels, with 256 gray levels per pixel. Fig. 5(a) depicts sample face images from a subject in this FERET subset.

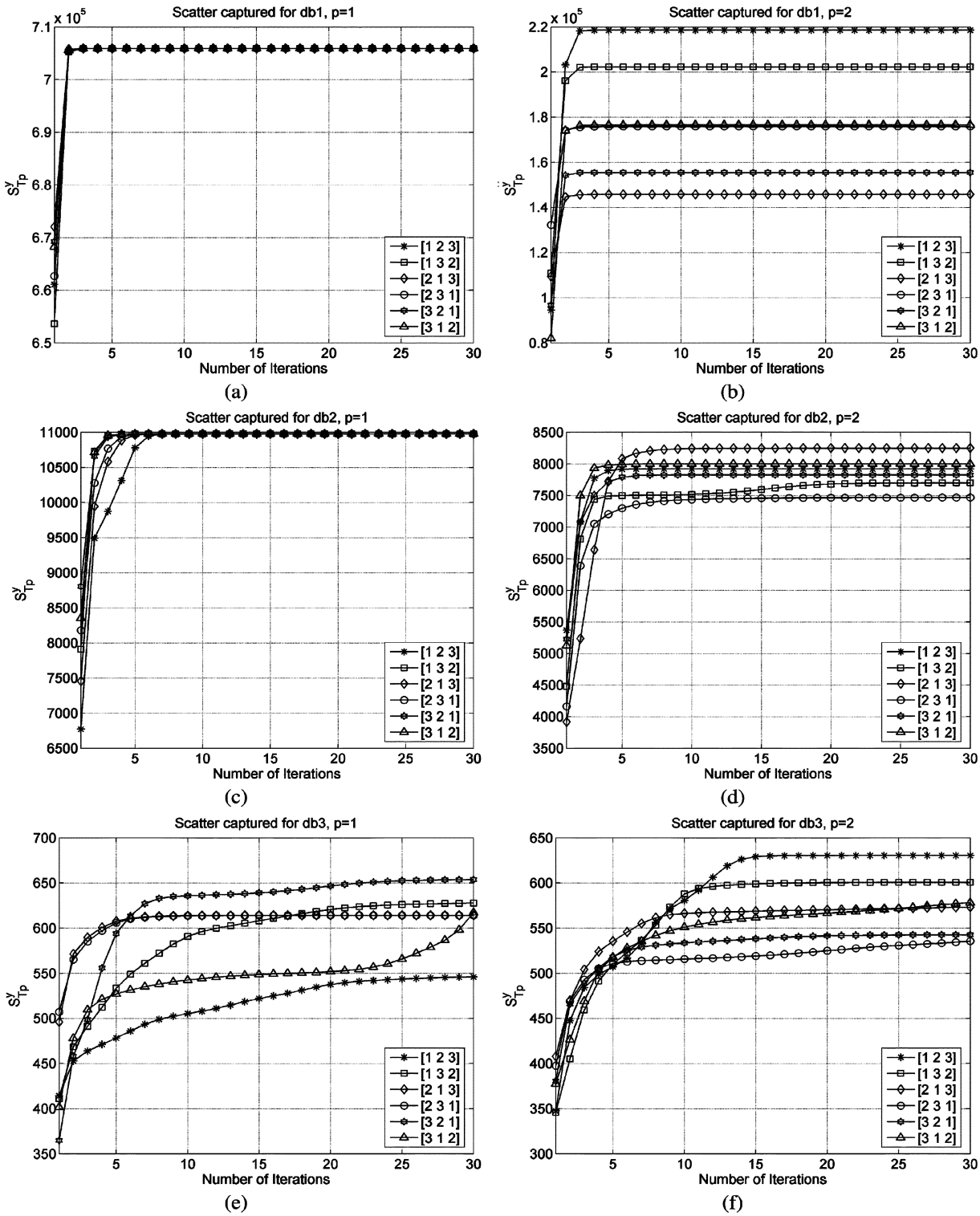


Fig. 3. Illustration of the effects of projection order on the scatter captured by UMPCA: on db1 with (a) $p = 1$ and (b) $p = 2$, on db2 with (c) $p = 1$ and (d) $p = 2$, and on db3 with (e) $p = 1$ and (f) $p = 2$.

The University of South Florida (USF) gait challenge data set version 1.7 consists of 452 sequences from 74 subjects walking in elliptical paths in front of the camera, with two viewpoints (left or right), two shoe types (A or B), and two surface types (grass or concrete). There are seven probes in this data set and probe A is chosen for the evaluation, with difference in viewpoint only. Thus, the gallery set is used as the training set and it is captured on grass surface, with shoe type A and from the right view, and probe A is used as the test set and it is captured on grass surface, with shoe type A and from the left view. Each set has only one sequence for a subject. Subjects are unique in the gallery and probe sets and there are no common sequences

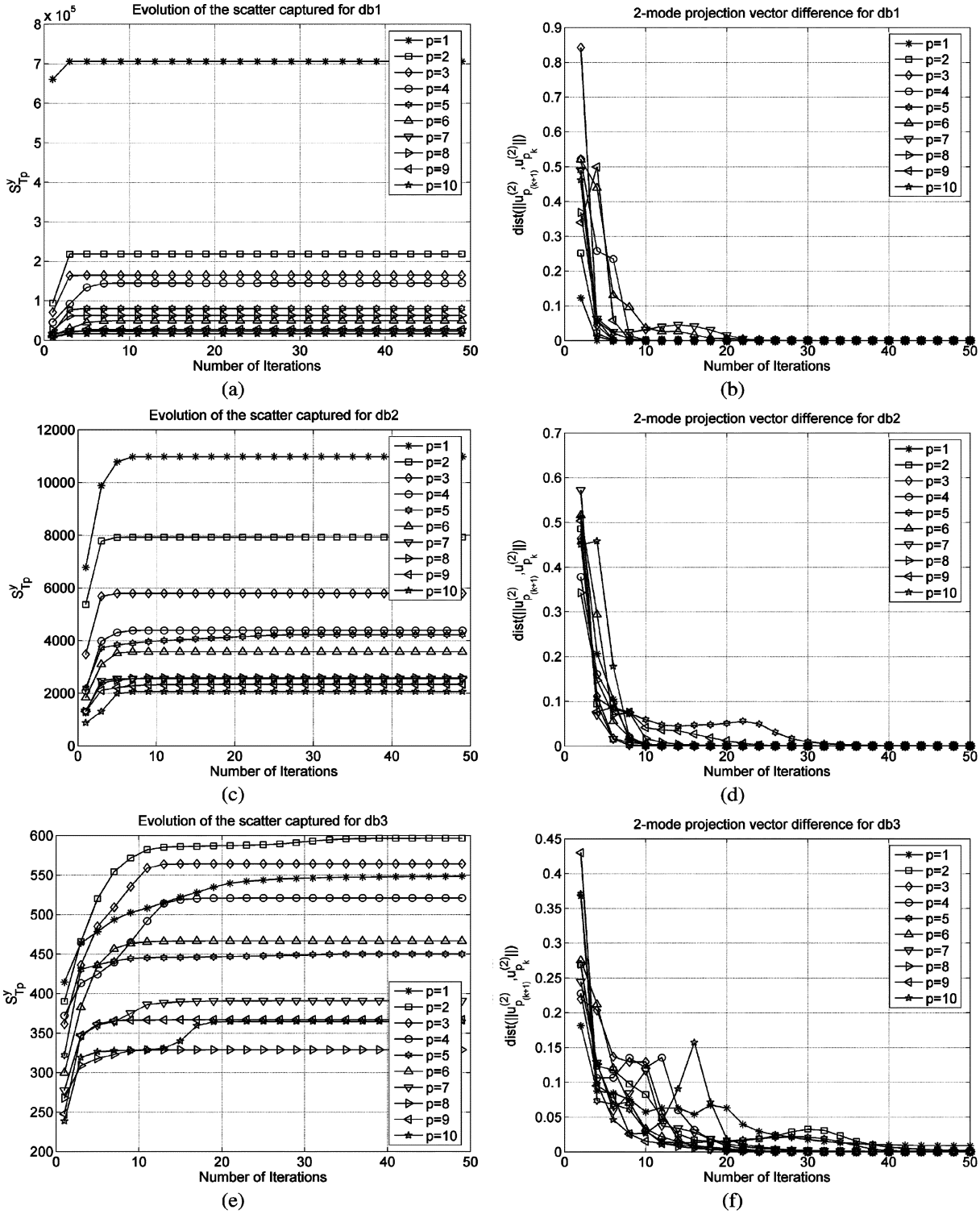


Fig. 4. Illustration of the convergence of UMPCA: the evolution of the total scatter captured on (a) db1, (c) db2, and (e) db3; and the evolution of $\mathrm{dist}(\mathbf{u}^{(2)}_{p(k)}, \mathbf{u}^{(2)}_{p(k-1)})$ on (b) db1, (d) db2, and (f) db3.

between the gallery set and the probe set. These gait data sets are employed to demonstrate the performance on third-order tensors since gait silhouette sequences are naturally 3-D data [9]. The procedures in [9] are followed to get gait samples from gait silhouette sequences and each gait sample is resized to a third-order tensor of $32 \times 22 \times 10$. There are 731 and 727 gait samples in the gallery set and probe A, respectively. Fig. 5(b) shows three gait samples from the USF gait database.

2) Algorithms and Their Settings in Performance Comparison: In the face and gait recognition experiments, the performance of the proposed UMPCA is compared against that of PCA [20], [21] and five existing multilinear PCA extensions,

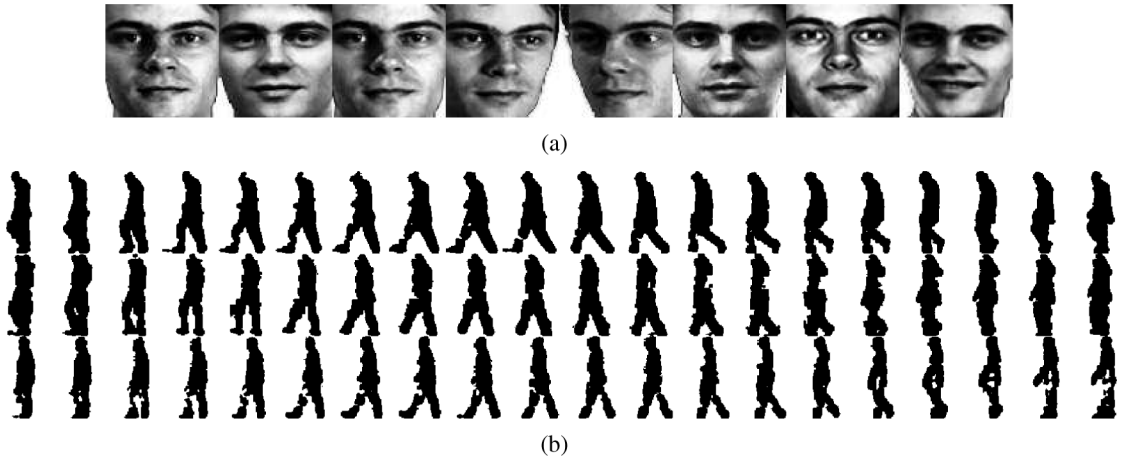


Fig. 5. Examples of the biometric data used in performance evaluation: (a) eight face samples from a subject in the FERET subset used, and (b) three gait samples from the USF gait database V.1.7, shown by concatenating frames in rows.

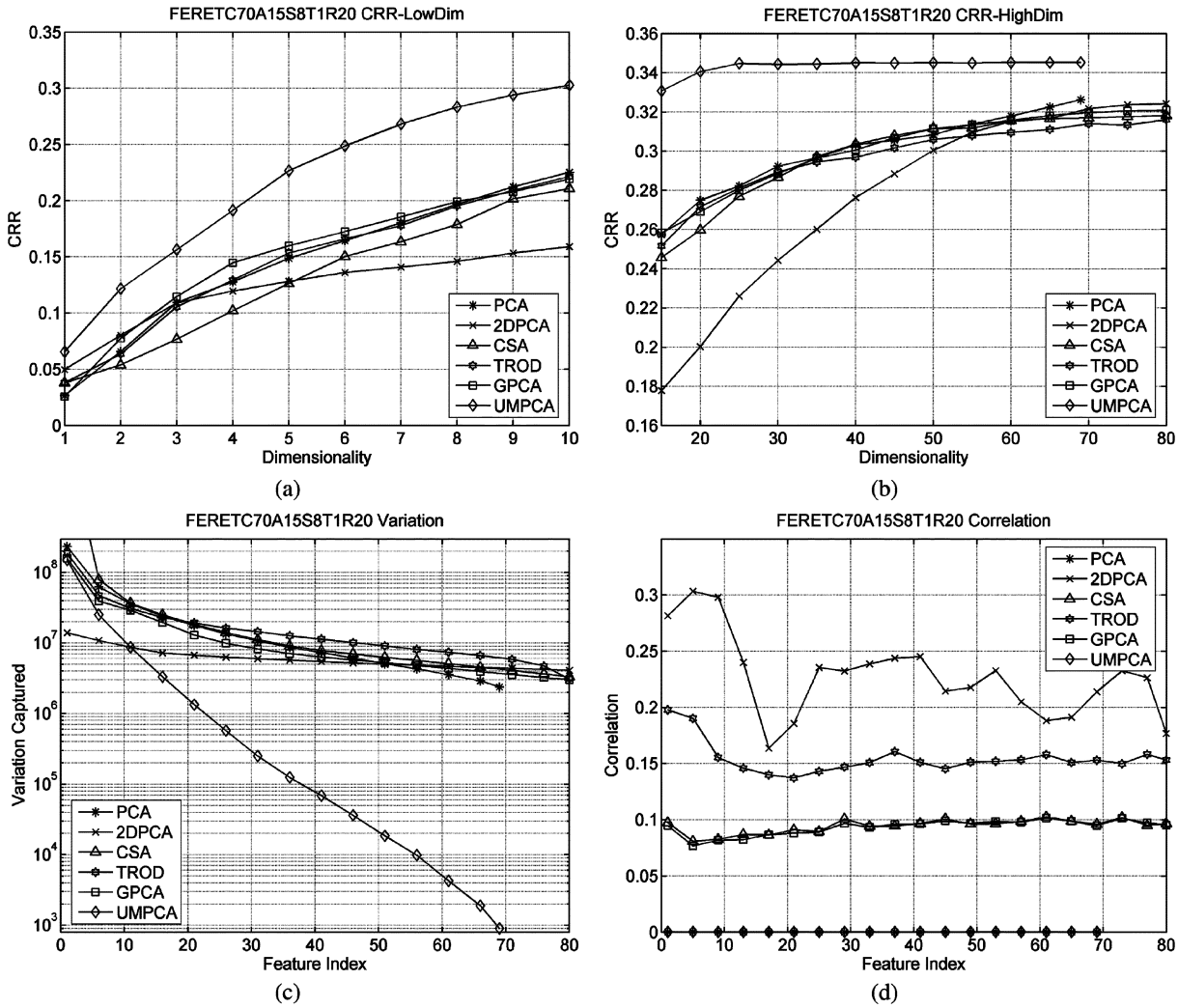


Fig. 6. Detailed face recognition results by PCA algorithms on the FERET database for $L = 1$: (a) performance curves for the low-dimensional projection case, (b) performance curves for the high-dimensional projection case, (c) the variation captured by individual features, and (d) the correlation among features.

2DPCA [17], CSA [18], TROD [22], GPCA [8], and MPCA [9]. It should be noted that 2DPCA and GPCA are only applied to the face recognition problem since it cannot handle the third-order tensors in gait recognition. In addition, 2DSVD [27] is used for initialization in GPCA so the GPCA algorithm tested is equivalent to MPCA with $N = 2$.

The recognition performance is evaluated by the identification rate calculated through similarity measurement between

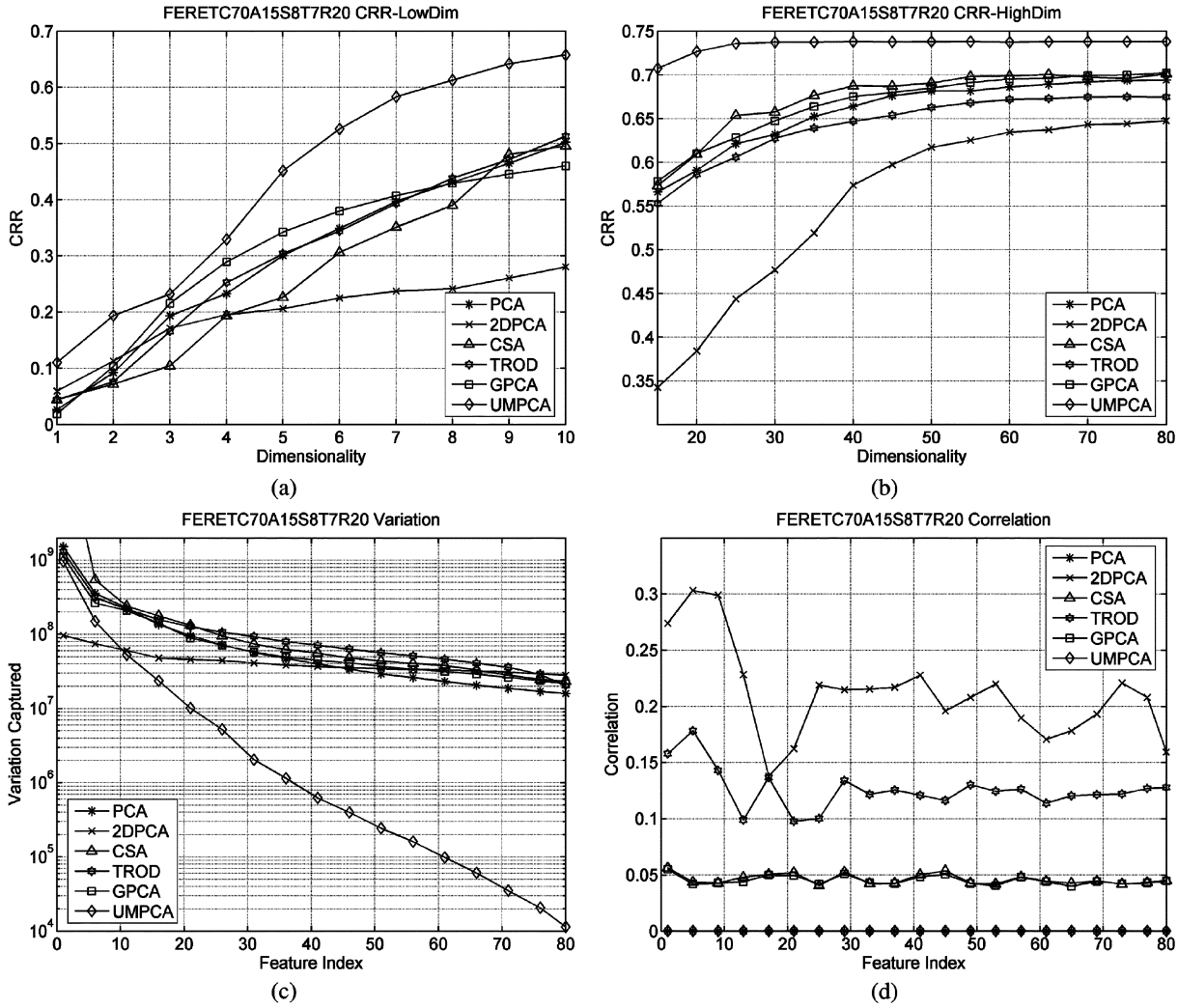


Fig. 7. Detailed face recognition results by PCA algorithms on the FERET database for $L = 7$: (a) performance curves for the low-dimensional projection case, (b) performance curves for the high-dimensional projection case, (c) the variation captured by individual features, and (d) the correlation among features.

feature vectors. The simple nearest neighbor classifier with Euclidean distance measure is used for classification of extracted features since the focus of this paper is on feature extraction. Among the algorithms considered here, 2DPCA, CSA, GPCA, and MPCA produce tensorial features, which need to be vectorized for classification. Hence, for these four algorithms, each entry in the projected tensorial features is viewed as an individual feature and the corresponding total scatter as defined in (4) is calculated. The tensorial features produced by these methods are then arranged into a vector in descending total scatter. All the iterative algorithms (CSA, TROD, GPCA, MPCA, and UMPCA) are terminated by setting the maximum number of iterations K for fair comparison and computational concerns. Since CSA, GPCA, and MPCA have very good convergence performance, K is set to 1. For TROD and UMPCA, K is set to 10 and the uniform initialization is used. Due to Corollary 1, up to 80 features are tested in the face recognition experiments and up to ten features are tested in the gait recognition experiments.

3) Face Recognition Results: Gray-level face images are naturally second-order tensors (matrices), i.e., $N = 2$. Therefore, they are input directly as 80×80 tensors to the multilinear algorithms (2DPCA, CSA, TROD, GPCA, MPCA, and UMPCA), while for PCA, they are vectorized to 6400×1 vectors as input. For each subject in a face recognition experiment, $L\ (= 1, 2, 3, 4, 5, 6, 7)$ samples are randomly selected for unsupervised training and the rest are used for testing. The results averaged over 20 such random splits (repetitions) are reported in terms of the correct recognition rate (CRR), i.e., the rank 1 identification rate. Figs. 6 and 7 show the detailed results for $L = 1$ and $L = 7$, respectively. $L = 1$ is the one training sample (per class) scenario [30], and $L = 7$ is the maximum number of training samples that can be used in this set of experiments. It should be noted that for PCA and UMPCA, there are at most 69 (uncorrelated) features when $L = 1$ since there are only 70 faces for training and the mean is zero. Figs. 6(a) and 7(a) plot the CRRs against P, the dimensionality of the subspace, for $P = 1, \ldots, 10$, and Figs. 6(b) and 7(b) plot those for P ranging from 15 to 80. From the figures, UMPCA outperforms the other five methods in both cases and across all Ps, indicating that the uncorrelated features extracted directly from the tensorial face data are effective in classification. The figures also show that for

TABLE III
FACE RECOGNITION RESULTS BY PCA ALGORITHMS ON THE FERET DATABASE: THE CRRS ($\mathrm{Mean} \pm \mathrm{Std}\%$) FOR VARIOUS Ls AND Ps

L	P	1	5	10	20	50	80
2	PCA	2.8±0.7	20.3±1.5	**31.5**±2.3	**38.4**±2.4	43.1±2.8	44.4±2.6
	2DPCA	**5.3**±0.9	15.9±1.9	19.5±2.1	26.0±3.2	40.4±2.8	44.2±2.4
	CSA	3.5±0.7	15.5±1.0	29.2±1.9	36.3±2.2	43.8±2.6	45.0±2.8
	TROD	3.5±0.8	19.9±3.4	30.1±2.1	37.7±2.4	42.3±2.1	43.8±2.5
	GPCA	2.6±0.6	**21.4**±1.5	28.4±1.8	38.2±2.2	**43.9**±2.7	**45.2**±2.7
	UMPCA	**7.9**±1.5	**30.0**±5.2	**41.7**±5.6	**46.1**±6.0	**46.7**±6.3	**46.7**±6.3
3	PCA	2.7±0.6	24.5±1.9	**38.0**±2.2	**46.3**±2.2	51.8±2.7	53.0±2.5
	2DPCA	**5.1**±0.9	17.3±1.5	22.3±1.8	30.5±2.7	47.4±2.6	51.8±2.2
	CSA	3.9±0.8	17.3±1.6	36.4±1.5	44.4±1.9	52.1±2.6	53.5±2.8
	TROD	3.9±0.7	23.2±3.3	36.4±2.3	45.1±2.4	50.5±2.7	52.2±2.6
	GPCA	2.4±0.6	**25.8**±1.6	34.9±2.3	45.8±2.2	**52.3**±2.7	**53.9**±2.8
	UMPCA	**7.5**±1.0	**35.3**±3.8	**49.7**±3.6	**56.0**±4.0	**56.7**±4.3	**56.6**±4.3
4	PCA	2.8±0.7	26.7±2.4	**42.5**±2.3	50.2±1.8	57.8±2.2	58.8±2.4
	2DPCA	**5.4**±0.6	18.3±1.1	24.3±1.7	34.1±4.3	51.7±2.5	56.4±2.5
	CSA	3.7±1.0	19.0±1.4	41.2±2.4	50.2±2.1	**58.4**±2.8	**59.6**±2.5
	TROD	3.8±0.9	25.3±2.6	42.2±3.1	50.0±2.6	55.6±2.1	57.6±2.4
	GPCA	2.3±0.6	**29.5**±2.3	40.4±2.4	**51.2**±2.5	58.3±2.5	**59.6**±2.3
	UMPCA	**8.1**±1.3	**40.1**±3.8	**56.9**±3.0	**63.3**±3.3	**64.0**±3.6	**64.0**±3.6
5	PCA	2.8±0.8	29.2±1.9	**47.0**±1.7	55.5±2.0	63.6±1.5	64.8±1.5
	2DPCA	**5.6**±1.2	19.9±1.6	26.4±2.4	36.4±3.5	57.0±2.5	61.6±2.3
	CSA	4.2±1.1	20.7±1.9	46.0±2.2	56.1±2.5	**64.8**±2.1	65.6±1.7
	TROD	4.2±1.1	28.9±3.0	46.7±2.9	55.6±2.1	61.6±1.9	63.7±1.8
	GPCA	2.6±0.7	**32.6**±2.1	43.0±2.6	**57.0**±2.2	64.4±2.1	**65.9**±1.7
	UMPCA	**8.5**±1.6	**42.5**±4.5	**61.0**±5.2	**67.7**±5.0	**68.7**±5.1	**68.7**±5.1
6	PCA	2.6±0.8	30.0±2.0	49.6±2.9	58.3±2.5	66.6±2.2	67.9±2.3
	2DPCA	**5.4**±1.4	20.9±1.9	27.9±2.7	38.3±2.9	58.1±1.9	63.2±2.4
	CSA	4.0±0.8	22.4±1.9	49.1±2.4	59.5±2.7	**68.0**±2.6	69.0±2.4
	TROD	4.3±0.7	28.5±2.8	**50.3**±2.3	58.7±2.7	64.8±2.3	66.7±2.0
	GPCA	2.2±1.0	**34.2**±2.4	46.4±2.6	**60.5**±2.6	67.5±2.5	**69.4**±2.3
	UMPCA	**9.0**±1.2	**44.5**±4.2	**63.1**±4.5	**70.4**±4.8	**71.4**±4.9	**71.3**±4.9

UMPCA, the recognition rate saturates around $P = 30$, which can be explained by observing the variation captured by individual features as shown in Figs. 6(c) and 7(c) (in log scale). These figures show that the variation captured by UMPCA is considerably lower than those captured by the other methods, due to its constraints of zero-correlation and the TVP. Despite capturing lower variation, UMPCA has superior performance in the recognition task performed. Nonetheless, when the variation captured is too low, those corresponding features are no longer descriptive enough to contribute in classification, leading to the saturation.

In addition, the average correlations of individual features with all the other features are plotted in Figs. 6(d) and 7(d). As supported by theoretical derivation, features extracted by PCA and UMPCA are uncorrelated. In contrast, the features extracted by all the other methods are correlated, with those extracted by 2DPCA and TROD having much higher correlation on average, which could be partly the reason of their poorer performance.

The recognition results for $P = 1, 5, 10, 20, 50, 80$ are listed in Table III for $L = 2, 3, 4, 5, 6$, with both the mean and the standard deviation (STD) over 20 repetitions indicated. The top two results for each L and P combination are highlighted in bold face for easy reading. From the table, UMPCA achieves the best recognition results in all cases reported. In particular, for smaller $P\,(1, 5, 10, 20)$, UMPCA outperforms the other methods significantly, demonstrating its superior capability in classifying faces in low-dimensional projection spaces.

4) Gait Recognition Results: In order to evaluate the recognition performance of UMPCA on third-order tensors, recognition experiments are also carried out on the 3-D gait data described in Section IV-B1. In these experiments, gait samples are input directly as third-order tensors to the multilinear algorithms, while for PCA, they are vectorized to 7040×1 vectors as input. The gallery set is used for training and the probe set A is used for testing. For the classification of individual gait samples, the CRRs are reported in Fig. 8(a). For the classification of gait sequences, both rank 1 and rank 5 recognition rates are reported in Fig. 8(b) and (c). The calculation of matching scores between two gait sequences follows that in [9]. From Fig. 8(a), starting from four features, UMPCA outperforms the other four methods consistently in terms of CRRs for individual gait samples. Fig. 8(b) and (c) demonstrates that starting from three features, UMPCA gives better recognition performance than all the other four algorithms. These results again show the superiority of UMPCA in low-dimensional projection space in unsupervised classification of tensorial samples.

On the other hand, in this experiment, the number of features that can be extracted by the UMPCA algorithm is limited to ten, the lowest mode dimension of the gait sample (from Corollary 1). Such a small number prevents higher recognition rate to be achieved. This limitation of UMPCA may be partly overcome through the combination with other features to enhance the recognition results. For example, MPCA may be combined with UMPCA through the relaxation of the zero-correlation constraint, or PCA may be combined with UMPCA through the relaxation of the projection. Beyond the maximum number features from UMPCA, the heuristic approach in TROD may also be adopted to generate more features. Furthermore, the aggregation scheme introduced in [6] could also be a possible future working direction.

C. Illustration of the UMPCA Projections

Finally, in order to provide some insights into the UMPCA algorithm, Fig. 9(a) and (b) depicts, as gray-level images, the first three EMPs obtained by UMPCA using the FERET database with $L = 4$, and the USF gait gallery set, respectively. From the EMPs for the face data, it can be seen that there is strong presence of structured information due to the multilinear nature of UMPCA, which is different from the information conveyed by the ghost-face-like bases produced by linear algorithms such as eigenface [21] or fisherface [41]. The maximum variation (the first EMP) mainly captures the difference between forehead area and the rest of the facial image [with reference to Fig. 5(a)]. The second and third EMPs indicate that there are also significant variations around the other facial feature areas, such as the eyes, nose, and mouth. In the EMPs for the gait data, which are third-order tensors displayed as their one-mode unfolded matrices in Fig. 9(b), structure is again observed across the three modes (column, row, and time). The first gait EMP indicates that

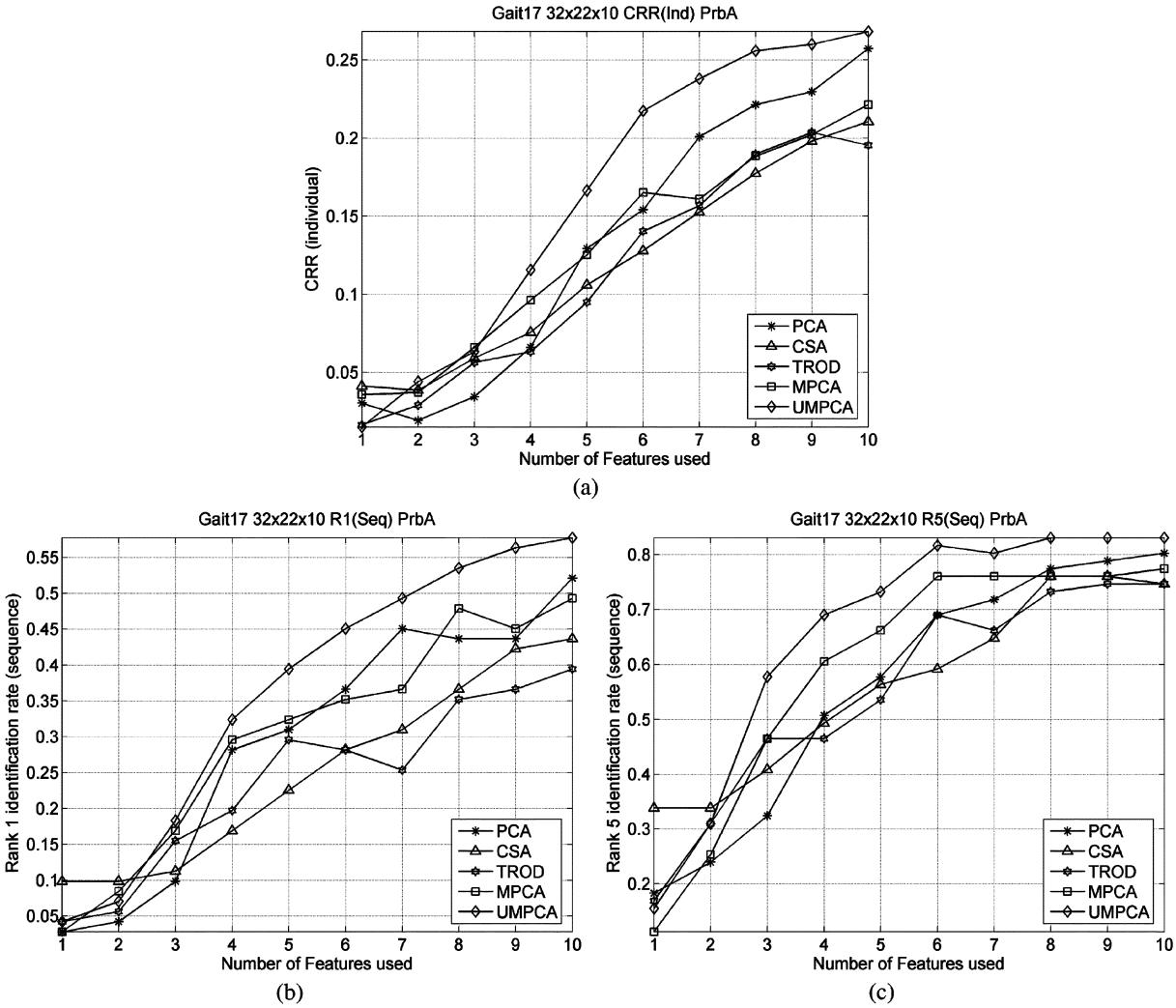
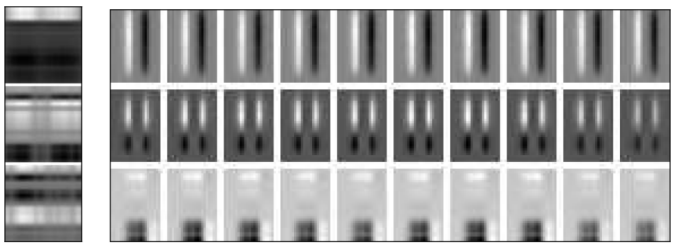


Fig. 8. Detailed gait recognition results by PCA algorithms on the USF gait database (probe A): (a) CRR for individual gait samples, (b) rank 1 recognition rate for gait sequences, and (c) rank 5 recognition rate for gait sequences.



(a) (b)

Fig. 9. Illustration of the first three EMPs (in top-to-down order) obtained by UMPCA from (a) the FERET database with $L = 4$, and (b) the USF gait gallery sequences (one-mode unfolding is shown).

the highlighted symmetric areas in frames encode the most variations, which roughly correspond to the boundary of the human silhouette [with reference to Fig. 5(b)]. The second gait EMP demonstrates that significant variation is encoded in the difference between the upper body and lower body movement. The third gait EMP shows that there is also considerable variation in the foot area, as expected. These observations provide insights into the nature of the features encoded by UMPCA and offer a better understanding of this algorithm's performance when applied to certain data sets.

V. Conclusion

This paper has introduced a novel unsupervised learning solution called UMPCA, which is capable of extracting uncorrelated features directly from tensorial representation through a TVP. This feature extraction problem is solved by successive maximization of variance captured by each elementary projection while enforcing the zero-correlation constraint. The solution is an iterative method utilizing an alternating projection method. A proof is provided regarding possible restrictions on the maximum number of uncorrelated features that can be extracted by the UMPCA procedure. Experimentation on face and gait recognition data sets demonstrates that compared with other unsupervised subspace learning algorithms including PCA, 2DPCA, CSA, TROD, GPCA, and MPCA, the proposed UMPCA solution has achieved the best overall results with the same number of features. Moreover, the low-dimensional projection space produced by UMPCA is particularly effective in these recognition tasks.

APPENDIX I
PROOF OF THEOREM 1

Proof: First, Lagrange multipliers can be used to transform the problem in (13) to the following to include all the constraints:

$$F\left(\mathbf{u}_p^{(n^*)}\right) = \mathbf{u}_p^{(n^*)^T}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - \nu\left(\mathbf{u}_p^{(n^*)^T}\mathbf{u}_p^{(n^*)} - 1\right) - \sum_{q=1}^{p-1}\mu_q\mathbf{u}_p^{(n^*)^T}\tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{g}_q \quad (21)$$

where ν and $\{\mu_q, q = 1, \ldots, p-1\}$ are Lagrange multipliers.

The optimization is performed by setting the partial derivative of $F(\mathbf{u}_p^{(n^*)})$ with respect to $\mathbf{u}_p^{(n^*)}$ to zero

$$\frac{\partial F\left(\mathbf{u}_p^{(n^*)}\right)}{\partial \mathbf{u}_p^{(n^*)}} = 2\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - 2\nu\mathbf{u}_p^{(n^*)} - \sum_{q=1}^{p-1}\mu_q\tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{g}_q = 0. \quad (22)$$

Multiplying (22) by $\mathbf{u}_p^{(n^*)^T}$ results in

$$2\mathbf{u}_p^{(n^*)^T}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - 2\nu\mathbf{u}_p^{(n^*)^T}\mathbf{u}_p^{(n^*)} = 0 \Rightarrow \nu = \frac{\mathbf{u}_p^{(n^*)^T}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)}}{\mathbf{u}_p^{(n^*)^T}\mathbf{u}_p^{(n^*)}} \quad (23)$$

which indicates that ν is exactly the criterion to be maximized, with the constraint on the norm of the projection vector incorporated.

Next, a set of $(p-1)$ equations are obtained by multiplying (22) by $\mathbf{g}_q^T\tilde{\mathbf{Y}}_p^{(n^*)^T}$, $q = 1, \ldots, p-1$, respectively

$$2\mathbf{g}_q^T\tilde{\mathbf{Y}}_p^{(n^*)^T}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - \sum_{q=1}^{p-1}\mu_q\mathbf{g}_q^T\tilde{\mathbf{Y}}_p^{(n^*)^T}\cdot\tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{g}_q = 0. \quad (24)$$

Let

$$\boldsymbol{\mu}_{p-1} = [\,\mu_1 \quad \mu_2 \quad \ldots \quad \mu_{p-1}\,]^T \quad (25)$$

and use (16) and (17), then the $(p-1)$ equations of (24) can be represented in a single matrix equation as the following:

$$2\mathbf{G}_{p-1}^T\tilde{\mathbf{Y}}_p^{(n^*)^T}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - \boldsymbol{\Phi}_p\boldsymbol{\mu}_{p-1} = 0. \quad (26)$$

Thus

$$\boldsymbol{\mu}_{p-1} = 2\boldsymbol{\Phi}_p^{-1}\cdot\mathbf{G}_{p-1}^T\tilde{\mathbf{Y}}_p^{(n^*)^T}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)}. \quad (27)$$

Since from (17) and (25)

$$\sum_{q=1}^{p-1}\mu_q\tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{g}_q = \tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{G}_{p-1}\boldsymbol{\mu}_{p-1} \quad (28)$$

equation (22) can be written as

$$\begin{aligned}
&2\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - 2\nu\mathbf{u}_p^{(n^*)} - \tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{G}_{p-1}\boldsymbol{\mu}_{p-1} = 0\\
&\Rightarrow \nu\mathbf{u}_p^{(n^*)} = \tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - \tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{G}_{p-1}\frac{\boldsymbol{\mu}_{p-1}}{2}\\
&= \tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)} - \tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{G}_{p-1}\boldsymbol{\Phi}_p^{-1}\cdot\mathbf{G}_{p-1}^T\tilde{\mathbf{Y}}_p^{(n^*)^T}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)}\\
&= \left[\mathbf{I}_{I_{n^*}} - \tilde{\mathbf{Y}}_p^{(n^*)}\mathbf{G}_{p-1}\boldsymbol{\Phi}_p^{-1}\mathbf{G}_{p-1}^T\tilde{\mathbf{Y}}_p^{(n^*)^T}\right]\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u}_p^{(n^*)}.
\end{aligned}$$

Using the definition in (15), an eigenvalue problem is obtained as $\boldsymbol{\Psi}_p^{(n^*)}\tilde{\mathbf{S}}_{T_p}^{(n^*)}\mathbf{u} = \nu\mathbf{u}$. Since ν is the criterion to be maximized, the maximization is achieved by setting $\mathbf{u}_p^{(n)^*}$ to be the (unit) eigenvector corresponding to the largest eigenvalue of (14). ■

APPENDIX II
PROOF OF COROLLARY 1

Proof: To prove the corollary, it is only needed to show that for any mode n, the number of bases that can satisfy the zero-correlation constraint is upper-bounded by $\min\{I_n, M\}$.

Considering only one mode n, the zero-correlation constraint for mode $n^* = n$ in (13) becomes

$$\mathbf{u}_p^{(n)^T}\tilde{\mathbf{Y}}_p^{(n)}\mathbf{g}_q = 0, \qquad q = 1, \ldots, p-1. \quad (29)$$

First, let $\hat{\mathbf{g}}_p^{(n)^T} = \mathbf{u}_p^{(n)^T}\tilde{\mathbf{Y}}_p^{(n)} \in \mathbb{R}^{1\times M}$ and the constraint becomes

$$\hat{\mathbf{g}}_p^{(n)^T}\mathbf{g}_q = 0, \qquad q = 1, \ldots, p-1. \quad (30)$$

Since $\mathbf{g}_q \in \mathbb{R}^{M\times 1}$, when $p = M+1$, the set $\mathbf{g}_q$, $q = 1, \ldots, M$, forms a basis for the M-dimensional space and there is no solution for (30). Thus, $P \le M$.

Second, let $\hat{\mathbf{u}}_q^{(n)} = \tilde{\mathbf{Y}}_p^{(n)}\mathbf{g}_q \in \mathbb{R}^{I_n\times 1}$ and the constraint becomes

$$\mathbf{u}_p^{(n)^T}\hat{\mathbf{u}}_q^{(n)} = 0, \qquad q = 1, \ldots, p-1. \quad (31)$$

Since $\mathbf{g}_q$, $q = 1, \ldots, p-1$, are orthogonal, $\hat{\mathbf{u}}_q^{(n)}$, $q = 1, \ldots, p-1$, are linearly independent if the elements of $\tilde{\mathbf{Y}}_p^{(n)}$ are not all zero. Since $\hat{\mathbf{u}}_q^{(n)} \in \mathbb{R}^{I_n\times 1}$, when $p = I_n + 1$, the set $\hat{\mathbf{u}}_q^{(n)}$, $q = 1, \ldots, p-1$, forms a basis for the I_n-dimensional space and there is no solution for (31). Thus, $P \le I_n$.

From the above, $P \le \min\{\min_n I_n, M\}$ if the elements of $\tilde{\mathbf{Y}}_p^{(n)}$ are not all zero, which is often the case as long as the projection basis is not initialized to zero and the elements of the training tensors are not all zero. ■

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful comments.

REFERENCES

[1] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[2] C. M. Cyr and B. B. Kimia, "3D object recognition using shape similiarity-based aspect graph," in *Proc. IEEE Conf. Comput. Vis.*, Jul. 2001, vol. 1, pp. 254–261.

[3] H. S. Sahambi and K. Khorasani, "A neural-network appearance-based 3-D object recognition using independent component analysis," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 138–149, Jan. 2003.

[4] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Comput. Vis. Image Understand.*, vol. 101, no. 1, pp. 1–15, Jan. 2006.

[5] S. Z. Li, C. Zhao, M. Ao, and Z. Lei, "Learning to fuse 3D + 2D based face recognition at both feature and decision levels," in *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures*, Oct. 2005, pp. 43–53.

[6] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 103–123, Jan. 2009.

[7] J. Ye, "Generalized low rank approximations of matrices," *Mach. Learn.*, vol. 61, no. 1–3, pp. 167–191, 2005.
[8] J. Ye, R. Janardan, and Q. Li, "GPCA: An efficient dimension reduction scheme for image compression and retrieval," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining*, 2004, pp. 354–363.
[9] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.
[10] C. Nolker and H. Ritter, "Visual recognition of continuous hand postures," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 983–994, Jul. 2002.
[11] C. Faloutsos, T. G. Kolda, and J. Sun, "Mining large time-evolving data using matrix and tensor tools," in *Int. Conf. Mach. Learn. Tutorial*, 2007 [Online]. Available: http://www.cs.cmu.edu/~christos/TALKS/ICML-07-tutorial/ICMLtutorial.pdf
[12] J. Sun, D. Tao, and C. Faloutsos, "Beyond streams and graphs: Dynamic tensor analysis," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining*, Aug. 2006, pp. 374–383.
[13] J. Sun, Y. Xie, H. Zhang, and C. Faloutsos, "Less is more: Sparse graph mining with compact matrix decomposition," *Statist. Anal. Data Mining*, vol. 1, no. 1, pp. 6–22, Feb. 2008.
[14] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos, "Incremental tensor analysis: Theory and applications," *ACM Trans. Knowl. Disc. Data*, vol. 2, no. 3, pp. 11:1–11:37, Oct. 2008.
[15] G. Shakhnarovich and B. Moghaddam, "Face recognition in subspaces," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds. New York: Springer-Verlag, 2004, pp. 141–168.
[16] M. H. C. Law and A. K. Jain, "Incremental nonlinear dimensionality reduction by manifold learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 377–391, Mar. 2006.
[17] J. Yang, D. Zhang, A. F. 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.
[18] D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, and T. S. Huang, "Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 36–47, Jan. 2008.
[19] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A taxonomy of emerging multilinear discriminant analysis solutions for biometric signal recognition," in *Biometrics: Theory, Methods, and Applications*, N. Boulgouris, K. N. Plataniotis, and E. Micheli-Tzanakou, Eds. New York: Wiley, 2009, ISBN: 978-0-470-24782-2.
[20] I. T. Jolliffe, *Principal Component Analysis*, ser. Springer Serires in Statistics, 2nd ed. : , 2002.
[21] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
[22] A. Shashua and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2001, vol. I, pp. 42–49.
[23] D. Tao, X. Li, X. Wu, W. Hu, and S. J. Maybank, "Supervised tensor learning," *Knowl. Inf. Syst.*, vol. 13, no. 1, pp. 1–42, Jan. 2007.
[24] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[25] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Tensor rank one discriminant analysis—a convergent method for discriminative multilinear subspace selection," *Neurocomputing*, vol. 71, no. 10–12, pp. 1866–1882, Jun. 2008.
[26] D. Xu, S. Yan, L. Zhang, H.-J. Zhang, Z. Liu, and H.-Y. Shum, "Concurrent subspaces analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. II, pp. 203–208.
[27] C. Ding and J. Ye, "2-dimensional singular value decomposition for 2D maps and images," in *Proc. SIAM Int. Conf. Data Mining*, 2005, pp. 24–34.
[28] J. Ye, R. Janardan, Q. Li, and H. Park, "Feature reduction via generalized uncorrelated linear discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1312–1322, Oct. 2006.
[29] J. Wang, K. N. Plataniotis, and A. N. Venetsanopoulos, "Selecting discriminant eigenfaces for face recognition," *Pattern Recognit. Lett.*, vol. 26, no. 10, pp. 1470–1482, 2005.
[30] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognit.*, vol. 39, no. 9, pp. 1746–1762, 2006.
[31] Y. Koren and L. Carmel, "Robust linear dimensionality reduction," *IEEE Trans. Vis. Comput. Graphics*, vol. 10, no. 4, pp. 459–470, Jul. –Aug. 2004.
[32] J. L. Rodgers, W. A. Nicewander, and L. Toothaker, "Linearly independent, orthogonal and uncorrelated variables," *Amer. Statistician*, vol. 38, no. 2, pp. 133–134, May 1984.
[33] T. K. Moon and W. C. Stirling, *Mathematical methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[34] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.
[35] Y. Wang and S. Gong, "Tensor discriminant analysis for view-based object recognition," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2006, vol. 3, pp. 33–36.
[36] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Elapsed time in human gait recognition: A new approach," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2006, vol. 2, pp. 177–180.
[37] S. Z. Li and A. K. Jain, "Introduction," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds. New York: Springer-Verlag, 2004, pp. 1–11.
[38] M. S. Nixon and J. N. Carter, "Automatic recognition by gait," *Proc. IEEE*, vol. 94, no. 11, pp. 2013–2024, Nov. 2006.
[39] R. Chellappa, A. Roy-Chowdhury, and S. Zhou, *Recognition of Humans and Their Activities Using Video*. San Rafael, CA: Morgan & Claypool, 2005.
[40] P. J. Phillips, H. Moon, S. A. Rizvi, and P. Rauss, "The FERET evaluation method for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[41] P. N. Belhumeur, J. P. Hespanha, 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.

Haiping Lu (S'02–M'09) received the B.Eng. and M.Eng. degrees in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2001 and 2004, respectively, and the Ph.D. degree in electrical and computer engineering from University of Toronto, Toronto, ON, Canada, in 2008.

Currently, he is a Research Fellow at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. Before joining A*STAR, he was a Postdoctoral Fellow at the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto. His current research interests include statistical pattern recognition, machine learning, multilinear algebra, tensor object processing, biometric encryption, and data mining.

Konstantinos N. (Kostas) Plataniotis (S'90–M'92–SM'03) received the B.Eng. degree in computer engineering from University of Patras, Patras, Greece, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from Florida Institute of Technology (Florida Tech), Melbourne, in 1992 and 1994, respectively.

He is a Professor with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, an Adjunct Professor with the School of Computer Science at Ryerson University, Toronto, ON, Canada, a member of The University of Toronto's Knowledge Media Design Institute, and the Director of Research for the Identity, Privacy and Security Initiative at the University of Toronto. His research interests include biometrics, communications systems, multimedia systems, and signal and image processing.

Dr. Plataniotis is the Editor in Chief (2009–2011) for the IEEE SIGNAL PROCESSING LETTERS, a registered professional engineer in the province of Ontario, and a member of the Technical Chamber of Greece. He is the 2005 recipient of IEEE Canada's Outstanding Engineering Educator Award "for contributions to engineering education and inspirational guidance of graduate students" and the corecipient of the 2006 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award for the published in 2003 paper entitled "Face Recognition Using Kernel Direct Discriminant Analysis Algorithms."

Anastasios N. Venetsanopoulos (S'66–M'69–SM'79–F'88) received the B.S. degree in electrical and mechanical engineering degree from the National Technical University of Athens (NTU), Athens, Greece, in 1965 and the M.S., M.Phil., and Ph.D. degrees in electrical engineering from Yale University, New Haven, CT, in 1966, 1968, and 1969, respectively.

He joined the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, in September 1968 as a Lecturer and he was promoted to Assistant Professor in 1970, Associate Professor in 1973, and Professor in 1981. He has served as Chair of the Communications Group and Associate Chair of the Department of Electrical Engineering. Between July 1997 and June 2001, he was Associate Chair (graduate studies) in the Department of Electrical and Computer Engineering and was Acting Chair during the spring term of 1998–1999. He served as the Inaugural Bell Canada Chair in Multimedia between 1999 and 2005. Between 2001 and 2006 he served as the 12th Dean of the Faculty of Applied Science and Engineering of the University of Toronto and on October 1, 2006, he accepted the position of founding vice-president research and innovation at Ryerson University, Toronto, ON, Canada. He was on research leave at the Imperial College of Science and Technology, the National Technical University of Athens, the Swiss Federal Institute of Technology, the University of Florence, and the Federal University of Rio de Janeiro, and has also served as Adjunct Professor at Concordia University. He has served as lecturer in 138 short courses to industry and continuing education programs and as consultant to numerous organizations; he is a contributor to 35 books and has published over 800 papers in refereed journals and conference proceedings on digital signal and image processing and digital communications.

Prof. Venetsanopoulos was elected as a Fellow of the IEEE "for contributions to digital signal and image processing"; he is also a Fellow of the Engineering Institute of Canada (EIC) and was awarded an honorary doctorate from the National Technical University of Athens in October 1994. In October 1996, he was awarded the "Excellence in Innovation Award" of the Information Technology Research Centre of Ontario and Royal Bank of Canada, "for innovative work in color image processing and its industrial applications." In November 2003, he was the recipient of the "Millennium Medal of IEEE." In April 2001 he became a Fellow of the Canadian Academy of Engineering. In 2003, he was the recipient of the highest award of the Canadian IEEE, the MacNaughton Award, and in 2003, he served as the Chair of the Council of Deans of Engineering of the Province of Ontario (CODE). In 2006, he was selected as the joint recipient of the 2003 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award.