Published in the Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Madison, WI, June, 2003, Vol 2: 93–99.

Multilinear Subspace Analysis of Image Ensembles

M. Alex O. Vasilescu[1,2] and **Demetri Terzopoulos**[2,1]

[1]Department of Computer Science, University of Toronto, Toronto ON M5S 3G4, Canada
[2]Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA

Abstract

Multilinear algebra, the algebra of higher-order tensors, offers a potent mathematical framework for analyzing ensembles of images resulting from the interaction of any number of underlying factors. We present a dimensionality reduction algorithm that enables subspace analysis within the multilinear framework. This N-mode orthogonal iteration algorithm is based on a tensor decomposition known as the N-mode SVD, the natural extension to tensors of the conventional matrix singular value decomposition (SVD). We demonstrate the power of multilinear subspace analysis in the context of facial image ensembles, where the relevant factors include different faces, expressions, viewpoints, and illuminations. In prior work we showed that our multilinear representation, called TensorFaces, yields superior facial recognition rates relative to standard, linear (PCA/eigenfaces) approaches. Here, we demonstrate factor-specific dimensionality reduction of facial image ensembles. For example, we can suppress illumination effects (shadows, highlights) while preserving detailed facial features, yielding a low perceptual error.

Keywords: nonlinear subspace analysis, N-mode component analysis, multilinear models, tensor decomposition, N-mode SVD, dimensionality reduction.

1 Introduction

Natural images are generated by the interaction of multiple factors related to scene structure, illumination, and imaging. Human perception remains robust despite significant variation of these factors. For example, people possess a remarkable ability to recognize faces despite a broad variety of expressive facial geometries, viewpoints, and lighting conditions. Our work confronts the challenge of learning tractable, nonlinear models of image ensembles useful in image compression and in difficult appearance-based recognition problems [5], such as facial recognition under varying conditions [1].

We have recently introduced a multilinear approach to the analysis of image ensembles that explicitly accounts for each of the multiple factors implicit in image formation [13, 12]. Our approach is motivated by the observation that multilinear algebra, the algebra of higher-order tensors, offers a potent mathematical framework for analyzing the multifactor structure of the image ensemble. It provides techniques for decomposing the ensemble in order to disentangle the constituent factors or modes.

The natural generalization of matrices (i.e., linear operators defined over a vector space), tensors define multilinear operators over a *set* of vector spaces. Hence, tensor analysis, which subsumes linear analysis as a special case, is a unifying mathematical framework suitable for addressing a variety of visual problems. In particular, we have introduced algorithms for learning multilinear models of facial image ensembles, called *TensorFaces* [13]. In facial recognition scenarios that involve varying viewpoint and illumination, TensorFaces yield dramatically improved recognition rates [12] over the linear facial recognition method known as eigenfaces [10].

This paper addresses subspace analysis within our multilinear framework, via dimensionality reduction over the multiple affiliated vector spaces. Multilinear dimensionality reduction generalizes the conventional version associated with linear principal components analysis (PCA), truncation of the singular value decomposition (SVD), whose optimality properties are well-known. Unfortunately, *optimal* dimensionality reduction is not straightforward in multilinear analysis. For multilinear dimensionality reduction, we present an N-mode orthogonal iteration algorithm based on a tensor decomposition known as the N-mode SVD. The latter is a natural extension to higher-order tensors of the conventional matrix SVD.

Following a review in Section 2 of the details of our multilinear approach, Section 3 presents the multilinear dimensionality reduction algorithm. In Section 4, we demonstrate factor-specific dimensionality reduction of facial image ensembles. In particular, we show that we can suppress illumination effects such as shadows and highlights, yet preserve detailed facial features, yielding a low perceptual error. Section 5 concludes the paper.

2 Synopsis of the Multilinear Approach

A *tensor* is a higher order generalization of a vector (first order tensor) and a matrix (second order tensor).

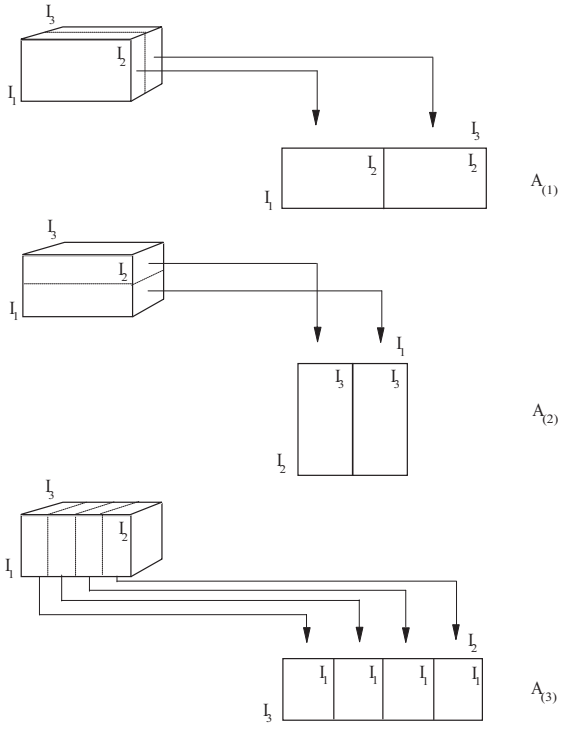


Figure 1: Flattening a (3rd-order) tensor. The tensor can be flattened in 3 ways to obtain matrices comprising its mode-1, mode-2, and mode-3 vectors.

2.1 Tensor Fundamentals

Tensors are multilinear mappings over a set of vector spaces. The *order* of tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$ is N.[1] Elements of $\mathcal{A}$ are denoted as $\mathcal{A}_{i_1 \ldots i_n \ldots i_N}$ or $a_{i_1 \ldots i_n \ldots i_N}$, where $1 \leq i_n \leq I_n$. In tensor terminology, matrix column vectors are referred to as mode-1 vectors and row vectors as mode-2 vectors. The mode-n vectors of an N$^{\text{th}}$ order tensor $\mathcal{A}$ are the I_n-dimensional vectors obtained from $\mathcal{A}$ by varying index i_n while keeping the other indices fixed. The mode-n vectors are the column vectors of matrix $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \ldots I_{n-1} I_{n+1} \ldots I_N)}$ that results by *mode-n flattening* the tensor $\mathcal{A}$ (Fig. 1). The *n-rank* of $\mathcal{A}$, denoted R_n, is defined as the dimension of the vector space generated by the mode-n vectors: $R_n = \text{rank}_n(\mathcal{A}) = \text{rank}(\mathbf{A}_{(n)})$.

A generalization of the product of two matrices is the product of a tensor and a matrix. The *mode-n product* of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N}$ whose entries are

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

[1] We denote scalars by lower case letters ($a, b, \ldots$), vectors by bold lower case letters ($\mathbf{a}, \mathbf{b}, \ldots$), matrices by bold upper-case letters ($\mathbf{A}, \mathbf{B}, \ldots$), and higher-order tensors by calligraphic upper-case letters ($\mathcal{A}, \mathcal{B}, \ldots$).

The mode-n product $\mathcal{B} = \mathcal{A} \times_n \mathbf{U}$ can be computed via the matrix multiplication $\mathbf{B}_{(n)} = \mathbf{U}\mathbf{A}_{(n)}$, followed by a re-tensorization to undo the mode-n flattening.[2]

The scalar product of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$, is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \ldots \sum_{i_N} a_{i_1 \ldots i_N} b_{i_1 \ldots i_N}$. The Frobenius norm of a tensor $\mathcal{A}$ is $\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$.

2.2 Tensor Decomposition of Image Ensembles

Image formation is the consequence of interactions among multiple factors—scene geometry, camera viewpoint, illumination conditions, etc. We formulate the analysis of an ensemble of images as a problem in multilinear algebra. Within this mathematical framework, the image ensemble is represented as a higher-order tensor. This image data tensor $\mathcal{D}$ must be decomposed in order to separate and parsimoniously represent the constituent factors.

To this end, we subject $\mathcal{D}$ to a generalization of matrix SVD. Matrix SVD orthogonalizes the column and row space, the two associated spaces of a matrix. An order $N > 2$ tensor $\mathcal{D}$ is an N-dimensional matrix comprising N spaces. N-mode SVD is a "generalization" of conventional matrix (i.e., 2-mode) SVD. It orthogonalizes these N spaces and decomposes the tensor as the mode-n product (1) of N-orthogonal spaces. Thus, a tensor can be expressed as a multilinear model of factors as follows:

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_n \mathbf{U}_n \ldots \times_N \mathbf{U}_N. \quad (2)$$

Tensor $\mathcal{Z}$, known as the *core tensor*, is analogous to the diagonal singular value matrix in conventional matrix SVD, but it does not have a simple, diagonal structure. The core tensor governs the interaction between the *mode matrices* $\mathbf{U}_1, \ldots, \mathbf{U}_N$. Mode matrix $\mathbf{U}_n$ contains the orthonormal vectors spanning the column space of matrix $\mathbf{D}_{(n)}$ resulting from the *mode-n flattening* of $\mathcal{D}$.

The **N-mode SVD algorithm** for decomposing $\mathcal{D}$ according to equation (2) is as follows:

1. For $n = 1, \ldots, N$, compute matrix $\mathbf{U}_n$ in (2) by computing the SVD of the flattened matrix $\mathbf{D}_{(n)}$ and setting $\mathbf{U}_n$ to be the left matrix of the SVD.[3]

2. Solve for the core tensor as follows:

$$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \ldots \times_n \mathbf{U}_n^T \ldots \times_N \mathbf{U}_N^T. \quad (3)$$

[2] The mode-n product of a tensor and a matrix is a special case of the inner product in multilinear algebra and tensor analysis. Note that for tensors and matrices of the appropriate sizes, $\mathcal{A} \times_m \mathbf{U} \times_n \mathbf{V} = \mathcal{A} \times_n \mathbf{V} \times_m \mathbf{U}$ and $(\mathcal{A} \times_n \mathbf{U}) \times_n \mathbf{V} = \mathcal{A} \times_n (\mathbf{V}\mathbf{U})$.

[3] For a non-square, $m \times n$ matrix $\mathbf{A}$, the matrix $\mathbf{U}$ in the SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ can be computed more efficiently, depending on which dimension of $\mathbf{A}$ is smaller, by decomposing either the $m \times m$ matrix $\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T$ and then computing $\mathbf{V}^T = \mathbf{\Sigma}^+\mathbf{U}^T\mathbf{A}$ or by decomposing the $n \times n$ matrix $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T$ and then computing $\mathbf{U} = \mathbf{A}\mathbf{V}\mathbf{\Sigma}^+$.

2.3 TensorFaces

The multilinear analysis of facial image ensembles leads to the TensorFaces representation. We illustrate the technique using a portion of the Weizmann face image database: 28 male subjects photographed in 5 viewpoints, 4 illuminations, and 3 expressions. Using a global rigid optical flow algorithm, we aligned the original 512×352 pixel images relative to one reference image. The images were then decimated by a factor of 3 and cropped as shown in Fig. 2, yielding a total of 7943 pixels per image within the elliptical cropping window.

Our facial image data tensor $\mathcal{D}$ is a $28 \times 5 \times 4 \times 3 \times 7943$ tensor (Fig. 2(c)). Applying multilinear analysis to $\mathcal{D}$, using our N-mode SVD algorithm with $N = 5$, we obtain

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{\text{people}} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{illums}} \times_4 \mathbf{U}_{\text{expres}} \times_5 \mathbf{U}_{\text{pixels}}, \quad (4)$$

where the $28 \times 5 \times 3 \times 3 \times 7943$ core tensor $\mathcal{Z}$ governs the interaction between the factors represented in the 5 mode matrices: The 28×28 mode matrix $\mathbf{U}_{\text{people}}$ spans the space of people parameters, the 5×5 mode matrix $\mathbf{U}_{\text{views}}$ spans the space of viewpoint parameters, the 4×4 mode matrix $\mathbf{U}_{\text{illums}}$ spans the space of illumination parameters and the 3×3 mode matrix $\mathbf{U}_{\text{expres}}$ spans the space of expression parameters. The 7943×1680 mode matrix $\mathbf{U}_{\text{pixels}}$ orthonormally spans the space of images. Reference [13] discusses the attractive properties of this analysis, some of which we now summarize.

Multilinear analysis subsumes linear, PCA analysis. As shown in Fig. 3, each column of $\mathbf{U}_{\text{pixels}}$ is an "eigenimage". Since they were computed by performing an SVD of the matrix $\mathbf{D}_{(\text{pixels})}$ obtained as the mode-5 flattened data tensor $\mathcal{D}$, these eigenimages are identical to the conventional eigenfaces [6, 10]. Eigenimages represent only the principal axes of variation over all the images. The big advantage of multilinear analysis beyond linear PCA is that TensorFaces explicitly represent how the various factors interact to produce facial images. Tensorfaces are obtained by forming the product $\mathcal{Z} \times_5 \mathbf{U}_{\text{pixels}}$ (Fig. 4(a)).

The facial image database comprises 60 images per person that vary with viewpoint, illumination, and expression. PCA represents each person as a set of 60 vector-valued coefficients, one from each image in which the person appears. The length of each PCA coefficient vector is $28 \times 5 \times 4 \times 3 = 1680$. By contrast, multilinear analysis enables us to represent each person, regardless of viewpoint, illumination, and expression, with the same coefficient vector of dimension 28 relative to the bases comprising the $28 \times 5 \times 4 \times 3 \times 7943$ tensor

$$\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{illums}} \times_4 \mathbf{U}_{\text{expres}} \times_5 \mathbf{U}_{\text{pixels}}, \quad (5)$$

some of which are shown in Fig. 4(b). This many-to-one mapping is useful for face recognition. Each column in the figure is a basis matrix that comprises 28 eigenvectors. In any column, the first eigenvector depicts the average person

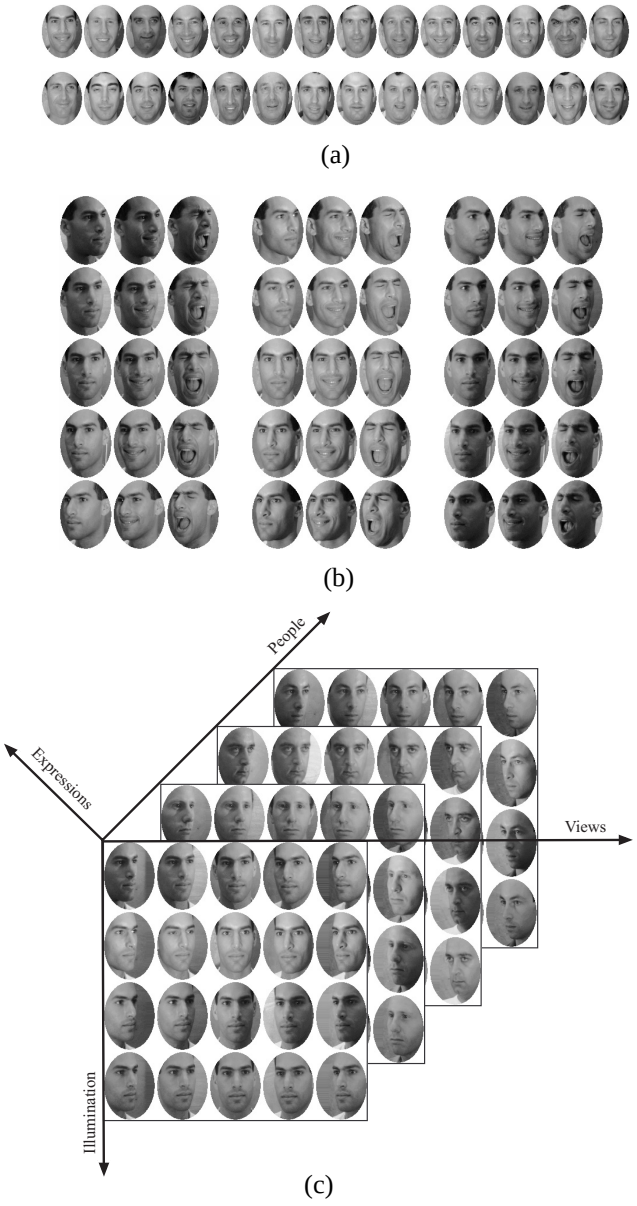
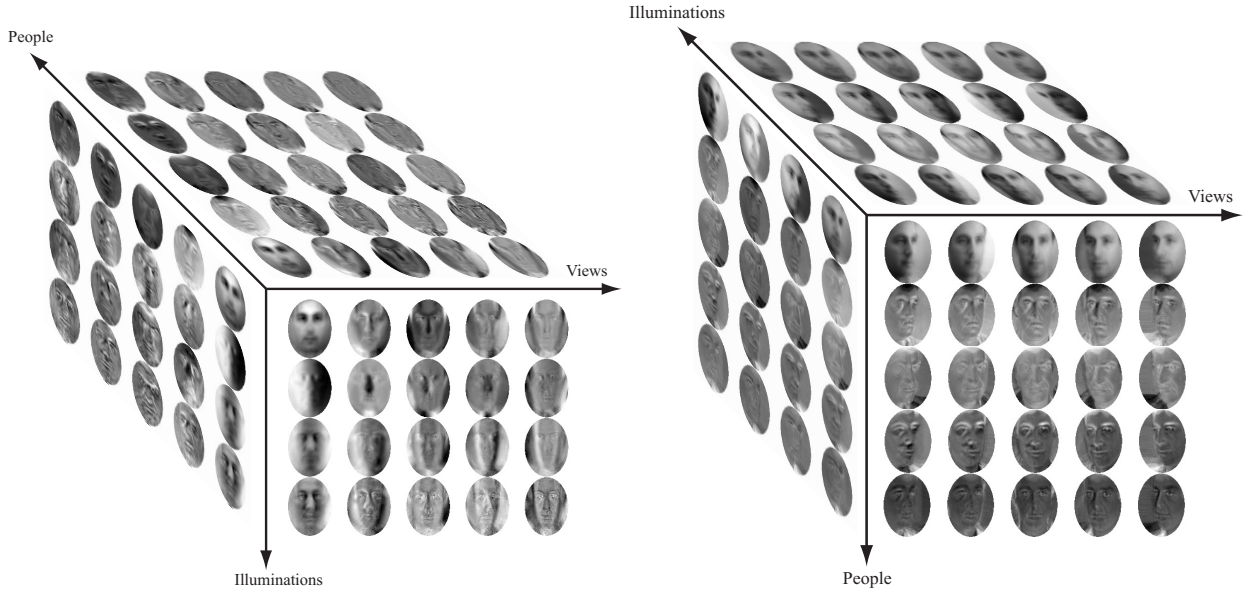


Figure 2: The facial image database (28 subjects, 60 images per subject). (a) The 28 subjects shown in expression 2 (smile), viewpoint 3 (frontal), and illumination 2 (frontal). (b) Part of the image set for subject 1. Left to right, the three panels show images captured in illuminations 1, 2, and 3. Within each panel, images of expressions 1, 2, and 3 (neural, smile, yawn) are shown horizontally while images from viewpoints 1, 2, 3, 4, and 5 are shown vertically. The image of subject 1 in (a) is the image situated at the center of (b). (c) The 5^{th}-order data tensor $\mathcal{D}$ for the image ensemble; only images in expression 1 (neutral) are shown.

Figure 3: $\mathbf{U}_{\text{pixels}}$ contains the PCA eigenvectors (eigenfaces), which are the principal axes of variation across all images.



(a) (b)

Figure 4: (a) A partial visualization of the $28 \times 5 \times 4 \times 3 \times 7943$ TensorFaces representation of $\mathcal{D}$, obtained as $\mathcal{T} = \mathcal{Z} \times_5 \mathbf{U}_{\text{pixels}}$ (only the subtensor of $\mathcal{T}$ associated with expression 1 (neutral) is shown). Note that the mode matrix $\mathbf{U}_{\text{pixels}}$ contains the conventional PCA eigenvectors or "eigenfaces", the first 10 of which are shown in Fig. 3, which are the principal axes of variation across all of the images. (b) A partial visualization of the $28 \times 5 \times 4 \times 3 \times 7943$ tensor $\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{illums}} \times_4 \mathbf{U}_{\text{expres}} \times_5 \mathbf{U}_{\text{pixels}}$ (again, only the subtensor associated with the neutral expression is shown), which defines 60 different bases for each combination of viewpoints, illumination and expressions. These bases have 28 eigenvectors which span the people space. The eigenvectors in any particular row play the same role in each column. The topmost plane depicts the average person, while the eigenvectors in the remaining planes capture the variability across people in the various viewpoint, illumination, and expression combinations.

and the remaining eigenvectors capture the variability over people, for the particular combination of viewpoint, illumination, and expression associated with that column. Each image is represented with a set of coefficient vectors representing the person, viewpoint, illumination and expression factors that generated the image. This is an important distinction that is relevant for image synthesis and recognition.

2.4 Face Recognition Using TensorFaces

We have proposed a recognition method based on multilinear analysis which uses the recognition bases in Fig. 4(b) (see [12] for the details). In our preliminary experiments with the Weizmann face image database, TensorFaces yields significantly better recognition rates than PCA (eigenfaces) in scenarios involving the recognition of people imaged in previously unseen viewpoints and illuminations.

In the first experiment, we trained our TensorFaces model on an ensemble comprising images of 23 people, captured from 3 viewpoints ($0, \pm 34$ degrees), with 4 illumination conditions (center, left, right, left+right). We tested our model on other images in this 23 person dataset acquired from 2 *different* viewpoints (± 17 degrees) under the same 4 illumination conditions. In this test scenario, the PCA method recognized the person correctly 61% of the time while TensorFaces recognized the person correctly 80% of the time.

In a second experiment, we trained our TensorFaces model on images of 23 people, 5 viewpoints ($0, \pm 17, \pm 34$ degrees), 3 illuminations (center light, left light, right light) and tested it on the 4^{th} illumination (left+right). PCA yielded a poor recognition rate of 27% while Tensorfaces achieved a recognition rate of 88%.

3 Dimensionality Reduction

Optimal dimensionality reduction in matrix PCA is obtained by truncating the SVD (i.e., deleting eigenvectors associated with the smallest eigenvalues). Unfortunately, optimal dimensionality reduction is not as straightforward in multilinear analysis.

3.1 Mode Matrix Truncation

A truncation of the mode matrices of the data tensor $\mathcal{D}$ results in an approximation $\hat{\mathcal{D}}$ with reduced ranks $R_1 \leq \bar{R}_1, R_2 \leq \bar{R}_2, \ldots, R_N \leq \bar{R}_N$, where $\bar{R}_n = \text{rank}_n(\mathcal{D}) = \text{rank}(\mathbf{D}_{(n)}) = \text{rank}(\mathbf{U}_n)$ is the n-rank of $\mathcal{D}$ for $1 \leq n \leq N$.

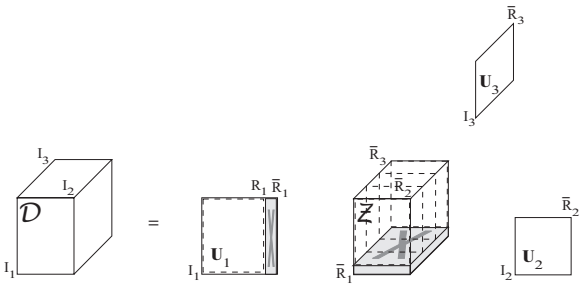


Figure 5: Data approximation through truncation. The data tensor $\mathcal{D}$ can be decomposed into the product of a core tensor $\mathcal{Z}$ and N mode matrices $\mathbf{U}_1 \ldots \mathbf{U}_N$; for the $N = 3$ case illustrated here, $\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$. Deletion of the last mode-1 eigenvector of $\mathbf{U}_1$ incurs an error in the approximation equal to $\sigma_{I_1}^2$, which equals the Frobenius norm of the (grey) subtensor of $\mathcal{Z}$ whose row vectors would normally multiply the eigenvector in the mode-1 product $\mathcal{Z} \times_1 \mathbf{U}_1$.

The error of this approximation is

$$\|\mathcal{D} - \hat{\mathcal{D}}\|^2 = \sum_{i_1=R_1+1}^{\bar{R}_1} \sum_{i_2=R_2+1}^{\bar{R}_2} \cdots \sum_{i_N=R_N+1}^{\bar{R}_N} \mathcal{Z}_{i_1 i_2 \ldots i_N}^2 \quad (6)$$

The error is bounded by the sum of squared singular values associated with the discarded singular vectors:

$$\|\mathcal{D} - \hat{\mathcal{D}}\|^2 \leq \sum_{i_1=R_1+1}^{\bar{R}_1} \sigma_{i_1}^2 + \sum_{i_2=R_2+1}^{\bar{R}_2} \sigma_{i_2}^2 + \cdots + \sum_{i_N=R_N+1}^{\bar{R}_N} \sigma_{i_N}^2. \quad (7)$$

Note that the singular value associated with the m^{th} singular vector in mode matrix $\mathbf{U}_n$ is equal to $\|\mathcal{Z}_{i_n=m}\|$; i.e., the Frobenius norm of subtensor $\mathcal{Z}_{i_n=m}$ of the core tensor $\mathcal{Z}$ (Fig. 5).

3.2 N-Mode Orthogonal Iteration Algorithm

Truncation of the mode matrices resulting from the N-mode SVD algorithm may yield a good reduced-dimensionality approximation $\hat{\mathcal{D}}$, but it is generally not optimal. A locally optimal dimensionality reduction scheme for tensors is to compute for $\mathcal{D}$ a *best rank-$(R_1, R_2, \ldots, R_N)$ approximation*[4] $\hat{\mathcal{D}} = \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \times_2 \hat{\mathbf{U}}_2 \ldots \times_N \hat{\mathbf{U}}_N$, with orthonormal $I_n \times R_n$ mode matrices $\hat{\mathbf{U}}_n$, for $n = 1, 2, \ldots, N$, which minimizes the error function [2, 3, 9]

$$e = \|\mathcal{D} - \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \ldots \times_N \hat{\mathbf{U}}_N\| + \sum_{i=1}^{N} \mathbf{\Lambda}_i \|\hat{\mathbf{U}}_i^T \hat{\mathbf{U}}_i - \mathbf{I}\|, \quad (8)$$

where the $\mathbf{\Lambda}_i$ are Lagrange multiplier matrices. To this end, our dimensionality-reducing N-*mode orthogonal iteration* algorithm (a higher-order extension of the orthogonal iteration for matrices) computes $\hat{\mathcal{D}}$ as follows:

1. Apply Step 1 of the N-mode SVD algorithm to $\mathcal{D}$; truncate each mode matrix $\mathbf{U}_n$, for $n = 1, 2, \ldots, N$, to R_n columns, thus obtaining the initial ($k = 0$) mode matrices $\mathbf{U}_1^0, \mathbf{U}_2^0, \ldots \mathbf{U}_N^0$.[5]

2. Iterate, for $k = 0, 1, \ldots$:

   2.1. Set $\tilde{\mathcal{U}}_1^{k+1} = \mathcal{D} \times_2 \mathbf{U}_2^{k\,T} \times_3 \mathbf{U}_3^{k\,T} \ldots \times_N \mathbf{U}_N^{k\,T}$; mode-1 flatten tensor $\tilde{\mathcal{U}}_1^{k+1}$ to obtain the matrix $\tilde{\mathbf{U}}_1^{k+1}$; set the columns of $\mathbf{U}_1^{k+1}$ to an orthonormal basis for the R_1-dimensional dominant subspace of $\tilde{\mathbf{U}}_1^{k+1}$.[6]

   2.2. Set $\tilde{\mathcal{U}}_2^{k+1} = \mathcal{D} \times_1 \mathbf{U}_1^{k+1\,T} \times_3 \mathbf{U}_3^{k\,T} \ldots \times_N \mathbf{U}_N^{k\,T}$; mode-2 flatten tensor $\tilde{\mathcal{U}}_2^{k+1}$ to obtain the matrix $\tilde{\mathbf{U}}_2^{k+1}$; set the columns of $\mathbf{U}_2^{k+1}$ to an orthonormal basis for the R_2-dimensional dominant subspace of $\tilde{\mathbf{U}}_2^{k+1}$.

   ...

   2.N. Set $\tilde{\mathcal{U}}_N^{k+1} = \mathcal{D} \times_1 \mathbf{U}_1^{k+1\,T} \times_2 \mathbf{U}_2^{k+1\,T} \ldots \times_{N-1} \mathbf{U}_{N-1}^{k+1\,T}$; mode-N flatten $\tilde{\mathcal{U}}_N^{k+1}$ to obtain the matrix $\tilde{\mathbf{U}}_N^{k+1}$; set the columns of $\mathbf{U}_N^{k+1}$ to an orthonormal basis for the R_N-dimensional dominant subspace of $\tilde{\mathbf{U}}_N^{k+1}$.

   until convergence: $||\mathbf{U}_n^{k+1\,T} \cdot \mathbf{U}_n^k||^2 > (1 - \epsilon) R_n$, for $1 \leq n \leq N$.

3. Set the converged mode matrices to $\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2, \ldots, \hat{\mathbf{U}}_N$. Compute the core tensor $\hat{\mathcal{Z}} = \tilde{\mathcal{U}}_N \times_N \hat{\mathbf{U}}_N^T$. The rank-reduced approximation of $\mathcal{D}$ is $\hat{\mathcal{D}} = \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \times_2 \hat{\mathbf{U}}_2 \ldots \times_N \hat{\mathbf{U}}_N$.

4 Dimensionality Reduction in Illumination

To illustrate the dimensionality reduction abilities of the N-mode orthogonal iteration algorithm presented in Section 3.2, we employ from the Weizmann facial image database an ensemble of images of 11 people, each photographed in neutral expression from a frontal viewpoint under 16 different illuminations. Fig. 6(a) shows three of the 176 original 7943-pixel images for one of the subjects.

[4] This best rank-$(R_1, R_2, \ldots, R_N)$ problem should not be confused with the classical "best rank-R" problem for tensors [4]: An N^{th}-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ has *rank* 1 when it is expressible as the outer product of N vectors: $\mathcal{A} = \mathbf{u}_1 \circ \mathbf{u}_2 \circ \ldots \circ \mathbf{u}_N$. The tensor element is expressed as $a_{ij\ldots m} = u_{1i} u_{2j} \ldots u_{Nm}$, where u_{1i} is the i^{th} component of $\mathbf{u}_1$, etc. The *rank* of a N^{th} order tensor $\mathcal{A}$, denoted $R = \text{rank}(\mathcal{A})$, is the minimal number of rank-1 tensors that yield $\mathcal{A}$ in a linear combination: $\mathcal{A} = \sum_{r=1}^{R} \sigma_r \mathbf{u}_1^{(r)} \circ \mathbf{u}_2^{(r)} \circ \ldots \circ \mathbf{u}_N^{(r)}$. Finding this minimal linear combination for a given tensor $\mathcal{A}$ is known as the best rank-R problem.

[5] The complexity of computing the SVD of an $m \times n$ matrix $\mathbf{A}$ (see Footnote 3) is $O(mn \min(m, n))$, which is costly when both m and n are large. However, we can compute the R leading singular factors of $\mathbf{A}$ efficiently by first computing the rank-R modified Gram-Schmidt (MGS) orthogonal factorization $\mathbf{A} \approx \mathbf{QR}$, where $\mathbf{Q}$ is $m \times R$ and $\mathbf{R}$ is $R \times n$, and then computing the SVD of $\mathbf{R}$ and multiplying it as follows: $\mathbf{A} \approx \mathbf{Q}(\tilde{\mathbf{U}}\mathbf{\Sigma}\mathbf{V}^T) = (\mathbf{Q}\tilde{\mathbf{U}})\mathbf{\Sigma}\mathbf{V}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

[6] We can compute $\mathbf{U}_1^{k+1}$ as the $I_1 \times R_1$ matrix whose columns are the first R_1 columns of the left matrix of the SVD of $\tilde{\mathbf{U}}_1^{k+1}$. For greater efficiency, we can proceed as suggested in Footnote 5.

Applying the N-mode orthogonal iteration algorithm, we compute in Step 1 an N-mode SVD of the $11\times16\times7943$ ensemble tensor $\mathcal{D}$ and obtain mode matrices $\mathbf{U}_{\text{people}}$, $\mathbf{U}_{\text{illums}}$, and $\mathbf{U}_{\text{pixels}}$ of dimension 11×11, 16×16, and 7943×176, respectively. We then truncate the illumination mode matrix $\mathbf{U}_{\text{illums}}$ from 16 columns to 3, thus obtaining the 16×3 reduced-rank matrix $\hat{\mathbf{U}}_{\text{illums}}$. We iterate in Step 2, updating $\hat{\mathbf{U}}_{\text{illums}}$ along with the other (non-truncated) mode matrices $\hat{\mathbf{U}}_{\text{people}}$ and $\hat{\mathbf{U}}_{\text{pixels}}$, until convergence (3 iterations).

Fig. 6(b) shows illumination-compressed images of the subject extracted from the dimensionality-reduced multilinear representation $\hat{\mathcal{D}} = \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_{\text{people}} \times_2 \hat{\mathbf{U}}_{\text{illums}} \times_3 \hat{\mathbf{U}}_{\text{pixels}}$. Note that the 81.25% reduction of the illumination dimensionality suppresses illumination effects such as shadows and highlights, but that it does not substantially degrade the appearance of the person, since the rank of the person mode matrix was not reduced. Increasing the illumination dimensionality to 6, the shadows and highlights begin to reappear, as shown in Fig. 6(c).

Thus, our multilinear model enables a *strategic* dimensionality reduction, which is more targeted than linear (PCA) dimensionality reduction. Fig. 7 compares TensorFaces image compression against PCA compression. Applying PCA compression, we retain in Fig. 7(b) the 11 (out of 176) most dominant eigenfaces and in Fig. 7(d) the 33 most dominant eigenfaces. Applying TensorFaces, we compress the dimensionality of the illumination mode from 16 to 1 ($R_{\text{illums}} = 1$) in Fig. 7(c) and from 16 to 3 ($R_{\text{illums}} = 3$) in Fig. 7(e). Since $R_{\text{people}} = 11$, in the first instance we retain 11×1 TensorFaces, while in the second we retain 11×3 TensorFaces, each time equaling the number of retained eigenfaces. Note that the total number of coefficients representing the compressed images is $11+1$ and $11+3$, respectively. Interestingly, the root mean squared errors (RMSE) relative to the original images, which are indicated in the figure, are higher for the TensorFaces compressions than they are for the PCA compressions. However, the "*perceptual error*" [8] of the TensorFaces compressions are significantly smaller, yielding substantially better image quality than PCA in subspaces of comparable dimension.

Figure 6: A subject was imaged under 16 different illuminations. (a) Three original images displaying different illumination conditions. (b) Compression of the images in (a) by reducing the illumination representation from 16 dimensions to 3 ($R_{\text{illums}} = 3$); i.e., $\hat{\mathbf{U}}_{\text{illums}}$ is reduced to a 16×3 matrix. This degrades the illumination effects (cast shadows, highlights). Arrows indicate the shadow cast by the nose in the original images (a) and the attenuated shadow in the compressed images (b). The shadow begins to reappear when the illumination dimensionality is increased to 6 ($R_{\text{illums}} = 6$) in (c); i.e., $\hat{\mathbf{U}}_{\text{illums}}$ a 16×6 matrix. Image sharpness and detailed facial features are well-preserved in both (b) and (c).

5 Conclusion

We have approached the analysis of an ensemble of images resulting from the confluence of multiple factors related to scene structure, illumination, and viewpoint as a problem in multilinear algebra. The ensemble is represented as a higher-order tensor. This image data tensor is decomposed into the product of a core tensor and several factor-specific mode matrices. The core tensor characterizes the interaction between the various factors, each of which is represented explicitly by a mode matrix whose orthonormal column vectors are factor-specific basis vectors.

We presented an N-mode orthogonal iteration algorithm for learning parsimonious, reduced-dimensionality multilinear models from raw image data tensors. The algorithm enables us to reduce the dimensionality (rank) of each mode matrix selectively. As an illustration of our technique, we demonstrated its ability to reduce significantly the dimensionality of the illumination subspace while not degrading other factors, such as facial appearance, preserving detailed facial features.

Our multilinear formulation accommodates any number of factors by exploiting tensor machinery. It subsumes as special cases the simple linear (1-factor) analysis known as principal components analysis (PCA), as well as bilinear (2-factor) analysis [7]. We are exploring several applications of multilinear analysis to computer vision and computer graphics; e.g., the synthesis and recognition of actions from human motion data [11] and the image-based rendering of textured surfaces [14].

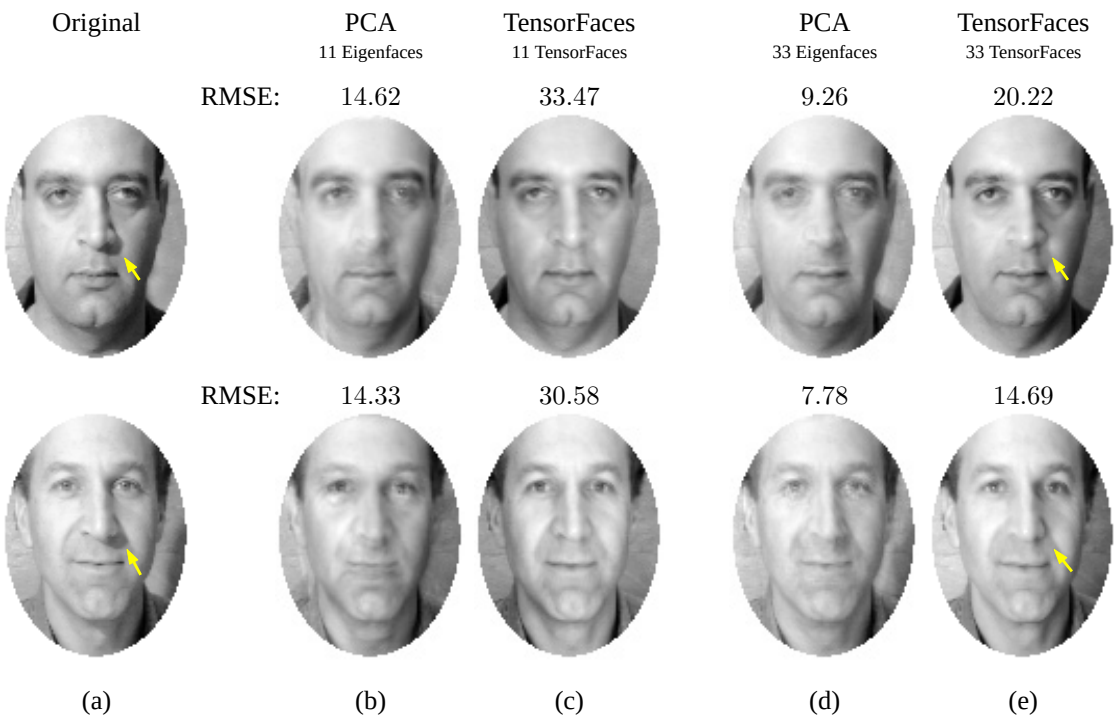


Figure 7: The "perceptual error" of TensorFaces compression of illumination is smaller than indiscriminate PCA compression in a subspace of comparable dimension. (a) Original images. (b) PCA image compression obtained by retaining the 11 most dominant eigenfaces. (c) TensorFaces image compression obtained by retaining 11 TensorFaces associated with $\hat{\mathbf{U}}_{\text{people}} \in \mathbb{R}^{11\times 11}$ and $\hat{\mathbf{U}}_{\text{illums}} \in \mathbb{R}^{16\times 1}$, which reduces the illumination representation from 16 dimensions to 1 ($R_{\text{illums}} = 1$). (d) PCA image compression obtained by retaining the 33 most dominant eigenfaces. (e) TensorFaces image compression obtained by retaining 33 TensorFaces associated with $\hat{\mathbf{U}}_{\text{people}} \in \mathbb{R}^{11\times 11}$ and $\hat{\mathbf{U}}_{\text{illums}} \in \mathbb{R}^{16\times 3}$, which reduces the illumination representation from 16 dimensions to 3 ($R_{\text{illums}} = 3$). Compared to the original images, the root mean squared errors (RMSE) of the PCA-compressed images are lower, yet the TensorFaces-compressed images have significantly better perceptual quality.

Acknowledgements

This research was funded in part by the Technical Support Working Group (TSWG) through the US Department of Defense's Combating Terrorism Technology Support Program.

References

[1] R. Chellappa, C.L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, May 1995.

[2] L. de Lathauwer, B. de Moor, and J. Vandewalle. On the best rank-1 and rank-($R_1, R_2, \ldots, R_n$) approximation of higher-order tensors. *SIAM Journal of Matrix Analysis and Applications*, 21(4):1324–1342, 2000.

[3] P.M. Kroonenberg and J. de Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1):69–97, 1980.

[4] J.B. Kruskal. Rank, decomposition, and uniqueness for 3-way and n-way arrays. In R. Coppi and S. Bolasco, editors, *Multiway Data Analysis*, pages 7–18, Amsterdam, 1989. North Holland.

[5] H. Murase and S. Nayar. Visual learning and recognition of 3D objects from appearance. *Int. J. of Computer Vision*, 14(1), 1995.

[6] L. Sirovich and M. Kirby. Low dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A.*, 4:519–524, 1987.

[7] J.B. Tenenbaum and W.T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1247–1283, 2000.

[8] P. Teo and D. Heeger. Perceptual image distortion. In *IEEE Conf. on Image Processing*, pages 982–986, Nov. 1994.

[9] L.R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[10] M.A. Turk and A.P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[11] M.A.O. Vasilescu. Human motion signatures: Analysis, synthesis, recognition. In *Proc. Int. Conf. on Pattern Recognition*, Quebec City, August 2002.

[12] M.A.O. Vasilescu and D. Terzopoulos. Multilinear analysis for facial image recognition. In *Proc. Int. Conf. on Pattern Recognition*, Quebec City, August 2002.

[13] M.A.O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: TensorFaces. In *Proc. European Conf. on Computer Vision (ECCV 2002)*, Copenhagen, Denmark, May 2002. 447-460.

[14] M.A.O. Vasilescu and D. Terzopoulos. TensorTextures. In *ACM SIGGRAPH 2003 Conf. Abstracts and Applications*, San Diego, August 2003.