To appear in the Proceedings of the *International Conference on Pattern Recognition (ICPR'02)*, Quebec City, Canada, August, 2002.

Multilinear Image Analysis for Facial Recognition

M. Alex O. Vasilescu
Department of Computer Science
University of Toronto
Toronto, ON M5S 3G4, Canada

Demetri Terzopoulos
Courant Institute
New York University
New York, NY 10003, USA

Abstract

Natural images are the composite consequence of multiple factors related to scene structure, illumination, and imaging. For facial images, the factors include different facial geometries, expressions, head poses, and lighting conditions. We apply multilinear algebra, the algebra of higher-order tensors, to obtain a parsimonious representation of facial image ensembles which separates these factors. Our representation, called TensorFaces, yields improved facial recognition rates relative to standard eigenfaces.

1 Introduction

People possess a remarkable ability to recognize faces when confronted by a broad variety of facial geometries, expressions, head poses, and lighting conditions. Developing a similarly robust computational model of face recognition remains a difficult open problem whose solution would have substantial impact on biometrics for identification, surveillance, human-computer interaction, and other applications.

Prior research has approached the problem of facial representation for recognition by taking advantage of the functionality and simplicity of linear algebra, the algebra of matrices. Principal components analysis (PCA) has been a popular technique in facial image recognition [1]. This method of linear algebra address single-factor variations in image formation. Thus, the conventional "eigenfaces" facial image recognition technique [9, 12] works best when person identity is the only factor that is permitted to vary. If other factors, such as lighting, viewpoint, and expression, are also permitted to modify facial images, eigenfaces face difficulty. Attempts have been made to deal with the shortcomings of PCA-based facial image representations in less constrained (multi-factor) situations; for example, by employing better classifiers [8].

Bilinear models have recently attracted attention because of their richer representational power. The *2-mode analysis* technique for analyzing (statistical) data matrices of scalar entries is described by Magnus and Neudecker [6]. 2-mode analysis was extended to vector entries by Marimont and Wandel [7] in the context of characterizing color surface and illuminant spectra. Tenenbaum and Freeman [10] applied this extension to three different perceptual tasks, including face recognition.

We have recently proposed a more sophisticated mathematical framework for the analysis and representation of image ensembles, which subsumes the aforementioned methods and which can account generally and explicitly for each of the multiple factors inherent to facial image formation [14]. Our approach is that of multilinear algebra—the algebra of higher-order tensors. The natural generalization of matrices (i.e., linear operators defined over a vector space), tensors define multilinear operators over a *set* of vector spaces. Subsuming conventional linear analysis as a special case, tensor analysis emerges as a unifying mathematical framework suitable for addressing a variety of computer vision problems. More specifically, we perform N-mode analysis, which was first proposed by Tucker [11], who pioneered 3-mode analysis, and subsequently developed by Kapteyn *et al.* [4, 6] and others, notably [2, 3].

In the context of facial image recognition, we apply a higher-order generalization of PCA and the singular value decomposition (SVD) of matrices for computing principal components. Unlike the matrix case for which the existence and uniqueness of the SVD is assured, the situation for higher-order tensors is not as simple [5]. There are multiple ways to orthogonally decompose tensors. However, one multilinear extension of the matrix SVD to tensors is most natural. We apply this *N-mode SVD* to the representation of collections of facial images, where multiple image formation factors, i.e., modes, are permitted to vary. Our *TensorFaces* representation separates the different modes underlying the formation of facial images. After reviewing TensorFaces in the next section, we demonstrate in Section 3 that TensorFaces show promise for use in a robust facial recognition algorithm.

2 TensorFaces

We have identified 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 [14]. Within this mathematical framework, the image ensemble is represented as a higher-dimensional tensor. This image data tensor $\mathcal{D}$ must be decomposed in order to separate and parsimoniously represent the constituent factors. To this end, we prescribe the *N-mode SVD* algorithm, a multilinear extension of the conventional matrix singular value decomposition (SVD).

Appendix A overviews the mathematics of our multilinear analysis approach and presents the N-mode SVD algorithm. In short, an order $N > 2$ tensor or N-way array $\mathcal{D}$ 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*, denoted $\times_n$ (see Equation (4) in Appendix A), of N-orthogonal spaces, as follows:

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_n \mathbf{U}_n \ldots \times_N \mathbf{U}_N. \quad (1)$$

Tensor $\mathcal{Z}$, known as the *core tensor*, is analogous to the diagonal singular value matrix in conventional matrix SVD (although 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 $\mathcal{D}$ (see Appendix A).

The multilinear analysis of facial image ensembles leads to the TensorFaces representation. To illustrate TensorFaces, we employed in our experiments a portion of the Weizmann face image database: 28 male subjects photographed in 5 viewpoints, 3 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. 1, yielding a total of 7943 pixels per image within the elliptical cropping window.

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

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{\text{people}} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{illums}} \times_4 \mathbf{U}_{\text{expres}} \times_5 \mathbf{U}_{\text{pixels}}, \quad (2)$$

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 3×3 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 7943×1260 mode matrix $\mathbf{U}_{\text{pixels}}$ orthonormally spans the space of images. Reference [14] discusses the attractive properties of this analysis, some of which we now summarize.

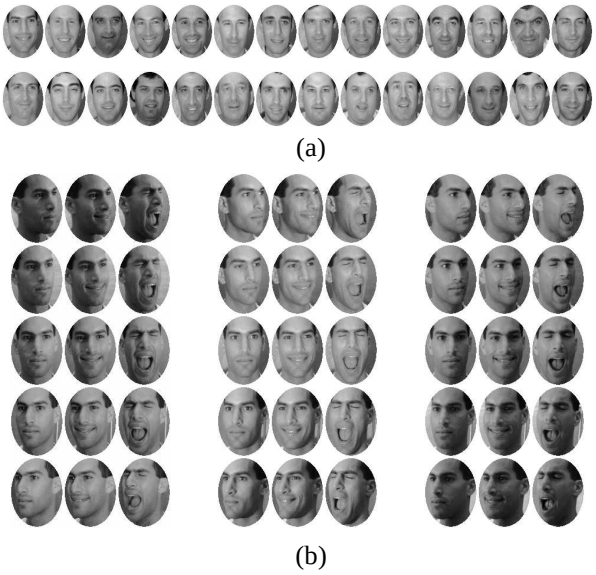

Figure 1: The facial image database (28 subjects $\times$ 45 images per subject). (a) The 28 subjects shown in expression 2 (smile), viewpoint 3 (frontal), and illumination 2 (frontal). (b) The full image set for subject 1. Left to right, the three panels show images captured in illuminations 1, 2, and 3. Within each panel, images of expressions 1, 2, and 3 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).

Our multilinear analysis subsumes linear, PCA analysis. As shown in Fig. 2(a), each column of $\mathbf{U}_{\text{pixels}}$ is an "eigenimage". These eigenimages are identical to conventional eigenfaces [9, 12], since the former were computed by performing an SVD on the mode-5 flattened data tensor $\mathcal{D}$ which yields the matrix $\mathbf{D}_{(\text{pixels})}$. The advantage of multilinear analysis, however, is that the core tensor $\mathcal{Z}$ can transform the eigenimages in $\mathbf{U}_{\text{pixels}}$ into TensorFaces, which represent the principal axes of variation across the various modes (people, viewpoints, illuminations, expressions) and represents how the various factors interact with each other to create the facial images. This is accomplished by simply forming the product $\mathcal{Z} \times_5 \mathbf{U}_{\text{pixels}}$. By contrast, the PCA basis vectors or eigenimages represent only the principal axes of variation across images.

Our facial image database comprises 45 images per person that vary with viewpoint, illumination, and expression. PCA represents each person as a set of 45 vector-valued coefficients, one from each image in which the person appears. The length of each PCA coefficient vector is $28 \times 5 \times 3 \times 3 = 1260$. By contrast, multilinear analysis enables us to represent each person with a single vector coefficient of dimension 28 relative to the bases comprising the $28 \times 5 \times 3 \times 3 \times 7943$ tensor

$$\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{illums}} \times_4 \mathbf{U}_{\text{expres}} \times_5 \mathbf{U}_{\text{pixels}}, \quad (3)$$

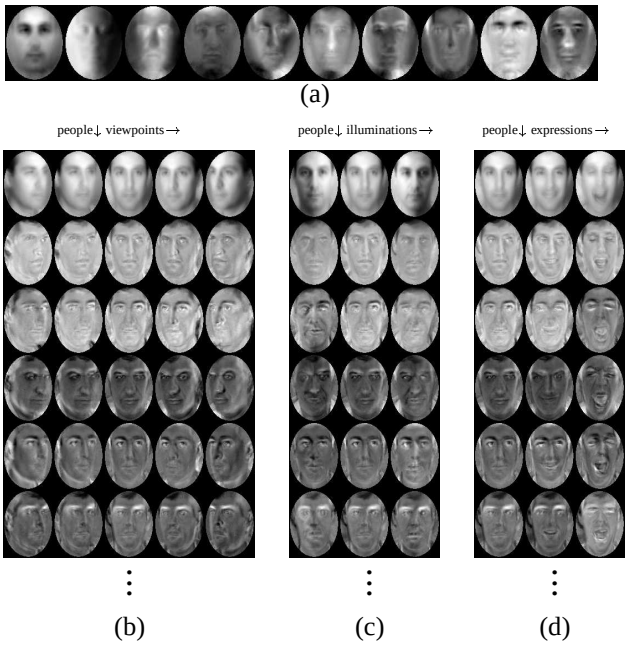

Figure 2: Some of the TensorFaces basis vectors resulting from the multilinear analysis of the facial image data tensor $\mathcal{D}$. (a) The first 10 PCA eigenvectors (eigenfaces), which are contained in the mode matrix $\mathbf{U}_{\text{pixels}}$, and are the principal axes of variation across all images. (b,c,d) A partial visualization of the $28 \times 5 \times 3 \times 3 \times 7943$ tensor $\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{illums}} \times_4 \mathbf{U}_{\text{expres}} \times_5 \mathbf{U}_{\text{pixels}}$, which defines 45 different bases for each combination of viewpoints, illumination and expressions, as indicated by the labels at the top of each array. 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 row across the three panels depicts the average person, while the eigenvectors in the remaining rows capture the variability across people in the various viewpoint, illumination, and expression combinations.

some of which are shown in Fig. 2(b–d). Each column in the figure is a basis matrix that comprises 28 eigenvectors. In any column, the first eigenvector depicts the average person and the remaining eigenvectors capture the variability across people, for the particular combination of viewpoint, illumination, and expression associated with that column.

3 Recognition Using TensorFaces

We propose a recognition method based on multilinear analysis analogous to the conventional one for linear PCA analysis. In the PCA or eigenface technique, one decomposes a data matrix $\mathbf{D}$ of known "training" facial images $\mathbf{d}_d$ into a reduced-dimensional basis matrix $\mathbf{B}_{\text{PCA}}$ and a matrix $\mathbf{C}$ containing a vector of coefficients $\mathbf{c}_d$ associated with each vectorized image $\mathbf{d}_d$. Given an unknown facial image $\mathbf{d}$, the projection operator $\mathbf{B}_{\text{PCA}}^{-1}$ linearly projects this new image into the reduced-dimensional space of image coefficients.

Our multilinear facial recognition algorithm performs the TensorFaces decomposition (2) of the tensor $\mathcal{D}$ of vectorized training images $\mathbf{d}_d$, extracts the matrix $\mathbf{U}_{\text{people}}$ which contains row vectors $\mathbf{c}_p^T$ of coefficients for each person p, and constructs the basis tensor $\mathcal{B}$ according to (3). We index into the basis tensor for a particular viewpoint v, illumination i, and expression e to obtain a subtensor $\mathcal{B}_{v,i,e}$ of dimension $28 \times 1 \times 1 \times 1 \times 7943$. We flatten $\mathcal{B}_{v,i,e}$ along the people mode to obtain the 28×7943 matrix $\mathbf{B}_{v,i,e(\text{people})}$. Note that a specific training image $\mathbf{d}_d$ of person p in viewpoint v, illumination i, and expression e can be written as $\mathbf{d}_{p,v,i,e} = \mathbf{B}_{v,i,e(\text{people})}^T \mathbf{c}_p$; hence, $\mathbf{c}_p = \mathbf{B}_{v,i,e(\text{people})}^{-T} \mathbf{d}_{p,v,i,e}$.

Now, given an unknown facial image $\mathbf{d}$, we use the projection operator $\mathbf{B}_{v,i,e(\text{people})}^{-T}$ to project $\mathbf{d}$ into a set of candidate coefficient vectors $\mathbf{c}_{v,i,e} = \mathbf{B}_{v,i,e(\text{people})}^{-T} \mathbf{d}$ for every v, i, e combination. Our recognition algorithm compares each $\mathbf{c}_{v,i,e}$ against the person-specific coefficient vectors $\mathbf{c}_p$. The best matching vector $\mathbf{c}_p$—i.e., the one that yields the smallest value of $||\mathbf{c}_{v,i,e} - \mathbf{c}_p||$ among all viewpoints, illuminations, and expressions—identifies the unknown image $\mathbf{d}$ as portraying person p.

As the following table shows, in our preliminary experiments with the Weizmann face image database, TensorFaces yields significantly better recognition rates than eigenfaces in scenarios involving the recognition of people imaged in previously unseen viewpoints (row 1) and under a previously unseen illumination (row 2):

Recognition Experiment	PCA	TensorFaces
Training: 23 people, 3 viewpoints ($0^\circ, \pm 34^\circ$), 4 illuminations Testing: 23 people, 2 viewpoints ($\pm 17^\circ$), 4 illuminations (center, left, right, left+right)	61%	80%
Training: 23 people, 5 viewpoints ($0^\circ, \pm 17^\circ, \pm 34^\circ$), 3 illuminations Testing: 23 people, 5 viewpoints ($0^\circ, \pm 17^\circ, \pm 34^\circ$), 4th illumination	27%	88%

4 Conclusion

We have approached the analysis of an ensemble of facial images resulting from the confluence of multiple factors related to scene structure, illumination, and viewpoint as a problem in multilinear algebra in which the image ensemble is represented as a higher-dimensional tensor. Using the "N-mode SVD" algorithm, a multilinear extension of the conventional matrix singular value decomposition (SVD), this image data tensor is decomposed in order to separate and parsimoniously represent the constituent factors. Our analysis subsumes as special cases the simple linear (1-factor) analysis associated with conventional SVD and principal components analysis (PCA), as well as the incrementally more general bilinear (2-factor) analysis that has recently been investigated in computer vision. Our completely general multilinear approach accommodates any number of factors by exploiting tensor machin-

ery and, in our experiments, it yields significantly better recognition rates than standard eigenfaces.

We plan to investigate dimensionality reduction in conjunction with TensorFaces (refer to the final paragraph $\Longrightarrow$ of Appendix A). See [13] in these proceedings for the application of multilinear analysis to the recognition of people and actions from human motion data.

A Multilinear Analysis

A *tensor* is a higher order generalization of a vector (first order tensor) and a matrix (second order tensor). Tensors are multilinear mappings over a set of vector spaces. The *order* of tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is N. 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 Nth 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 $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N)}$ that results by *mode-n flattening* the tensor $\mathcal{A}$ (see Fig. 1 in [14]).

A generalization of the product of two matrices is the product of a tensor and a matrix. The *mode-n product* of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n \times \ldots \times I_N}$ by a matrix $\mathbf{M} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{M}$, is the $I_1 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N$ tensor

$$(\mathcal{A} \times_n \mathbf{M})_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} m_{j_n i_n}. \quad (4)$$

The mode-n product can be expressed in terms of flattened matrices as $\mathbf{B}_{(n)} = \mathbf{M}\mathbf{A}_{(n)}$.[1]

Our *N-mode SVD algorithm* for decomposing $\mathcal{D}$ according to equation (1) is:

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

2. Solve for the core tensor as follows:

$$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \ldots \times_n \mathbf{U}_n^T \ldots \times_N \mathbf{U}_N^T. \quad (5)$$

[1] 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 $(\mathcal{A} \times_n \mathbf{U}) \times_n \mathbf{V} = \mathcal{A} \times_n (\mathbf{V}\mathbf{U})$.

[2] When $\mathbf{D}_{(n)}$ is a non-square matrix, the computation of $\mathbf{U}_n$ in the singular value decomposition (SVD) $\mathbf{D}_{(n)} = \mathbf{U}_n \mathbf{\Sigma} \mathbf{V}_n^T$ can be performed efficiently, depending on which dimension of $\mathbf{D}_{(n)}$ is smaller, by decomposing either $\mathbf{D}_{(n)} \mathbf{D}_{(n)}^T = \mathbf{U}_n \mathbf{\Sigma}^2 \mathbf{U}_n^T$ and then computing $\mathbf{V}_n^T = \mathbf{\Sigma}^+ \mathbf{U}_n^T \mathbf{D}_{(n)}$ or by decomposing $\mathbf{D}_{(n)}^T \mathbf{D}_{(n)} = \mathbf{V}_n \mathbf{\Sigma}^2 \mathbf{V}_n^T$ and then computing $\mathbf{U}_n = \mathbf{D}_{(n)} \mathbf{V}_n \mathbf{\Sigma}^+$.

Dimensionality reduction in matrix principal component analysis is obtained by truncation of the singular value decomposition (i.e., deleting eigenvectors associated with the smallest eigenvalues). Unfortunately, this does not have a trivial multilinear counterpart. According to [3], a useful generalization to tensors involves an optimal rank-$(R_1, R_2, \ldots, R_N)$ approximation which iteratively optimizes each of the modes of the given tensor, where each optimization step involves a best reduced-rank approximation of a positive semi-definite symmetric matrix. This technique is a higher-order extension of the orthogonal iteration for matrices.

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. A multilinear singular value decomposition. *SIAM Journal of Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[3] 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.

[4] A. Kapteyn, H. Neudecker, and T. Wansbeek. An approach to n-mode component analysis. *Psychometrika*, 51(2):269–275, June 1986.

[5] T. G. Kolda. Orthogonal tensor decompositions. *SIAM J. on Matrix Analysis and Applications*, 23(1):243–255, 2001.

[6] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York, New York, 1988.

[7] D.H. Marimont and B.A. Wandell. Linear models of surface and illuminance spectra. *J. Optical Society of America, A.*, 9:1905–1913, 1992.

[8] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1994.

[9] 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.

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

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

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

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

[14] 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. In press.