Multilinear Projection for Appearance-Based Recognition in the Tensor Framework

M. Alex O. Vasilescu
Massachusetts Institute of Technology
www.media.mit.edu/~maov

Demetri Terzopoulos
University of California, Los Angeles
www.cs.ucla.edu/~dt

Abstract

Numerical multilinear (tensor) algebra is a principled mathematical approach to disentangling and explicitly and parsimoniously representing the essential factors or modes of image formation, among them illumination, scene geometry, and imaging, thereby dramatically improving the performance of appearance-based recognition. Generalizing concepts from linear (matrix) algebra, we define the identity tensor and the pseudo-inverse tensor and we employ them to develop a multilinear projection algorithm, which is natural for performing recognition in the tensor algebraic framework. Our multilinear projection algorithm simultaneously projects an unlabeled test image into multiple constituent mode spaces spanned by learned, mode-specific basis sets in order to infer its mode labels. Multilinear projection is applied to unconstrained facial image recognition, where the mode labels are person identity, viewpoint, illumination, etc.

1. Introduction

Appearance-based methods have attracted a great deal of attention in computer vision. Traditional appearance-based image analysis and recognition employs principal components analysis (PCA) to parsimoniously model the variation apparent in training image ensembles, typically using the well-known matrix singular value decomposition (SVD) and dimensionality reduction through the omission of higher-order singular vectors. Conventional PCA, which is a linear method, models the apparent variation as if it had resulted from a single contributory factor. However, natural images result from the interaction of *multiple* factors or modes related to scene structure, illumination, and imaging. For example, facial images are the result of facial geometry (person, expression), the pose of the head relative to the camera, the lighting conditions, and the type of camera employed. This multifactor variation causes severe difficulties for conventional appearance-based face recognition methods. In particular, the PCA approach and its variants adequately address face recognition only under tightly constrained conditions—e.g., frontal images, fixed light-sources, fixed expression—where person identity is the only factor that is allowed to vary.

We have introduced a more powerful, tensor algebraic framework for the appearance-based analysis and recognition of images that elegantly and effectively deals with the multifactor variation inherent to image formation [9, 8]. Our nonlinear approach exploits multilinear algebra—the algebra of higher-order tensors. The TensorFaces method [9] employs multilinear PCA (MPCA) and a natural multilinear extension of the matrix SVD to tensors, known as the N*-mode SVD*, to analyze ensembles of facial images in which multiple modes are permitted to vary. The method separates and parsimoniously represents each of the different modes underlying the formation of facial images, enabling mode-specific dimensionality reduction. A facial image is represented by a set of coefficient vectors, one for each constituent mode or factor—person, viewpoint, illumination, expression, etc. The tensor algebraic approach was shown to be promising for facial recognition [8].

Our contribution in this paper is an extension of the tensor approach that is significant for recognition. Generalizing concepts from linear algebra, we introduce the *mode-*n *identity tensor* and the *mode-*n *pseudo-inverse tensor* and employ them to develop a *multilinear projection algorithm* for recognition in the tensor framework. This novel algorithm simultaneously projects an unlabeled test image into multiple constituent mode spaces in order to infer its set of mode labels. We demonstrate our multilinear projection in the context of unconstrained facial image recognition, where the mode labels are the person identity, the viewpoint, the illumination, the expression, etc.

The remainder of this paper is organized as follows: For readers unfamiliar with the topic from our prior papers, we cover the terminology and relevant fundamentals of multilinear algebra in Appendix A. In Section 2, we review our multilinear image analysis approach and, in particular, the TensorFaces method. Section 3 develops our multilinear

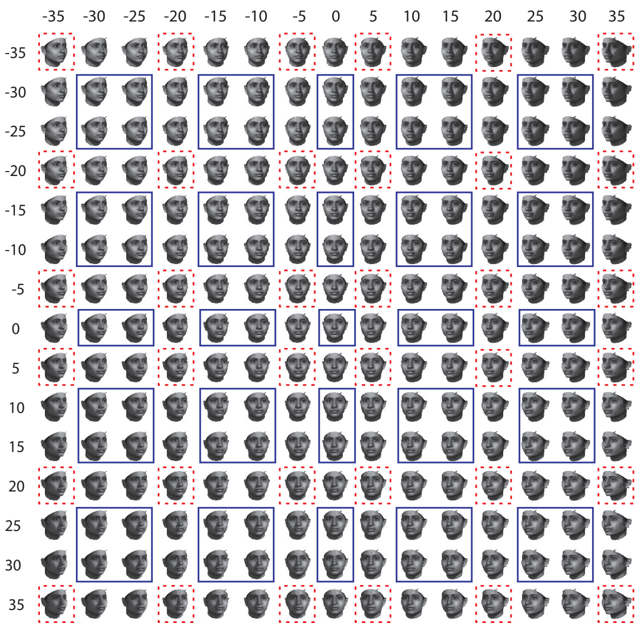


Figure 1. Facial image data. 3D scans of 75 subjects were acquired using a Cyberware 3030PS laser scanner as part of the University of Freiburg 3D morphable faces database [1]. The facial images from one of the subjects are shown, viewed from 15 different viewpoints (across) ($\theta = -35^\circ$ to $+35^\circ$ in 5° steps on the horizontal plane $\phi = 0^\circ$) under 15 different illuminations (down) ($\theta = -35^\circ$ to $+35^\circ$ in 5° steps on an inclined plane $\phi = 45^\circ$). In our recognition experiments, the dash-boxed images served as training images; the solid-boxed images served as test images.

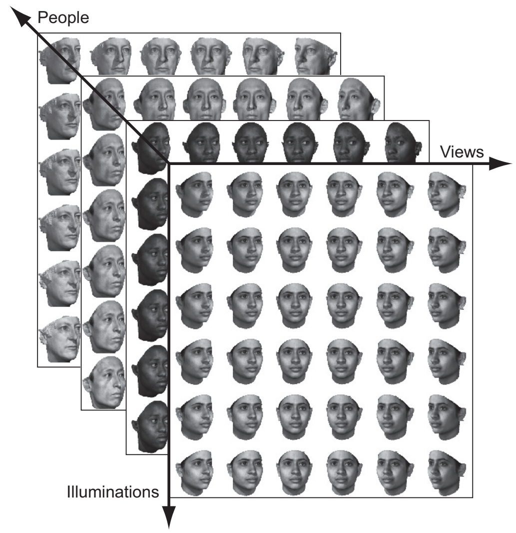


Figure 2. A portion of the fourth-order data tensor $\mathcal{D}$ for the image ensemble formed from the dash-boxed images of each subject. Only 4 of the 75 people are shown.

projection algorithm. Section 4 reviews its application to facial recognition under unconstrained conditions.

2. Multilinear Facial Image Analysis

Progressing beyond PCA, which can model only single-factor variations in image ensembles, we have been confronting the fact that natural images result from the interaction of multiple factors related to scene structure, illumination, and imaging. The analysis of an ensemble of images resulting from the confluence of multiple factors, or modes, is a problem that is addressable using multilinear algebra. Within our multilinear framework, the image ensemble is organized as a higher-order data tensor. This image data tensor $\mathcal{D}$ is decomposed in order to separate and, through dimensionality reduction, parsimoniously represent the constituent factors. The TensorFaces method [9] is a particular multilinear method that addresses facial image ensembles, whose relevant factors include different facial geometries and expressions, head poses, and lighting conditions.

We illustrate the TensorFaces analysis using gray-level facial images of 75 subjects (Fig. 1). Each subject is imaged from 15 different viewpoints under 15 different illuminations. Fig. 1(b) shows the set of 225 images for one of the subjects, with viewpoints arrayed horizontally and illuminations arrayed vertically. The image set was rendered from a 3D scan of the subject. Each image is $80 \times 107 = 8560$ pixels in size. The 75 scans were acquired using a Cyberware 3030PS laser scanner and are part of the 3D morphable faces database created at the University of Freiburg [1].

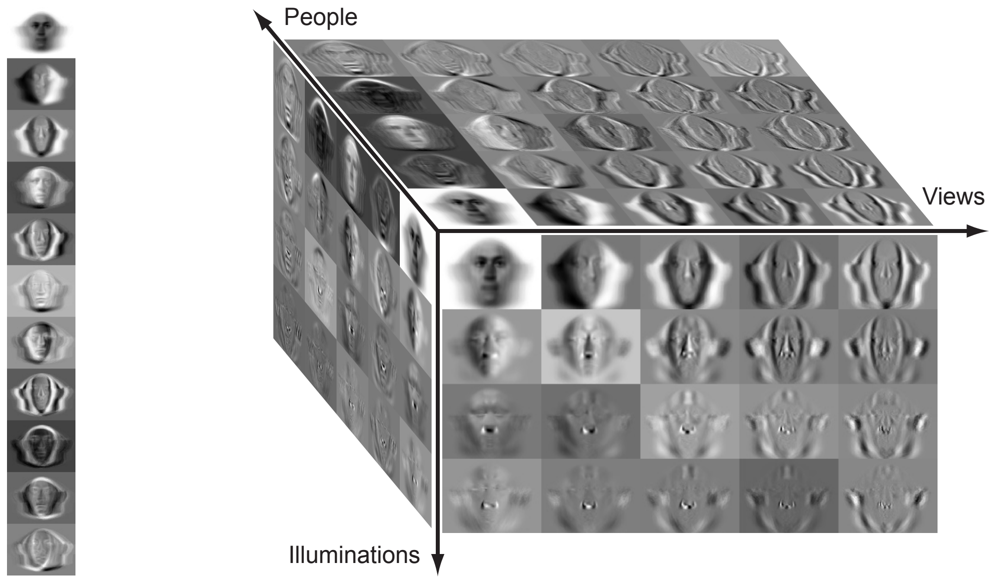
We select an ensemble of images from the dataset comprising for each person the dash-boxed images shown in Fig. 1. Thus, our facial image data tensor $\mathcal{D}$ is a $6 \times 6 \times 75 \times 8560$ tensor (Fig. 2). Applying multilinear analysis, the N-mode SVD (refer to Appendix A.3), with $N = 4$, of $\mathcal{D}$ is

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{\text{illums}} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{people}} \times_4 \mathbf{U}_{\text{pixels}}, \quad (1)$$

where the $6 \times 6 \times 75 \times 8560$ core tensor $\mathcal{Z}$ governs the interaction between the factors represented in the mode matrices—the 6×6 mode matrix $\mathbf{U}_{\text{illums}}$ spans the space of illumination parameters and contains row vectors $\mathbf{c}_l^T$ of coefficients for each illumination direction l, the 6×6 mode matrix $\mathbf{U}_{\text{views}}$ spans the space of viewpoint parameters and contains row vectors $\mathbf{c}_v^T$ of coefficients for each view direction v, the 75×75 mode matrix $\mathbf{U}_{\text{people}}$ spans the space of people parameters and contains row vectors $\mathbf{c}_p^T$ of coefficients for each person p, and the 8560×2700 mode matrix $\mathbf{U}_{\text{pixels}}$ orthonormally spans the space of images, and its columns are conventional "eigenfaces" (Fig. 3(a)).

The big advantage of multilinear analysis over linear PCA (i.e., the eigenfaces basis shown in Fig. 3(a)) is that the TensorFaces bases (Fig. 3(b)) explicitly represent how the various factors interact to produce facial images. TensorFaces, which are expressed as

$$\mathcal{T} = \mathcal{Z} \times_4 \mathbf{U}_{\text{pixels}}, \quad (2)$$



(a) (b)

Figure 3. Eigenfaces and TensorFaces bases for an ensemble of 2,700 facial images spanning 75 people, each imaged under 6 viewing and 6 illumination conditions. (a) PCA eigenvectors (eigenfaces) $\mathbf{U}_{\text{pixels}}$, which are the principal axes of variation across all images. (b) A partial visualization of the $6 \times 6 \times 75 \times 8560$ TensorFaces $\mathcal{T} = \mathcal{Z} \times_4 \mathbf{U}_{\text{pixels}}$ representation obtained from $\mathcal{D}$. Tensorfaces captures the illumination-mode, view-mode, and person-mode variation across all images.

explicitly represent the mode-specific variation across all images. Note that, in practice, $\mathcal{T}$ can be computed directly by the N-mode SVD without explicitly computing the usually large $\mathbf{U}_{\text{pixels}}$.

Our facial image database comprises 75 people and 36 images per person that vary with viewpoint and illumination. Note that conventional PCA represents each image with a unique coefficient vector and each person with a set of 36 coefficient vectors, one for each image in which the person appears. By contrast, multilinear analysis enables us to represent each person, regardless of viewpoint and illumination by a single invariant coefficient vector of dimension 75, while each image is represented with a set of coefficient vectors representing the person, viewpoint, and illumination factors associated with the image. This important distinction between PCA and MPCA is relevant for facial image recognition, a topic that we will consider next.

3. Multilinear Projection

3.1. Motivation

We will now discuss how to infer from an unlabeled test image the coefficient vectors associated with the multiple image formation factors. The recognition algorithm for TensorFaces proposed in [8] obtained coefficient vectors based on a linear projection approach. It computed a set of linear projection operators, which yielded a set of candidate coefficient vectors for recognition. Multiple linear projections are less than ideal. We will now develop a novel projection method for appearance-based recognition that fully exploits the multilinear structure of the tensor framework.

Our unified, multilinear projection method simultaneously infers the mode coefficient vectors of an unknown test image, by projecting it from the pixel space into the N different constituent mode spaces that are of relevance. Fig. 4 illustrates this in the context of face recognition, where we analyze facial images and the relevant modes are person identity, illumination, viewpoint, etc.

For concreteness, we will continue the development of our superior approach using the facial image dataset of Fig. 1 and MPCA. Given the data tensor $\mathcal{D}$ (Fig. 2) of labeled, vectorized training images $\mathbf{d}_{p,v,l}$, where the subscripts are person, view, and illumination labels, our method first performs an MPCA decomposition (1), extracting mode matrices $\mathbf{U}_{\text{illums}}$, $\mathbf{U}_{\text{views}}$, and $\mathbf{U}_{\text{people}}$, as well as the TensorFaces basis $\mathcal{T}$ (2) that governs the interaction between them (Fig. 3(b)). Then the method represents an unlabeled, test image $\mathbf{d}$ by the relevant set of coefficient vectors as follows:

$$\mathbf{d} = \mathcal{T} \times_1 \mathbf{c}_l^T \times_2 \mathbf{c}_v^T \times_3 \mathbf{c}_p^T, \tag{3}$$

a multilinear representation that is illustrated in Fig. 5(a).

Given $\mathbf{d}$ and $\mathcal{T}$, we must determine the unknown variables, $\mathbf{c}_l$, $\mathbf{c}_v$, and $\mathbf{c}_p$ in order to determine the illumination, viewpoint, and person associated with the test image. This involves computing an inverse (or pseudo-inverse) of tensor

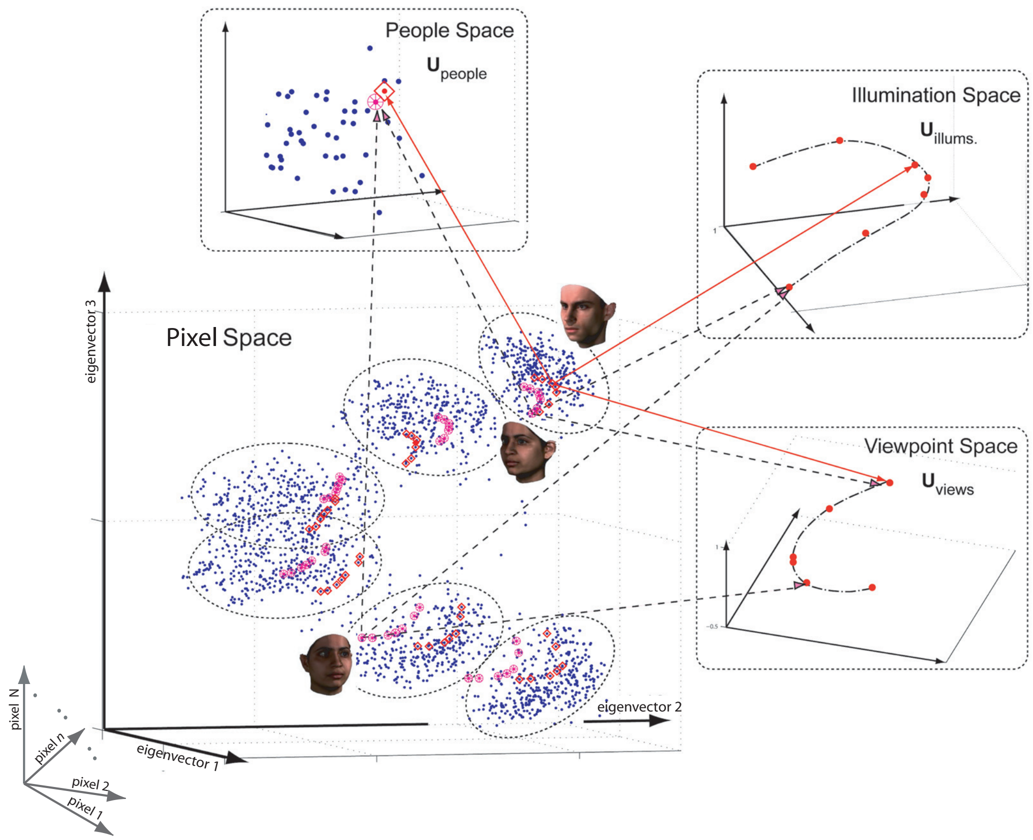
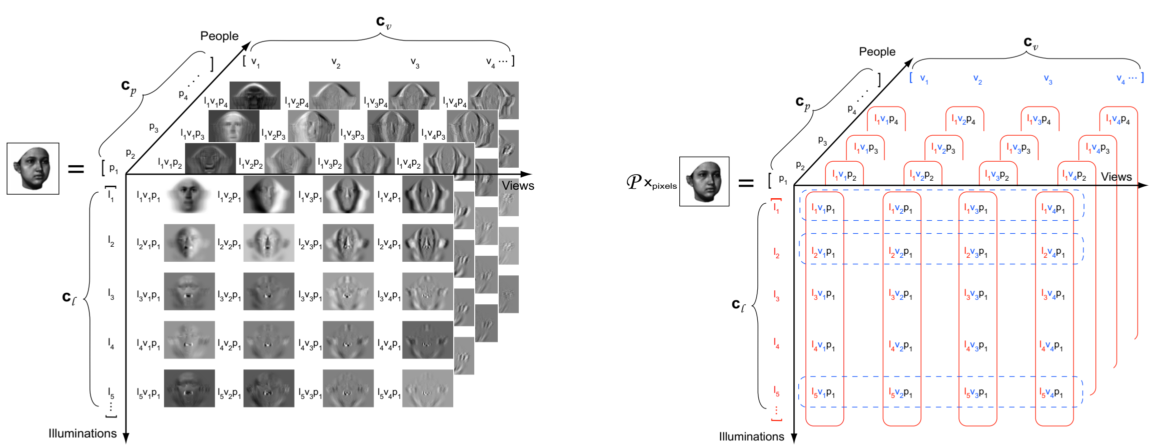
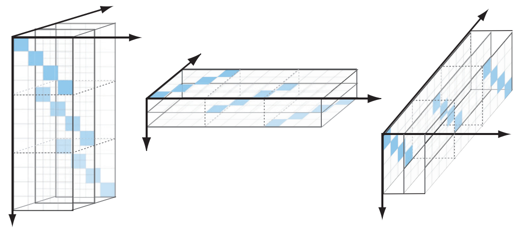


Figure 4. An illustration of the multilinear projection. In this example, 75 people were imaged from 6 different viewing directions under 7 different illumination conditions. Each blue dot in the pixel space represents one of the 3150 acquired images. For clarity, each image is represented by 3 coefficients associated with the 3 most significant eigenvectors. The multilinear projection simultaneously projects each image to multiple mode spaces (person, illumination, viewpoint). Note the manifold structure of the analyzed images. They form 6 clusters in the pixel space, each comprising all the images acquired from the same viewing direction. The clusters evidently have a semi-circular arrangement within the pixel space that can be parameterized by the viewing direction. Similarly, within each cluster, images for each individual also exhibit a semi-circular arrangement that can be parameterized by the illumination direction. The red diamonds denote images of one person under various illuminations, while the purple circles denote images of another person under various illuminations.



(a) (b)

Figure 5. (a) Image representation $\mathbf{d} = \mathcal{T} \times_1 \mathbf{c}_l^T \times_2 \mathbf{c}_v^T \times_3 \mathbf{c}_p^T$. (b) Given an unlabeled test image $\mathbf{d}$, the associated coefficient vectors $\mathbf{c}_l$, $\mathbf{c}_v$, $\mathbf{c}_p$ are computed by decomposing the response tensor $\mathcal{R} = \mathcal{P} \times_4 \mathbf{d}^T$ using the N-mode SVD algorithm.



(a) (b) (c)

Figure 6. The three identity tensors of order 3; (a) mode-1 identity tensor; (b) mode-2 identity tensor; (c) mode-3 identity tensor.

$\mathcal{T}$, which raises some related questions. How does one invert a tensor? When one multiplies a tensor with its inverse tensor, what should the resulting identity tensor be? Unfortunately, the identity tensor is not the obvious generalization of the identity matrix; i.e., it is not a tensor with ones along the super diagonal. We address these questions next.

3.2. Identity and (Pseudo-) Inverse Tensors

First, we will extend the definition of the mode-n product of a tensor with a matrix (Definition 8 in Appendix A.1) to the mode-n product between two tensors:

Definition 1 (Generalized Mode-n Product, $\times_n$) *The mode-n product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_n\times\ldots\times I_N}$ and a tensor $\mathcal{B} \in \mathbb{R}^{J_1\times J_2\times\ldots\times J_n\times\ldots\times J_N}$, where $I_n = J_1J_2\ldots J_{n-1}J_{n+1}\ldots J_N$, is denoted by $\mathcal{C} = \mathcal{A}\times_n\mathcal{B}$, where $\mathcal{C} \in \mathbb{R}^{I_1\times\ldots\times I_{n-1}\times J_n\times I_{n+1}\times\ldots\times I_N}$ can be expressed in terms of matrices as $\mathbf{C}_{(n)} = \mathbf{B}_{(n)}\mathbf{A}_{(n)}$.*

Second, unlike the identity matrix, identity tensors are not diagonal tensors.

Definition 2 (Mode-n Identity Tensor) *$\mathcal{I}_n$ is a mode-n multiplicative identity tensor iff $\mathcal{I}_n\times_n\mathcal{A} = \mathcal{A}$, where $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_n\times\ldots\times I_N}$ and $\mathcal{I}_n \in \mathbb{R}^{I_1\times I_2\ldots\times J_n\ldots I_N}$, where $J_n = I_{n+1}\ldots I_NI_1I_2\ldots I_{n-1}$. The mode-$n$ identity tensor $\mathcal{I}_n$ is a tensorized identity matrix of dimensionality $J_n\times(J_{n+1}\ldots J_NJ_1\ldots J_{n-1})$.*

Fig. 6 illustrates the structure of the three identity tensors of order 3. Although the existence of an identity tensor for every mode might seem a construct peculiar to multilinear algebra, one should recall that there exist left and right identity matrices in linear algebra for every matrix $\mathbf{A} \in \mathbb{R}^{I_1\times I_2}$. Whereas the left and right identity matrices have different dimensions, they share the same diagonal structure, unlike the case for higher-order tensors.

Next, generalizing the definition of an inverse matrix, we define the mode-n inverse tensor as follows:

Definition 3 (Mode-n Inverse Tensor, $\mathcal{A}^{-_n}$) *Tensor $\mathcal{B}$ is a mode-n inverse of tensor $\mathcal{A}$ if and only if $\mathcal{A}\times_n\mathcal{B} = \mathcal{I}_n$ and $\mathcal{B}\times_n\mathcal{A} = \mathcal{I}_n$, where $\mathcal{I}_n$ is the mode-n identity tensor.*

Finally, we define a *mode-n pseudo-inverse tensor* that generalizes the left pseudo-inverse matrix $\mathbf{A}^+\mathbf{A}$ and right pseudo-inverse matrix $\mathbf{A}\mathbf{A}^+$ from linear algebra.

Definition 4 (Mode-n Pseudo-Inverse Tensor) *The mode-n pseudoinverse tensor, $\mathcal{A}^{+_n}$, of tensor $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_N}$ satisfies:*

1. $(\mathcal{A}\times_n\mathcal{A}^{+_n})\times_n\mathcal{A} = \mathcal{A}$
2. $(\mathcal{A}^{+_n}\times_n\mathcal{A})\times_n\mathcal{A}^{+_n} = \mathcal{A}^{+_n}$

The mode-n pseudoinverse tensor $\mathcal{A}^{+_n}$ of $\mathcal{A}$ is the tensorized pseudoinverse of $\mathbf{A}_{(n)}$; i.e., the mode-n flattened version of $\mathcal{A}^{+_n}$ is $\mathbf{A}^+_{(n)}$.

3.3. Multilinear Projection Algorithm

Since an image is a point (vector) in pixel space, then all mathematical operations must be performed in that space. Hence, $\mathcal{T}$ must be inverted with respect to the pixel mode to obtain the projection tensor $\mathcal{P}$. Projecting the unlabeled test image $\mathbf{d}$ onto the pixel mode of $\mathcal{T}$ yields the image response tensor

$$\mathcal{R} = \mathcal{P}\times_N\mathbf{d}^T, \tag{4}$$

where $\mathcal{P}$ is obtained by re-tensorizing matrix $\mathbf{P}_{\text{(pixels)}} = \mathbf{T}^{+T}_{\text{(pixels)}}$, matrix $\mathbf{T}_{\text{(pixels)}}$ being the pixel-mode flattened $\mathcal{T}$.

Lemma 1 *$\mathcal{R}$ has rank-$(1,\ldots,1)$.*
Proof:

$$\begin{aligned}\mathcal{R} &= \mathcal{P}\times_N\mathbf{d}^T && (5)\\ &\approx \mathcal{I}_1\times_N(\mathbf{c}^T_{N-1}\otimes\mathbf{c}^T_{N-2}\ldots\otimes\mathbf{c}^T_1) && (6)\\ &= \mathbf{c}_1\circ\mathbf{c}_2\ldots\circ\mathbf{c}_{N-1}. && (7)\end{aligned}$$

Since $\mathcal{R}$ is expressible as an outer product of coefficient vectors associated with the factors inherent to $\mathbf{d}$, it is of rank-$(1,\ldots,1)$.

In principle then, we can compute the coefficient vectors by applying the N-mode SVD to decompose the response tensor as follows:

$$\mathcal{R} = \mathcal{Z}\times_1\mathbf{c}_1\ldots\times_n\mathbf{c}_n\ldots\times_{N-1}\mathbf{c}_{N-1}. \tag{8}$$

The *multilinear projection* algorithm is detailed in Fig. 7. The reason for Step 5 in the algorithm is that, due to numerical errors, $\mathcal{R}$ may not be precisely rank-$(1,\ldots,1)$ in practice, so it is generally necessary to compute an optimal rank-$(1,\ldots,1)$ approximation, which is accomplished in an iterative manner, one mode at a time, by the alternating least squares method.

The multilinear projection algorithm

Given a (TensorFaces) basis set $\mathcal{T}$ and an unknown test image $\mathbf{d}$:

1. Compute the projection operator $\mathcal{P}$. In matrix form $\mathbf{P}_{\text{(pixels)}} = \mathbf{T}^{T+}_{\text{(pixels)}}$.
2. Compute the response tensor $\mathcal{R} = \mathcal{P} \times_N \mathbf{d}^T$.
3. Compute mode matrices $\mathbf{U}_n$, for $n = 1, 2, \ldots, N-1$, by decomposing $\mathcal{R}$ using the N-mode SVD algorithm as follows: $\mathcal{R} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_{N-1} \mathbf{U}_{N-1}$.
4. Truncate each mode matrix to one column, thus obtaining the initial coefficient vectors $\mathbf{U}^0_1, \mathbf{U}^0_2, \ldots, \mathbf{U}^0_{N-1}$.
5. *Alternating Least Squares*: Iterate for $k = 1, 2, \ldots$, until $\|\mathcal{R} - \mathcal{Z}^k \times_1 \mathbf{U}^k_1 \ldots \times_n \mathbf{U}^k_n \ldots \mathbf{U}^k_{N-1}\| \leq \epsilon$:
   (a) For $n = 1, 2, \ldots, N-1$:
      i. Set $\tilde{\mathcal{U}}^k_n = \mathcal{R} \times_1 \mathbf{U}^{k}_1{}^T \ldots \times_{n-1} \mathbf{U}^k_{n-1}{}^T \times_{n+1} \mathbf{U}^{k-1}_{n+1}{}^T \ldots \times_{N-1} \mathbf{U}^{k-1}_{N-1}{}^T$.
      ii. Mode-n flatten tensor $\tilde{\mathcal{U}}^k_n$ to obtain $\tilde{\mathbf{U}}^k_n$.
      iii. Set $\mathbf{U}^k_n$ to the first column of the left singular matrix of the SVD of $\tilde{\mathbf{U}}^k_n$.
   (b) Compute core tensor $\mathcal{Z}^k = \tilde{\mathcal{U}}^k_{N-1} \times_{N-1} \mathbf{U}^k_{N-1}$.
6. Letting the converged $\mathbf{U}^k_n$ be denoted by the coefficient vectors $\mathbf{c}_1, \ldots, \mathbf{c}_{N-1}$ representing $\mathbf{d}$, then the projection of $\mathbf{d}^T$ is $\hat{\mathbf{d}}^T = \mathcal{T} \times_1 \mathbf{c}^T_1 \times_2 \mathbf{c}^T_2 \ldots \times_{N-1} \mathbf{c}^T_{N-1}$.

Figure 7.

4. Application to Face Recognition

Applying the multilinear projection algorithm to the specific face recognition scenario described earlier, in Step 4 of the algorithm the truncated $\mathbf{U}_n$ where $1 \leq n \leq 3$ (i.e., $\mathbf{U}_{\text{people}}$, $\mathbf{U}_{\text{views}}$, and $\mathbf{U}_{\text{illums}}$), become the initial approximations to the illumination $\mathbf{c}_l$, view $\mathbf{c}_v$, and person $\mathbf{c}_p$ coefficient vectors. These are the modes inherent to $\mathbf{d}$.

Projecting the unlabeled test image $\mathbf{d}$ onto the pixel mode of $\mathcal{T}$ yields the image response tensor (Fig. 5(b)):

$$\mathcal{R} = \mathcal{P} \times_4 \mathbf{d}^T \approx \mathbf{c}_l \circ \mathbf{c}_v \circ \mathbf{c}_p.^{1} \tag{9}$$

The rank-$(1,1,1)$ structure of tensor $\mathcal{R}$ and the fact that $\mathbf{c}_l$, $\mathbf{c}_v$, and $\mathbf{c}_p$ are unit vectors enables us to compute these three coefficient vectors via a tensor decomposition using the N-mode SVD algorithm. This is because the mode-n vectors of $\mathcal{R}$ are multiples of their corresponding coefficient vectors ($\mathbf{c}_l$, $\mathbf{c}_v$, $\mathbf{c}_p$) (cf. the boxed rows/columns in Fig. 5(b)). Thus, flattening $\mathcal{R}$ in each mode yields rank-1 matrices, enabling the N-mode SVD to compute the corresponding coefficient vector. The N-mode SVD thus maps $\mathcal{R}$ into N different mode spaces that explicitly account for the contribution of each mode—illumination, viewpoint, and person.

In particular, note that the person coefficient vector $\mathbf{c}_p$ is the left singular matrix of the SVD of $\mathbf{R}_{\text{(people)}}$. To recognize the person in the unknown test image $\mathbf{d}^T$, we can employ a normalized nearest neighbor classification scheme by computing normalized scalar products between $\mathbf{c}^T_p$ and each of the row vectors of the people mode matrix $\mathbf{U}_{\text{people}}$.

We have applied our multilinear projection algorithm to face recognition in experiments with the $16,875$ images captured from the University of Freiberg 3D Morphable Faces Database (Fig. 1). Using multilinear ICA (MICA) bases [11] that were learned from a training ensemble of 2,700 images, multilinear projection and nearest neighbor classification, we obtained recognition rates slightly greater than 98% for test subjects whose faces were imaged in previously unseen viewpoints and illuminations.

[1]From Lemma 1, $\mathcal{P} \times_4 \mathbf{d}^T \approx \mathcal{I}_4 \times_4 (\mathbf{c}^T_p \otimes \mathbf{c}^T_v \otimes \mathbf{c}^T_l)$, where $\mathcal{I}_4$ is the re-tensorized identity matrix $\mathbf{I}_{\text{(pixels)}} \approx \mathbf{T}^T_{\text{(pixels)}} \mathbf{T}^{+T}_{\text{(pixels)}} = \mathbf{T}^T_{\text{(pixels)}} \mathbf{P}_{\text{(pixels)}}$.

Acknowledgements

The 3D Morphable faces database was obtained courtesy of Prof. Sudeep Sarkar of the University of South Florida (USF) as part of the USF HumanID 3D Database.

A. Multilinear Algebraic Fundamentals

A *tensor*, or n-way array, is a higher-order generalization of a scalar (zeroth-order tensor), a vector (first-order tensor), and a matrix (second-order tensor). We denote scalars by lowercase italic letters $(a, b, \ldots)$, vectors by bold lowercase letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices by bold uppercase letters $(\mathbf{A}, \mathbf{B}, \ldots)$, and higher-order tensors by calligraphic uppercase letters $(\mathcal{A}, \mathcal{B}, \ldots)$.

A.1. Basic Definitions

Definition 5 (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. *An element of* $\mathcal{A}$ *is denoted as* $\mathcal{A}_{i_1 \ldots i_n \ldots i_N}$ *or* $a_{i_1 \ldots i_n \ldots i_N}$, *where* $1 \leq i_n \leq I_n$.

Definition 6 (Mode-n Vectors) *The mode-n vectors of an* N^{th}*−order tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *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 (a.k.a. fibers) 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 result from *flattening* the tensor $\mathcal{A}$ (Fig. 8).

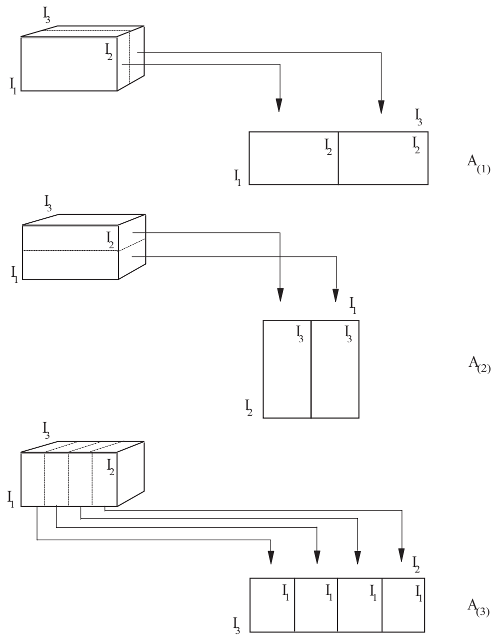


Figure 8. Flattening a third-order tensor. The tensor can be flattened in 3 ways to obtain matrices comprising its 1-mode, 2-mode, and 3-mode vectors.

Definition 7 (Mode-n Rank) *The mode-n rank of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, denoted R_n, is defined as the dimension of the vector space generated by the mode-n vectors:*

$$R_n = \mathrm{rank}_n(\mathcal{A}) = \mathrm{rank}(\mathbf{A}_{(n)}). \tag{10}$$

A generalization of the product of two matrices is the product of a tensor and a matrix.

Definition 8 (Mode-n Product, $\times_n$) *The mode-n product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n \times \ldots \times I_N}$ and a matrix $\mathbf{M} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{M}$, is a tensor of dimensionality $\mathbb{R}^{I_1 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N}$ whose entries are computed by $(\mathcal{A} \times_n \mathbf{M})_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} m_{j_n i_n}$.*

The mode-n product can be expressed in tensor notation as

$$\mathcal{B} = \mathcal{A} \times_n \mathbf{M}, \tag{11}$$

or in terms of flattened matrices as

$$\mathbf{B}_{(n)} = \mathbf{M}\mathbf{A}_{(n)}. \tag{12}$$

A matrix representation of the mode-n product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$ and a set of N matrices, $\mathbf{F}_n \in \mathbb{R}^{J_n \times I_n}$ can be obtained as follows:

$$\begin{aligned} \mathcal{B} &= \mathcal{A} \times_1 \mathbf{F}_1 \ldots \times_n \mathbf{F}_n \ldots \times_N \mathbf{F}_N \qquad (13) \\ \mathbf{B}_{(n)} &= \mathbf{F}_n \mathbf{A}_{(n)} (\mathbf{F}_{n-1} \otimes \ldots \mathbf{F}_1 \otimes \mathbf{F}_N \otimes \ldots \mathbf{F}_{n+1})^T, \end{aligned}$$

where $\otimes$ denotes the matrix Kronecker product. Given a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$ and two matrices, $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$ and $\mathbf{V} \in \mathbb{R}^{K_n \times J_n}$, then

$$\mathcal{A} \times_n (\mathbf{V}\mathbf{U}) = (\mathcal{A} \times_n \mathbf{U}) \times_n \mathbf{V}. \tag{14}$$

The *Frobenius norm* of a tensor $\mathcal{A}$ is given by $\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$, where $\langle \cdot, \cdot \rangle$ denotes the scalar product.

A.2. Tensor Decompositions

There are two types of higher-order tensor decompositions, but neither has all the nice properties of the matrix SVD. The tensor rank-R decomposition is a natural generalization of the matrix rank-R decomposition, but it does not compute the orthonormal subspace associated with each mode. The second, known as the rank-$(R_1, R_2, \ldots, R_N)$ decomposition, does not reveal the rank of the tensor, but it naturally generalizes the orthonormal subspaces corresponding to the left/right singular matrices computed by the matrix SVD [7, 5, 4, 3].

Theorem 1 (Rank-$(R_1, R_2, \ldots, R_N)$ Decomposition) *Let $\mathcal{A}$ be a $I_1 \times I_2 \times \times I_n \ldots I_N$ tensor where $1 \leq n \leq N$. Every such tensor can be decomposed as follows:*

$$\mathcal{A} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_n \mathbf{U}_n \ldots \times_N \mathbf{U}_N \tag{15}$$

where $\mathbf{U}_n$ are orthonormal mode-n matrices of dimensionality $I_n \times I_n$, where $\mathbf{u}_n^{(i_n)}$ is the i_n^{th} column in $\mathbf{U}_n$ for $1 \leq n \leq N$, and where $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_n \ldots \times I_N}$. The subtensors $\mathcal{Z}_{i_n=a}$ and $\mathcal{Z}_{i_n=b}$ obtained by fixing the n^{th} index to a and b are orthogonal for all values of n, a, and b when $a \neq b$. The $\|\mathcal{Z}_{j_n=a}\| = \sigma_a^{(n)}$ is the a^{th} mode-n singular value of $\mathcal{A}$ and the a^{th} column vector of $\mathbf{U}_n$, such that $\|\mathcal{Z}_{i_n=1}\| \geq \|\mathcal{Z}_{i_n=2}\| \geq \ldots \|\mathcal{Z}_{i_n=I_n}\| \geq 0$.

This is illustrated in Fig. 9 for the case $N = 3$. 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 $\mathbf{A}_{(n)}$ resulting from the *mode-n flattening* of $\mathcal{A}$ (Fig. 8).

For matrices, the rank-R decomposition and the multilinear rank-(R_1, R_2) decomposition (with $R_1 = R_2 = R$ necessarily) are equivalent; hence, there is no need to distinguish between them. They are both computed by the matrix SVD (along with the orthonormal subspaces associated with the column space $\mathbf{U}_1$ and row space $\mathbf{U}_2$). Unfortunately, however, there does not exist a single higher order SVD for tensors that has all the nice properties of the matrix SVD.

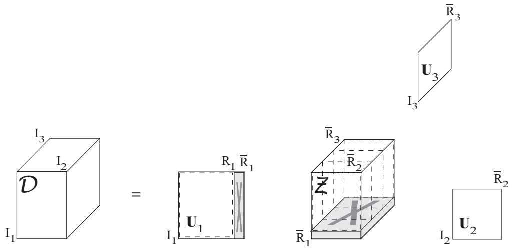


Figure 9. The rank-(R_1, R_2, R_3) decomposition expresses the data tensor $\mathcal{D}$ as the product of a core tensor $\mathcal{Z}$ and N orthogonal mode matrices $\mathbf{U}_1 \ldots \mathbf{U}_N$; for the $N = 3$ case illustrated here, $\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$. Deletion of the last mode-1 eigenvector of $\mathbf{U}_1$ incurs an error in the approximation equal to $\sigma^2_{I_1}$, 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 N-mode SVD algorithm

1. For $n = 1, \ldots, N$, compute matrix $\mathbf{U}_n$ in (15) 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. Solve for the core tensor:
$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \ldots \times_n \mathbf{U}_n^T \ldots \times_N \mathbf{U}_N^T$.

Figure 10.

A.3. N-Mode SVD and Dimensionality Reduction

The rank-$(R_1, R_2, \ldots R_N)$ decomposition (15) may be computed using the N*-mode SVD* algorithm (Fig. 10), a multilinear extension of the conventional matrix SVD [7, 5, 2].

There is no trivial multilinear counterpart to dimensionality reduction in the linear case. Truncation of the mode matrices resulting from the N-mode SVD algorithm may yield a good reduced-dimensionality approximation $\hat{\mathcal{D}}$ (Fig. 9), but it is generally not optimal.

A locally optimal dimensionality reduction scheme for tensors is to compute a *best rank-*$(R_1, R_2, \ldots, R_N)$ *approximation*[2] $\hat{\mathcal{D}} = \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \times_2 \hat{\mathbf{U}}_2 \ldots \times_N \hat{\mathbf{U}}_N$, with orthonormal $I_n \times R_n$ mode matrices $\hat{\mathbf{U}}_n$, for $n = 1, 2, \ldots, N$, which minimizes the error function

$$e = \|\mathcal{D} - \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \ldots \times_N \hat{\mathbf{U}}_N\| + \sum_{i=1}^{N} \mathbf{\Lambda}_i \|\hat{\mathbf{U}}_i^T \hat{\mathbf{U}}_i - \mathbf{I}\|, \tag{16}$$

where the $\mathbf{\Lambda}_i$ are Lagrange multiplier matrices. The (local) minimization can be accomplished in an iterative manner, optimizing 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. The algorithm is given in [10, 2, 5].

[2] This best rank-$(R_1, R_2, \ldots, R_N)$ problem should not be confused with the classical "best rank-R" problem for tensors [6]: 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.

References

[1] V. Blanz and T. Vetter. Morphable model for the synthesis of 3D faces. In *SIGGRAPH 99 Conference Proceedings*, pages 187–194. ACM SIGGRAPH, 1999. 2

[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. 8

[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. 7

[4] A. Kapteyn, H. Neudecker, and T. Wansbeek. An approach to n-mode component analysis. *Psychometrika*, 51(2):269–275, June 1986. 7

[5] P. Kroonenberg and J. de Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45:69–97, 1980. 7, 8

[6] 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. 8

[7] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966. 7, 8

[8] M. Vasilescu and D. Terzopoulos. Multilinear analysis for facial image recognition. In *Proc. Int. Conf. on Pattern Recognition (ICPR 2002)*, volume 2, pages 511–514, Quebec City, August 2002. 1, 3

[9] M. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: TensorFaces. In *Proc. European Conf. on Computer Vision (ECCV 2002)*, pages 447–460, Copenhagen, Denmark, May 2002. 1, 2

[10] M. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2003)*, volume II, pages 93–99, Madison, WI, 2003. 8

[11] M. Vasilescu and D. Terzopoulos. Multilinear independent components analysis. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, volume I, pages 547–553, San Diego, CA, 2005. 6