Multilinear Projection For Face Recognition Via Canonical Decomposition

M. Alex O. Vasilescu
University of California, Los Angeles
maov@cs.ucla.edu

Abstract

This paper introduces a new multilinear projection algorithm for appearance-based recognition in a tensor framework. The multilinear projection simultaneously maps an unlabeled image from the pixel space into multiple causal factors underlying image formation, including illumination, imaging, and scene structure. For facial recognition, the most relevant aspect of scene structure is the specific person whose face has been imaged. Our new multilinear projection algorithm, which is based on the canonical decomposition of tensors, is superior to our previously proposed multilinear projection algorithm that is based on an M-mode SVD. To develop our new algorithm, we extend and formalize the definition of the mode-m product, the mode-m identity tensor, and the mode-m pseudo-inverse tensor. We demonstrate our multilinear projection in the context of facial image recognition and compare its results in simultaneously inferring the identity, view, illumination, etc., coefficient vectors of an unlabeled test image against those obtained using multilinear projection based on the M-mode SVD, as well as against the results obtained using a set of multiple linear projections. Finally, we present a strategy for developing a practical biometric system that can enroll an uncooperative subject using a one or more images and then recognize that subject in unconstrained test images.

1. Introduction

The goal of many statistical data analysis problems, among them those arising in the domains of computer vision and machine learning, is to find a suitable representation of multivariate data that facilitates the analysis, visualization, compression, approximation, recognition and/or interpretation of the observed data. This is often done by applying a suitable transformation to the space in which the observational data reside.

Representations that are derived through *linear transformations* of the original observed data have traditionally been preferred due to their conceptual and computational simplicity. Principal components analysis (PCA), one of the most valuable results from applied linear algebra, is used broadly in many forms of data analysis, including the analysis of facial image data, because it is a simple, non-parametric method for extracting relevant information from complex data sets. PCA provides a dimensionality reduction methodology that aspires to reveal a meaningful causal factor underlying data formation. Whether derived through second-order or higher-order statistical considerations, however, linear transformations, such as PCA [12, 14, 11] and independent components analysis (ICA) [2, 8], are limited in their ability to support the analysis of *multifactor* data formation, since linear transformations are best suited to modeling observational data that results from single-factor linear variation or from the linear combination of multiple sources.

Vasilescu and Terzopoulos [16, 15] have argued that since natural images result from the interaction between multiple causal factors related to the imaging process, the illumination, and the scene geometry, a principled mathematical approach to disentangling and explicitly representing these causal factors essential to image formation is through numerical multilinear algebra, the algebra of higher-order tensors. The *multilinear transformations* that are involved in this approach lead to generative models that explicitly capture how the observed data are influenced by multiple causal factors. In general, these causal factors may be fundamental physical, behavioral, or biological processes that cause patterns of variation in the observational data, which comprise a set of measurements or response variables that are affected by the causal factors. Facial images in particular are the result of specific facial geometry (person, facial expression, etc.), the pose of the head relative to the camera, the lighting conditions, and the type of camera employed. A multilinear transformation computes a unique representation for each causal factor and an image is represented as a collection of causal factor representations.

In this paper, we develop a multilinear projection method for appearance-based recognition through canonical decomposition, the so-called the CANDECOMP/PARAFAC (CP) decomposition [5, 7]. Given an unlabeled facial image, the multilinear projection maps it from the measurement, pixel space to the multiple causal factor spaces, thus simultaneously inferring the identity of the person, the viewpoint of the camera, the illumination conditions, etc., coefficient vectors of the test image. Due to differences in the optimization constraints, our multilinear projection based on CP decomposition (MP-CP) is superior to our earlier multilinear projection based on multilinear PCA (MP-MPCA) [19]. The earlier approach must first compute an intermediate rank-$(R_1, \dots, R_M)$ decomposition that optimizes for orthonormal mode matrices and that must be dimensionally reduced [10, 9, 6, 17]. This intermediate step, which requires the imposition of an orthonormality constraint, can introduce bias, resulting in a suboptimal solution. Our new approach employs a modified CP algorithm, which computes the best fitting rank-1 decomposition without an intermediate step that biases the solution. To develop our new multilinear projection algorithm, we extend and formalize the definition of the mode-m product, the mode-m identity tensor, and the mode-m pseudo-inverse tensor. There have been two previous attempts at such a generalization [1, 19], which were informal and/or incomplete.

We demonstrate multilinear projection via canonical decompostion in the context of facial image recognition and compare its results against those associated with multilinear projection via multilinear PCA [19] as well as against the results associated with performing a set of multiple linear projections [16].

Finally, we discuss a strategy for developing a practical biometric system that can enroll an uncooperative subject from a small number of surveillance images, by representing his/her facial image(s) relative to the statistics encoded in a multilinear model that is trained on a set of cooperative subjects, and that can then recognize the uncooperative subject in unconstrained test images.

The remainder of this paper is organized as follows: In Section 2, we review our multilinear image analysis approach and, in particular, the TensorFaces method. Section 3 develops our multilinear projection algorithms. Section 4 applies the multilinear projection to face recognition and discusses the development of a realistic face recognition system where uncooperative subjects must be enrolled from only one or a few images and then recognized in unconstrained test images.

2. Multilinear Image Representation

The multilinear analysis framework for appearance-based image representation offers a potent mathematical approach to analyzing the multifactor structure of image ensembles and for addressing the fundamental yet difficult

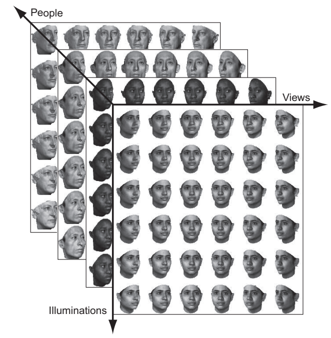


Figure 1. A facial image dataset. (a) 3D scans of 75 subjects, recorded using a Cyberware$^{\mathrm{TM}}$ 3030PS laser scanner as part of the University of Freiburg 3D morphable faces database [3]. A portion of the 4^{th}-order data tensor $\boldsymbol{\mathcal{D}}$ of the image ensemble used for training. Only 4 of the 75 people are shown.

problem of disentangling the causal factors.[1]

Multilinear transformations lead to generative models that explicitly capture how the observed data are influenced by multiple underlying causal factors. A multilinear transformation is a *nonlinear* function or mapping from not just one, but a *set* of M domain vector spaces $\mathbb{R}^{m_i}$, $1 \leq i \leq M$, to a range vector space $\mathbb{R}^n$:

$$\boldsymbol{\mathcal{T}} : \{\mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_M}\} \mapsto \mathbb{R}^n. \qquad (1)$$

Given the data tensor $\boldsymbol{\mathcal{D}}$ of labeled, vectorized training images $\mathbf{d}_{pvle}$, where the subscripts denote person p, view v, illumination l, and expression e labels, we can apply the MPCA algorithm [17, 15] to compute causal mode matrices $\mathbf{U}_{\mathrm{P}}$, $\mathbf{U}_{\mathrm{V}}$, $\mathbf{U}_{\mathrm{L}}$, and $\mathbf{U}_{\mathrm{E}}$ as well as the TensorFaces basis $\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{D}} \times_{\mathrm{P}} \mathbf{U}_{\mathrm{P}}^{\mathrm{T}} \times_{\mathrm{V}} \mathbf{U}_{\mathrm{V}}^{\mathrm{T}} \times_{\mathrm{L}} \mathbf{U}_{\mathrm{L}}^{\mathrm{T}} \times_{\mathrm{E}} \mathbf{U}_{\mathrm{E}}^{\mathrm{T}}$ that governs the interaction between them (Figure 2(a)). Then the method represents an image $\mathbf{d}_{pvl}$ by the relevant set of person, view, and illumination coefficient vectors as follows:

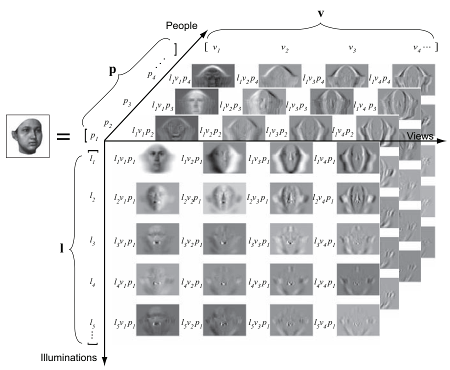
$$\mathbf{d}_{pvle} = \boldsymbol{\mathcal{T}} \times_{\mathrm{P}} \mathbf{p}_p^{\mathrm{T}} \times_{\mathrm{V}} \mathbf{v}_v^{\mathrm{T}} \times_{\mathrm{L}} \mathbf{l}_l^{\mathrm{T}} \times_{\mathrm{E}} \mathbf{e}_e^{\mathrm{T}}. \qquad (2)$$

Alternatively, we can apply the MICA algorithm [18, 20], which employs higher-order statistics to compute an MICA basis tensor $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{D}} \times_{\mathrm{P}} \mathbf{C}_{\mathrm{P}}^{+} \times_{\mathrm{V}} \mathbf{C}_{\mathrm{V}}^{+} \times_{\mathrm{L}} \mathbf{C}_{\mathrm{L}}^{+} \times_{\mathrm{E}} \mathbf{C}_{\mathrm{E}}^{+}$. Analogous to the MPCA case, an image can be represented with respect to the MICA basis, as follows:

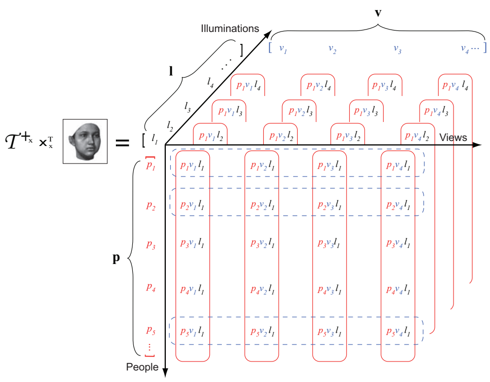
$$\mathbf{d}_{pvle} = \boldsymbol{\mathcal{M}} \times_{\mathrm{P}} \mathbf{p}_p^{\mathrm{T}} \times_{\mathrm{V}} \mathbf{v}_v^{\mathrm{T}} \times_{\mathrm{L}} \mathbf{l}_l^{\mathrm{T}} \times_{\mathrm{E}} \mathbf{e}_e^{\mathrm{T}}. \qquad (3)$$

By comparison to linear approaches where an individual has a representation for every image in which they appear, in the multilinear approaches discussed above, such

[1] An observation comprises a set of measurements or *response variables* whose values are influenced by multiple underlying *causal factors*. The causal factors are not directly measurable, but they are of interest, and the variables extracted by data analysis in order to represent them are known as *explanatory variables*. For example, an image is an observation whose measurements are pixels, the values of which vary with changes in the causal factors—scene structure, illumination, view, etc.



(a)



(b)

Figure 2. (a) MPCA image representation $\mathbf{d} = \mathcal{T} \times_{\mathrm{P}} \mathbf{p}^{\mathrm{T}} \times_{\mathrm{V}} \mathbf{v}^{\mathrm{T}} \times_{\mathrm{L}} \mathbf{l}^{\mathrm{T}}$. (b) Given an unlabeled test image $\mathbf{d}$, the associated coefficient vectors $\mathbf{p}$, $\mathbf{v}$, $\mathbf{l}$ are estimated by decomposing the response tensor $\mathcal{R} = \mathcal{T}^{+_{\mathrm{x}}} \times_{\mathrm{x}}^{\mathrm{T}} \mathbf{d}$ using a multilinear projection algorithm.

as MPCA and MICA, an individual has the same representation regardless of viewpoint, illumination, expression, etc. This is an important advantage of multilinear models over linear ones on which our recognition system capitalizes for superior results.

3. Multilinear Projection

Given an unlabeled test image (probe) $\mathbf{d}$ and $\mathcal{T}$ or $\mathcal{M}$, we must determine the unknown coefficient vectors, $\mathbf{p}_p$, $\mathbf{v}_v$, $\mathbf{l}_l$, and $\mathbf{e}_e$ in order to recognize the person, view, illumination, and expression associated with the test image. Solving for these vectors in (2) or (3) will, in principle, require the computation of a pseudo-inverse of tensor $\mathcal{T}$ or $\mathcal{M}$. In analogy with matrix algebra, this raises the following questions: How does one "invert" a tensor? When one "multiplies" a tensor with its "inverse tensor", what should be the resulting "identity tensor"? We will next show that an M^{th}-order tensor has M *pseudo-inverse tensors*, one with respect to each mode, and that there are M *identity tensors*, one per mode, whose structure is not diagonal with ones along the main diagonal.

3.1. Identity and Pseudo-Inverse Tensors

First, we generalize the definition of the mode-m product[2] of a tensor and a matrix to two tensors:[3]

Definition 3.1 (Generalized Mode-m Product) *The generalized mode-m product between two tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m \times \cdots \times I_M}$ and $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_m \times \cdots \times J_M}$ is expressed as follows:*

1. $\mathcal{A} \times_m \mathcal{B} = \mathcal{C} \in \mathbb{R}^{I_1 \times \cdots \times I_{m-1} \times J_m \times I_{m+1} \times \cdots \times I_M}$, *where* $I_m = J_1 \ldots J_{m-1} J_{m+1} \ldots J_M$, *can be expressed in matrix form as* $\mathbf{C}_{[m]} = \mathbf{B}_{[m]} \mathbf{A}_{[m]}$.

2. $\mathcal{A} \times_m^{\mathrm{T}} \mathcal{B} = \mathcal{C} \in \mathbb{R}^{I_1 \times \cdots \times I_{m-1} \times K_m \times I_{m+1} \times \cdots \times I_M}$, *where* $K_m = J_1 \ldots J_{m-1} J_{m+1} \ldots J_M$ *and* $I_m = J_m$, *can be expressed in matrix form as* $\mathbf{C}_{[m]} = \mathbf{B}_{[m]}^{\mathrm{T}} \mathbf{A}_{[m]}$.

3. $\mathcal{A} \;{}^{\mathrm{T}}\!\!\times_m \mathcal{B} = \mathcal{C} \in \mathbb{R}^{I_m \times J_m}$, *where* $I_1 \ldots I_{m-1} I_{m+1} \ldots I_M = J_1 \ldots J_{m-1} J_{m+1} \ldots J_M$, *can be expressed in matrix form as* $\mathbf{C}_{[m]}^{\mathrm{T}} = \mathbf{B}_{[m]} \mathbf{A}_{[m]}^{\mathrm{T}}$.

4. $\mathcal{A} \;{}^{\mathrm{T}}\!\!\times_m^{\mathrm{T}} \mathcal{B} = \mathcal{C} \in \mathbb{R}^{J_1 \times \cdots \times J_{m-1} \times I_m \times J_{m+1} \times \cdots \times J_M}$, *where* $J_m = I_1 \ldots I_{m-1} I_{m+1} \ldots I_M$, *can be expressed in matrix form as* $\mathbf{C}_{[m]}^{\mathrm{T}} = \mathbf{B}_{[m]}^{\mathrm{T}} \mathbf{A}_{[m]}^{\mathrm{T}}$.

With the above generalization, we define the mode-m identity tensor as follows:

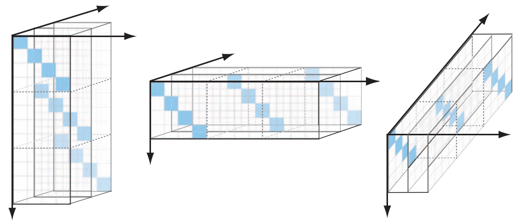
Definition 3.2 (Mode-m Identity Tensor) *Tensor $\mathcal{I}_m$ is a mode-m multiplicative identity tensor if and only if $\mathcal{I}_m \times_m \mathcal{A} = \mathcal{A}$, where $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_m \times \cdots \times I_M}$ and $\mathcal{I}_m \in \mathbb{R}^{I_1 \times \cdots \times I_{m-1} \times J_m \times I_{m+1} \times \cdots \times I_M}$, where $J_m = I_1 I_2 \ldots I_{m-1} I_{m+1} \ldots I_M$.*

While a mode-wise identity tensor might seem to be a construct peculiar to multilinear algebra, one should recall that in linear algebra there exist left and right identity matrices for every rectangular 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. By contrast, the mode-m identity tensors are not diagonal tensors. Figure 3 illustrates the structure of the three identity tensors of order 3.

The mode-m identity tensor can be used to tensorize a matrix or a row vector via a mode-m product. It does not change the values of the matrix/vector but simply reorders its elements. In particular, it can re-tensorize a matrix obtained by matrixizing a tensor; i.e., given

[2] $[\mathcal{A} \times_m \mathbf{B}]_{i_1 \ldots i_{m-1} j_m i_{m+1} \ldots i_M} = \sum_{i_m} a_{i_1 \ldots i_{m-1} i_m i_{m+1} \ldots i_M} b_{j_m i_m}$.

[3] Note that there have been two previous attempts at such a generalization [1, 19], which were informal and/or incomplete.



(a) (b) (c)

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

a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_m \times \cdots \times I_M}$ and an identity tensor, $\mathcal{I}_m \in \mathbb{R}^{I_1 \times \cdots \times I_{m-1} \times J_m \times I_{m+1} \times \cdots \times I_M}$ with $J_m = I_1 I_2 \ldots I_{m-1} I_{m+1} \ldots I_M$, then

$$\mathcal{I}_m \times_m \mathbf{A}_{[m]} = \mathcal{A}. \quad (4)$$

We now define a mode-m pseudo-inverse tensor that generalizes the pseudo-inverse matrix from linear algebra.

Definition 3.3 (Mode-m Pseudo-Inverse Tensor) *The mode-m pseudoinverse tensor $\mathcal{A}^{+_m}$ of tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ satisfies:*

1. $(\mathcal{A} \times^{\mathrm{T}}_m \mathcal{A}^{+_m}) \times_m \mathcal{A} = \mathcal{A}$
2. $(\mathcal{A}^{+_m}{}^{\mathrm{T}} \times_m \mathcal{A})\ {}^{\mathrm{T}}\times^{\mathrm{T}}_m \mathcal{A}^{+_m} = \mathcal{A}^{+_m}$

The mode-m pseudoinverse tensor $\mathcal{A}^{+_m}$ of $\mathcal{A}$ is the tensorized version of $\mathbf{A}^{+\mathrm{T}}_{[m]}$; i.e., $\mathbf{A}^{+\mathrm{T}}_{[m]} = [\mathcal{A}^{+_m}]_{[m]}$.

3.2. Multilinear Projection Algorithms

To determine the coefficient vectors that represent an unlabeled observation (image), which is a point (vector) in the (pixel) measurement space, we must map the observation from the measurement space to the causal factor spaces (Figure 2). Given an unlabeled test (probe) image $\mathbf{d}$ and a learned TensorFaces model $\mathcal{T}$, the image is represented as follows:

$$\mathbf{d} \;=\; \mathcal{T} \times_{\mathrm{P}} \mathbf{r}^{\mathrm{T}}_{\mathrm{P}} \times_{\mathrm{V}} \mathbf{r}^{\mathrm{T}}_{\mathrm{V}} \times_{\mathrm{L}} \mathbf{r}^{\mathrm{T}}_{\mathrm{L}} \times_{\mathrm{E}} \mathbf{r}^{\mathrm{T}}_{\mathrm{E}} + \boldsymbol{\rho}, \quad (5)$$

where $\mathbf{d} \in \mathbb{R}^{I_{\mathrm{x}} \times 1 \times \cdots \times 1}$ and $\boldsymbol{\rho}$ is a residual vector that lies outside the range of the multilinear generative model. Thus, $\boldsymbol{\rho}$ is orthogonal to the TensorFaces basis $\mathcal{T}$ and $\boldsymbol{\rho} = \mathbf{0}$ when $\mathbf{d}$ lies in the subspace spanned by the basis. Thus, to compute the coefficient vector representations, $\mathbf{r}_{\mathrm{P}}$, $\mathbf{r}_{\mathrm{V}}$, $\mathbf{r}_{\mathrm{L}}$, and $\mathbf{r}_{\mathrm{E}}$, needed to recognize the person, view, illumination, and expression depicted in test image $\mathbf{d}$, we must pseudo-invert $\mathcal{T}$ with respect to the (pixel) measurement mode—i.e., compute $\mathcal{T}^{+_{\mathrm{x}}}$.

In view of the above considerations, we will now derive a general multilinear projection algorithm. To this end, we will temporarily revert back to numbering modes for full generality and assume that mode 1 is the measurement (e.g., pixel) mode. The general, M-mode form of (5) is

$$\mathbf{d} = \mathcal{T} \times_2 \mathbf{r}^{\mathrm{T}}_2 \cdots \times_M \mathbf{r}^{\mathrm{T}}_M + \boldsymbol{\rho}. \quad (6)$$

Performing a mode-1 product of both sides of this equation by the mode-1 pseudo-inverse of the TensorFaces bases, we obtain a *response tensor*

$$\begin{aligned}
\mathcal{R} &= \mathcal{T}^{+_1} \times^{\mathrm{T}}_1 \mathbf{d} && (7)\\
&= \mathcal{T}^{+_1} \times^{\mathrm{T}}_1 (\mathcal{T} \times_2 \mathbf{r}^{\mathrm{T}}_2 \cdots \times_m \mathbf{r}^{\mathrm{T}}_m \cdots \times_M \mathbf{r}^{\mathrm{T}}_M + \boldsymbol{\rho}) && (8)\\
&= (\mathcal{T}^{+_1} \times^{\mathrm{T}}_1 \mathcal{T}) \times_2 \mathbf{r}^{\mathrm{T}}_2 \cdots \times_m \mathbf{r}^{\mathrm{T}}_m \cdots \times_M \mathbf{r}^{\mathrm{T}}_M + (\mathcal{T}^{+_1} \times^{\mathrm{T}}_1 \boldsymbol{\rho}) &&\\
&\simeq \mathcal{I}_1 \times_2 \mathbf{r}^{\mathrm{T}}_2 \cdots \times_m \mathbf{r}^{\mathrm{T}}_m \cdots \times_M \mathbf{c}^{\mathrm{T}}_M + \mathbf{0} && (9)\\
&= \mathit{1} \times_2 \mathbf{r}_2 \cdots \times_m \mathbf{r}_m \cdots \times_M \mathbf{r}_M \quad \text{Rank-}(1,\ldots,1) && (10)\\
&= \mathbf{r}_2 \circ \mathbf{r}_3 \cdots \circ \mathbf{r}_M, \quad \text{Rank-1} && (11)
\end{aligned}$$

where $\mathbf{d} \in \mathbb{R}^{I_1 \times 1 \times \cdots \times 1}$ and $\mathbf{d}_{[1]} = \mathbf{d}^{\mathrm{T}}$, where $(\mathcal{T}^{+_1} \times^{\mathrm{T}}_1 \mathcal{T}) \simeq \mathcal{I}_1$ when $I_1 < I_2 I_3 \ldots I_M$, otherwise $(\mathcal{T}^{+_1} \times^{\mathrm{T}}_1 \mathcal{T}) = \mathcal{I}_1$, and where $\mathcal{I}_1 \in \mathbb{R}^{(I_2 I_3 \ldots I_M) \times I_2 \times \cdots \times I_M}$. The three equalities (9)-(11) can be derived using the definition of the mode-m product [4] and the vec-Kronecker property [5]. The rank-$(1,\ldots,1)$/rank-1 structure of the response tensor $\mathcal{R}$ is amenable to a tensor decomposition using the MPCA algorithm or a modified CP algorithm in order to determine the $\mathbf{r}_m$ coefficient vector representations.

The multilinear projection algorithm can employ a modified CANDECOMP/PARAFAC (CP) algorithm to compute the best fitting rank-1 term for the response tensor. Like the MPCA algorithm, the CP algorithm takes advantage of the structure of $\mathcal{R}$. The mode-m vectors of $\mathcal{R}$ are multiples of $\mathbf{r}_m$ and the scalar multiples have a well defined structure that the CP algorithm exploits. Given the structure of $\mathcal{R}$, the outer product of coefficient vectors $\mathbf{r}_m$ may be expressed in matrix form as:

$$\begin{aligned}
\mathbf{R}_{[m]} &\simeq (\mathbf{r}_2 \circ \mathbf{r}_3 \cdots \circ \mathbf{r}_M)_{[m]} && (12)\\
&= \mathbf{r}_m(\mathbf{r}_M \otimes \cdots \otimes \mathbf{r}_{m+1} \otimes \mathbf{r}_{m-1} \otimes \cdots \otimes \mathbf{r}_2)^{\mathrm{T}} &&\\
&= \mathbf{r}_m(\mathbf{r}_M \odot \cdots \odot \mathbf{r}_{m+1} \odot \mathbf{r}_{m-1} \odot \cdots \odot \mathbf{r}_2)^{\mathrm{T}} &&\\
&= \mathbf{r}_m \mathbf{y}^{\mathrm{T}}_m, && (13)
\end{aligned}$$

where $\circ$ is the outer-product, $\otimes$ is the Kronecker product[6], $\odot$ is the Khatri-Rao product [7], and $\mathbf{y}_m = (\mathbf{r}_M \odot \cdots \odot \mathbf{r}_{m+1} \odot \mathbf{r}_{m-1} \odot \cdots \odot \mathbf{r}_2)$. Therefore, each coefficient vector representation is given by

$$\mathbf{r}_m = \mathbf{R}_{[m]}\mathbf{y}_m(\mathbf{y}^{\mathrm{T}}_m \mathbf{y}_m)^{-1} = \mathbf{R}_{[m]}\mathbf{y}_m / \|\mathbf{y}_m\|^2. \quad (14)$$

Given the form of $\mathbf{y}_m$, we can compute its norm efficiently using the relationship $(\mathbf{U} \odot \mathbf{V})^{\mathrm{T}}(\mathbf{U} \odot \mathbf{V}) = (\mathbf{U}^{\mathrm{T}}\mathbf{U}) \circledast (\mathbf{V}^{\mathrm{T}}\mathbf{V})$ between the Khatri-Rao and Hadamard products [8],

[4] The mode-m product $\mathcal{C} = \mathcal{A} \times_1 \mathbf{B}_1 \cdots \times_m \mathbf{B}_m \cdots \times_M \mathbf{B}_M$ can be expressed in matrix form as $\mathbf{C}_{[m]} = \mathbf{B}_m \mathbf{A}_{[m]} (\mathbf{B}_M \otimes \cdots \otimes \mathbf{B}_{m+1} \otimes \mathbf{B}_{m-1} \otimes \cdots \otimes \mathbf{B}_1)^{\mathrm{T}}$ where $\otimes$ is the Kronecker product.

[5] The vec-Kronecker property: $\mathrm{vec}(\mathbf{a} \circ \mathbf{b}) = \mathrm{vec}(\mathbf{a}\mathbf{b}^{\mathrm{T}}) = \mathbf{b} \otimes \mathbf{a}$

[6] The Kronecker product is $\mathbf{U} \otimes \mathbf{V} = \begin{bmatrix} u_{11}\mathbf{V} & \ldots & u_{1J}\mathbf{V} \\ \vdots & \ddots & \vdots \\ u_{I1}\mathbf{V} & \ldots & u_{IJ}\mathbf{V} \end{bmatrix}$.

[7] The Khatri-Rao product is a columnwise Kronecker product: $\mathbf{U} \odot \mathbf{V} = [(\mathbf{u}^{(1)} \otimes \mathbf{v}^{(1)}) \ldots (\mathbf{u}^{(l)} \otimes \mathbf{v}^{(l)}) \ldots (\mathbf{u}^{(L)} \otimes \mathbf{v}^{(L)})]$.

[8] The Hadamard product is an element-wise product defined as $[\mathbf{U} \circledast \mathbf{V}]_{ij} = u_{ij}v_{ij}$.

Algorithm 3.1 Multilinear projection (MP) algorithm with CP, rank-1 decomposition

Input a TensorFaces basis tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$,[a] where mode $m = 1$ is the measurement mode, and an unlabeled test observation (image) $\mathbf{d}$.

1. Compute the pseudo-inverse $\boldsymbol{\mathcal{T}}^{+_1}$ (in matrix form, $\mathbf{T}^{+\mathrm{T}}_{[1]}$).
2. Compute the response tensor $\boldsymbol{\mathcal{R}} := \boldsymbol{\mathcal{T}}^{+_1} \times_1 \mathbf{d}^{\mathrm{T}}$.
3. Initialize $\mathbf{y}_m$ to the column norms of $\mathbf{R}_{[m]}$.
4. For $m := 2, \ldots, M$, set $\mathbf{r}_m := \mathbf{R}_{[m]}\mathbf{y}_m / \|\mathbf{y}_m\|^2$.
5. *Local optimization via alternating least squares*:
   Iterate for $n := 1, \ldots, N$

   For $m := 2, \ldots, M$,
   $\mathbf{r}_m := \mathbf{R}_{[m]} \left(\mathbf{r}_M \odot \cdots \odot \mathbf{r}_{m-1} \odot \mathbf{r}_{m+1} \odot \cdots \odot \mathbf{r}_2\right) / \|\mathbf{r}_M\|^2 \ldots \|\mathbf{r}_{m+1}\|^2 \|\mathbf{r}_{m-1}\|^2 \ldots \|\mathbf{r}_2\|^2.$
   Set $\boldsymbol{\mathcal{Z}} := \boldsymbol{\mathcal{R}} \times_2 \mathbf{r}_2^{\mathrm{T}} \cdots \times_M \mathbf{r}_M^{\mathrm{T}}$.[b]

   until convergence.[c]

Output the converged causal factor representation vectors $\mathbf{r}_2 \ldots \mathbf{r}_M$.

[a] Or given a MICA basis tensor $\boldsymbol{\mathcal{M}}$.

[b] Note that $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{1^M}$ is a degenerate tensor of order M; i.e., a scalar.

[c] Note that N is a prespecified maximum number of iterations. A possible convergence criterion is to compute at each iteration the approximation error $e_n := \|\boldsymbol{\mathcal{D}} - \tilde{\boldsymbol{\mathcal{D}}}\|^2$ and test if $e_{n-1} - e_n \leq \epsilon$ for a sufficiently small tolerance ϵ.

as follows:

$$\begin{aligned}
\|\mathbf{y}_m\|^2 &= \mathbf{y}_m^{\mathrm{T}}\mathbf{y}_m && (15)\\
&= \mathbf{r}_M^{\mathrm{T}}\mathbf{r}_M \circledast \ldots \circledast \mathbf{r}_{m+1}^{\mathrm{T}}\mathbf{r}_{m+1} \circledast \mathbf{r}_{m-1}^{\mathrm{T}}\mathbf{r}_{m-1} \circledast \ldots \mathbf{r}_2^{\mathrm{T}}\mathbf{r}_2\\
&= (\mathbf{r}_M^{\mathrm{T}}\mathbf{r}_M) \ldots (\mathbf{r}_{m+1}^{\mathrm{T}}\mathbf{r}_{m+1})(\mathbf{r}_{m-1}^{\mathrm{T}}\mathbf{r}_{m-1}) \ldots (\mathbf{r}_2^{\mathrm{T}}\mathbf{r}_2)\\
&= \|\mathbf{r}_M\|^2 \ldots \|\mathbf{r}_{m+1}\|^2 \|\mathbf{r}_{m-1}\|^2 \ldots \|\mathbf{r}_2\|^2. && (16)
\end{aligned}$$

Algorithm 3.1 is a *multilinear projection algorithm via Rank-1 analysis*, which employs a modified CP algorithm, and computes the best fitting rank-1 decomposition of $\boldsymbol{\mathcal{R}}$. Each $\mathbf{r}_m$ is computed efficiently according to (14) with (16) and in an iterative manner by holding all other coefficient vectors fixed.

Next, we will summarize the multilinear projection algorithm [19] that computes the coefficient vectors by applying the MPCA algorithm (MP-MPCA). Since, in principle, the mode-m vectors of $\boldsymbol{\mathcal{R}}$ are multiples of the $\mathbf{r}_m$ coefficient vectors (e.g., for facial image recognition, $\mathbf{r}_{\mathrm{P}}$, $\mathbf{r}_{\mathrm{V}}$, $\mathbf{r}_{\mathrm{L}}$, $\mathbf{r}_{\mathrm{E}}$; cf. the framed rows/columns in Figure 2(b)), matrixizing $\boldsymbol{\mathcal{R}}$ in each mode yields rank-1 matrices, enabling the M-mode SVD algorithm to compute the corresponding coefficient vector. When dealing with facial images, the person coefficient vector $\mathbf{r}_{\mathrm{P}}$ is the leading left-singular vector of the SVD. In practice, the M-mode SVD [15] of $\boldsymbol{\mathcal{R}}$ may not result in a rank-$(1, \ldots, 1)$ decomposition. The MP algorithm introduced in [19] exploits the MPCA algorithm [17] which achieves a locally optimal dimensionality reduction through alternating least squares and computes a *Rank-*$(1, \ldots, 1)$ *Multilinear Projection (MP-MPCA)*. MP-MPCA first optimizes for the computation of orthonormal mode-matrices, which is unnecessary and can result in a rank-$(R_2, \ldots, R_M)$, followed by dimensionality reduction to achieve a rank-$(1, \ldots, 1)$ decomposition. The first step biases the MP-MPCA algorithm. In contrast, the CP initialization computes a rank-1 analysis from the start, which is a better initial condition since $\boldsymbol{\mathcal{R}}$ is a rank-1 tensor.

The MP-MPCA method is sensitive to the sign indeterminacy of the decomposition of the response tensor; i.e., for any pair of factor representation vectors,

$$\begin{aligned}
\boldsymbol{\mathcal{R}} &\simeq \mathbf{r}_2 \circ \cdots \circ \mathbf{r}_i \circ \cdots \circ \mathbf{r}_j \circ \cdots \circ \mathbf{r}_M && (17)\\
&= \mathbf{r}_2 \circ \cdots \circ -\mathbf{r}_i \circ \cdots \circ -\mathbf{r}_j \circ \cdots \circ \mathbf{r}_M, && (18)
\end{aligned}$$

and alternative decompositions can be obtained by flipping the signs of any number of vector pairs. Sign consistency in the MP-MPCA can be achieved analogously to how one might achieve consistency in choosing PCA basis vectors [4]. Note, however, that the MP-CP method starts with a consistent initialization condition, so it is less prone to sign indeterminacy.

The application of the MP-PCA Algorithm or the MP-CP Algorithm 3.1 to an unlabeled test image $\mathbf{d}$ yields causal factor representation vectors $\mathbf{r}_2, \ldots, \mathbf{r}_M$. For recognition, we assign causal mode labels to $\mathbf{d}$ by computing a cosine similarity measure between $\mathbf{r}_m$ and each of the I_m rows $\mathbf{c}_i^{\mathrm{T}}$ of $\mathbf{U}_m$:

$$\arg\max_i \frac{\mathbf{c}_i^{\mathrm{T}}\mathbf{r}_m}{\|\mathbf{c}_i\|\|\mathbf{r}_m\|}. \qquad (19)$$

The probe $\mathbf{d}$ is assigned the label i, where $1 \leq i \leq I_m$, of the signature $\mathbf{c}_i^{\mathrm{T}}$ that maximizes (19). In the particular context of facial image recognition, we denote the $\mathbf{r}_m$ vectors computed by the MP algorithms as $\mathbf{r}_{\mathrm{P}}$, $\mathbf{r}_{\mathrm{V}}$, $\mathbf{r}_{\mathrm{L}}$, and $\mathbf{r}_{\mathrm{E}}$, in association with the people, views, illuminations, and expressions modes, respectively. To recognize the unlabeled test image $\mathbf{d}$, we maximize the set of similarity measures

$$\arg\max_p \frac{\mathbf{p}_p^{\mathrm{T}}\mathbf{r}_{\mathrm{P}}}{\|\mathbf{p}_p\|\|\mathbf{r}_{\mathrm{P}}\|}; \quad \arg\max_v \frac{\mathbf{v}_v^{\mathrm{T}}\mathbf{r}_{\mathrm{V}}}{\|\mathbf{v}_v\|\|\mathbf{r}_{\mathrm{V}}\|};$$
$$\arg\max_l \frac{\mathbf{l}_l^{\mathrm{T}}\mathbf{r}_{\mathrm{L}}}{\|\mathbf{l}_l\|\|\mathbf{r}_{\mathrm{L}}\|}; \quad \arg\max_e \frac{\mathbf{e}_e^{\mathrm{T}}\mathbf{r}_{\mathrm{E}}}{\|\mathbf{e}_e\|\|\mathbf{r}_{\mathrm{E}}\|}; \qquad (20)$$

that for $\mathbf{r}_{\mathrm{P}}$, $\mathbf{r}_{\mathrm{V}}$, $\mathbf{r}_{\mathrm{L}}$, and $\mathbf{r}_{\mathrm{E}}$ find the best matching signatures; i.e., rows $\mathbf{p}_p^{\mathrm{T}}$, $\mathbf{v}_v^{\mathrm{T}}$, $\mathbf{l}_l^{\mathrm{T}}$, and $\mathbf{e}_e^{\mathrm{T}}$ of the causal mode matrices $\mathbf{U}_{\mathrm{P}}$, $\mathbf{U}_{\mathrm{V}}$, $\mathbf{U}_{\mathrm{L}}$, and $\mathbf{U}_{\mathrm{E}}$, respectively. Evaluating the set of similarity measures together enables us to recognize the probe image $\mathbf{d}$ as depicting person p in view v, illumination l, and expression e.

Figure 4 illustrates the architecture of our multilinear recognition system, showing the TensorFaces (MPCA) model using (20). Of course, if we are interested only in recognizing the person depicted in $\mathbf{d}$, we can achieve savings by storing only the person signatures $\mathbf{U}_{\text{P}}$ and performing only the first similarity optimization in (20).

4. Facial Image Recognition Experiments

We will now evaluate the recognition algorithms that we have developed in this paper. There are a number of meaningful experimental scenarios: We employ the Freiburg image database (Figure 1). We represent images by using an MPCA (TensorFaces) model, and in the testing phase we recognize unlabeled test images (probes) by first inferring their coefficient vector representations and then using a similarity measure to label the probes, thus achieving recognition. The inference step may be accomplished using (i) the multiple linear projection (MLP) method [16], or (ii) the multilinear projection (MP) method implemented either (a) by MP-MPCA [19] or (b) by our MP-CP (Algorithm 3.1).

In all the experiments reported below, people depicted in unlabeled test images (probes), which were not part of the training set, are recognized by inferring the person representation associated with the test image and choosing the person label using the similarity methods $\arg\min_{p;\,vle} \|\mathbf{p}_p - \mathbf{r}_{vle}\|$, for MLP and (19) for MP.

Our next experiments employed the Freiburg facial image dataset of 16,875 images and the data tensor $\boldsymbol{\mathcal{D}}$ illustrated in Figure 1.

The trained MPCA (TensorFaces) basis and mode matrices have dimensions $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{8560\times74\times3\times1}$, $\mathbf{U}_{\text{P}} \in \mathbb{R}^{75\times74}$, $\mathbf{U}_{\text{V}} \in \mathbb{R}^{6\times3}$, and $\mathbf{U}_{\text{L}} \in \mathbb{R}^{6\times1}$. Thus, the TensorFaces basis tensors contains 222 basis vectors. However, an image is represented by $74+3+1 = 78$ parameters. The MPCA image representations and response tensors are shown in Figure 2.

We trained the TensorFaces model and obtained recognition results employing the different projection algorithms to compute the person representations of unlabeled test images (probes). Table 1 compares the recognition rates obtained when applying the multiple linear projections (MLP) method and when applying the multilinear projection the MP-MPCA algorithm or with our MP-CP algorithm. Note that our MP-CP algorithm outperforms the MP-MPCA algorithm used in the previous recognition experiment.

Table 2 provides a detailed study of how dimensionality reduction in the trained TensorFaces model affects recognition rates when using the MP-CP algorithm. The table shows recognition percentage rates obtained for the number of people, view, and illumination basis vectors retained as indicated along each axis.

people dimension \ viewpoint dimension	1	2	3	4	5	6
71	21.48	61.60	95.89	25.28	0.02	0
72	21.68	62.86	95.94	94.54	0	0.03
73	19.65	63.13	95.62	93.56	69.53	0.25
74	19.85	61.55	96.81	92.33	0.23	0.21
75	21.43	61.45	97.09	95.93	0.02	0

illumination dimension \ viewpoint dimension	1	2	3	4	5	6
1	21.43	61.45	97.09	95.93	0.02	0
2	2.78	9.42	14.16	5.43	0.26	1.20
3	3.28	1.07	5.25	4.05	35.52	0.13
4	1.56	1.81	3.08	2.29	11.44	65.81
5	1.51	3.52	2.16	2.16	1.78	14.73
6	2.60	5.00	2.04	1.51	3.97	2.34

Table 2. Recognition rates obtained by the MP-CP, rank-1 recognition algorithm with the MPCA (TensorFaces) model subject to various dimensionality reductions. The table shows percentage recognition rates obtained for the number of people, view, and illumination basis vectors retained that are indicated along each axis.

4.1. Practical Face Recognition

Our multilinear framework is clearly relevant to biometric systems. A big challenge for a face recognition system in real-world use is enrolling unwilling participants from one or more unconstrained surveillance images and recognizing them under different imaging conditions. Such a system might be structured as shown in Figure 4. First, a multilinear model is learned from a (hopefully large and representative) set of cooperative participants representative of the human population, each of whom ideally supply a complete set of training images acquired from multiple views, under multiple illuminations, in multiple expressions, etc., thus learning how the different causal factors interact to form an image. An uncooperative subject whose face is detected in one or more surveillance images can then be enrolled into the system, by representing his/her facial image(s) relative to the statistics encoded in the learned model.

Since an individual has the same person representation in a multilinear framework regardless of other imaging conditions, the person representation is extracted by decomposing the surveillance image(s). The multilinear basis tensor $\boldsymbol{\mathcal{T}}$ or $\boldsymbol{\mathcal{M}}$ learnt from the training data is employed and a surveillance image is decomposed by employing multilinear projection (Algorithm 3.1). When more than one image is available, a more robust person representation can be computed by exploiting all the available images and Step 4 of the MP-CP (Algorithm 3.1) is modified for the person mode. Thus, we compute a $\boldsymbol{\mathcal{R}}_i$ for every image i and an associated $\mathbf{y}_{m,i}$ for every image when mode m is the person mode (Steps 2-3 of Algorithm 3.1). Step 4 integrates all this information by defining $\mathbf{R}_{[m]} = [\mathbf{R}_{[m]1} \dots \mathbf{R}_{[m]i} \dots]$, $\mathbf{y}_m = [\mathbf{y}_{m,1}^{\mathrm{T}} \dots \mathbf{y}_{m,i}^{\mathrm{T}} \dots]^{\mathrm{T}}$ and $\mathbf{r}_m := \mathbf{R}_{[m]}\mathbf{y}_m / \|\mathbf{y}_m\|^2$ when m is the person mode. The collection of people representations $\mathbf{U}_{\text{P}}$ can be augmented and employed in recognition. Since a person representation is invariant of the other causal factor associated with image formation, the

MPCA: MLP vs. MP-MPCA vs. MP-CP Recognition Experiment	MLP	MP-MPCA	MP-CP
Training: 75 people, 6 views ($\theta = \pm 35, \pm 20, \pm 5$, $\phi = 0$), 6 illuminations ($\theta = 45$, $\phi = 90 + \delta$, $\delta = \pm 35, \pm 20, \pm 5$) **Testing:** 75 people, 9 views ($\phi = 0 \pm 10, \pm 15, \pm 25, \pm = 30$), 9 illuminations ($\theta = 90 + \delta$, $\delta = \pm 35, \pm 20, \pm 5, \theta = 0$)	92.67%	92.65%	96.81%

Table 1. Facial recognition rates when using the image dataset in Figure 1 to train a TensorFaces model and using the MLP and MP recognition methods.

person representation extracted from a few surveillance images should enable the recognition of the subject from an arbitrary, unlabeled image, as illustrated at the bottom of the figure. This should also make it possible, in principle, to synthesize a complete image set of training images for the subject.

Implementing a prototype recognition system of this kind, especially one that would also support the multilinear fusion of multimodality biometric data (e.g., video and speech) would be a worthwhile future goal.

5. Conclusion

We have introduced a multilinear projection via a rank-1 analysis that employs a modified CP algorithm by extending the underlying mathematics through a generalization of the mode-m product, mode-m identity tensor, and mode-m pseudo-inverse tensor. Our multilinear projection algorithm was applied to facial image recognition and its results were compared with the results achieved by applying a set of multiple linear projections (MLP) [16] as well as the results achieved with multilinear projection that employs MPCA (MP-MPCA) [19]. A detailed study was made of the recognition rates obtained with the Tensorfaces model using the MP-CP and rank-1 algorithms subject to various dimensionality reductions. A strategy for practical biometric systems was also discussed, where an uncooperative subject can be enrolled into the biometric system from a small number of images and then recognized in unconstrained test images.

Acknowledgements

I am grateful to Demetri Terzopoulos for his valuable comments on drafts of this paper, which greatly improved the presentation.

References

[1] B. W. Bader and T. G. Kolda. Efficient MATLAB computations with sparse and factored tensors. *SIAM Journal on Scientific Computing*, 30(1):205–231, December 2007.

[2] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

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

[4] R. Bro, E. Acar, and T. G. Kolda. Resolving the sign ambiguity in the singular value decomposition. *Journal of Chemometrics*, 22:135–140, February 2008.

[5] J. D. Carroll and J. J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, 35:283–319, 1970.

[6] 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] R. Harshman. Foundations of the PARAFAC procedure: Model and conditions for an explanatory factor analysis. Technical Report UCLA Working Papers in Phonetics 16, University of California, Los Angeles, Los Angeles, CA, December 1970.

[8] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, New York, 2001.

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

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

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

[12] 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.

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

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

[15] M. A. O. Vasilescu. Human motion signatures: Analysis, synthesis, recognition. In *Proc. Int. Conf. on Pattern Recognition*, volume 3, pages 456–460, Quebec City, August 2002.

[16] M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis for facial image recognition. In *Proc. Int. Conf. on Pattern Recognition*, volume 2, pages 511–514, Quebec City, August 2002.

[17] M. A. O. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In *Proc. IEEE Conf. on*

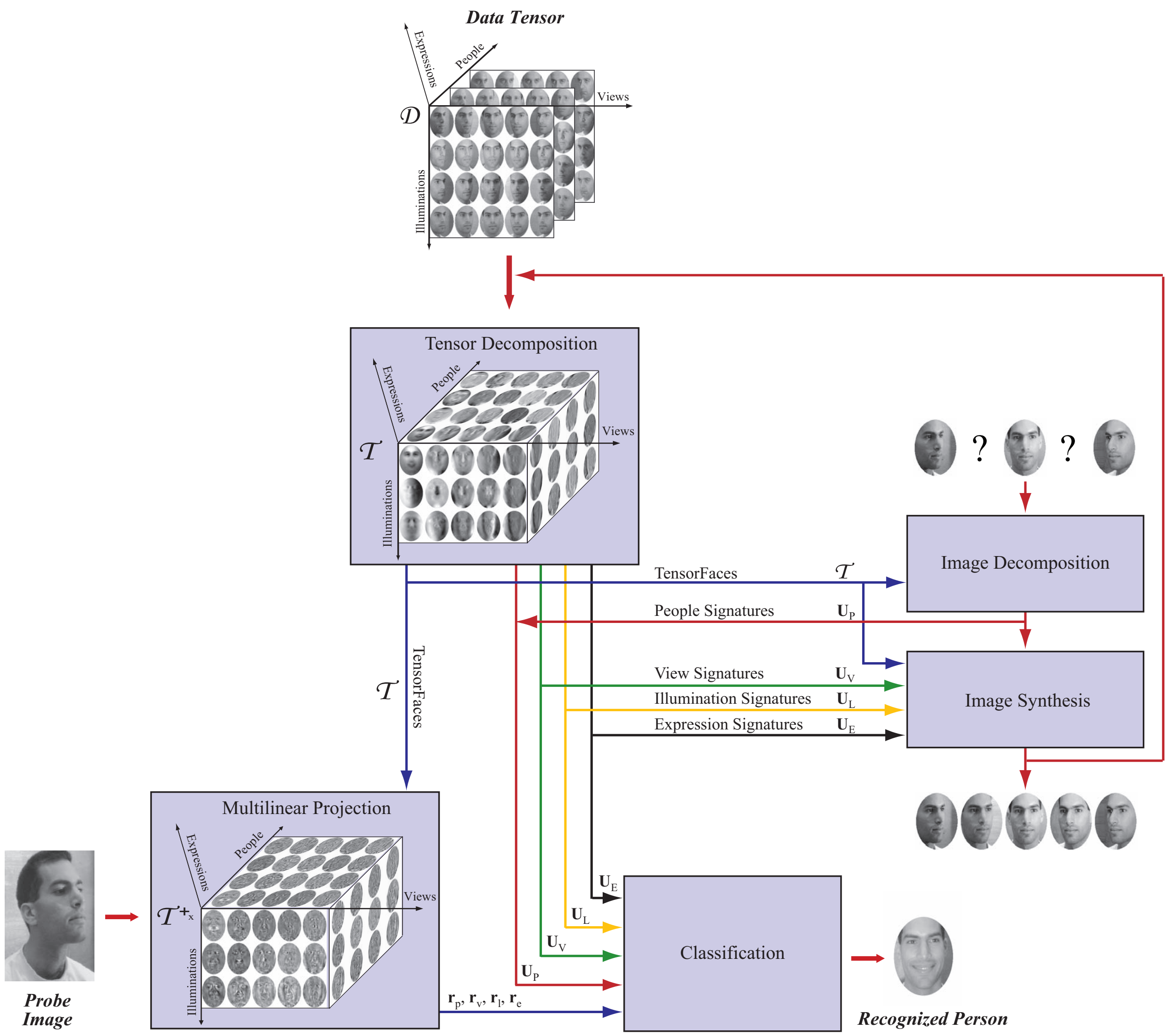


Figure 4. Architecture of a multilinear facial image recognition system. A facial training image ensemble including different people, expressions, views, illuminations, and expressions is organized as a data tensor. The data tensor made up of willing participants is decomposed in the (offline) learning phase to train a multilinear model. An uncooperative subject whose face is detected in one or more surveillance images can then be enrolled into the system, by representing his/her facial image(s) relative to the statistics encoded in the learned model. In the (online) recognition phase, the model recognizes a previously unseen probe image as one of the known people in the database. In principle, the trained generative model can also synthesize novel images of known or unknown persons from one or more of their facial images.

Computer Vision and Pattern Recognition, volume II, pages 93–99, Madison, WI, 2003.

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

[19] M. A. O. Vasilescu and D. Terzopoulos. Multilinear projection for appearance-based recognition in the tensor framework. In *Proc. 11th IEEE International Conf. on Computer Vision (ICCV'07)*, pages 1–8, Rio de Janeiro, Brazil, 2007.

[20] M. A. O. Vasilescu and D. Terzopoulos. Multilinear (tensor) ICA and dimensionality reduction. In *Independent Component Analysis and Signal Separation, Proc. 7th International Conf. on Independent Component Analysis (ICA 2007)*, volume 4666 of *Lecture Notes in Computer Science*, pages 818–826. Springer, 2007.