Multilinear Independent Components Analysis

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

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

Abstract

Independent Components Analysis (ICA) maximizes the statistical independence of the representational components of a training image ensemble, but it cannot distinguish between the different factors, or modes, inherent to image formation, including scene structure, illumination, and imaging. We introduce a nonlinear, multifactor model that generalizes ICA. Our Multilinear ICA *(MICA) model of image ensembles learns the statistically independent components of multiple factors. Whereas ICA employs linear (matrix) algebra, MICA exploits multilinear (tensor) algebra. We furthermore introduce a* multilinear projection algorithm *which projects an unlabeled test image into the N constituent mode spaces to simultaneously infer its mode labels. In the context of facial image ensembles, where the mode labels are person, viewpoint, illumination, expression, etc., we demonstrate that the statistical regularities learned by MICA capture information that, in conjunction with our multilinear projection algorithm, improves automatic face recognition.*

1 Introduction

A key problem in data analysis for pattern recognition and signal processing is finding a suitable representation. For historical and computational reasons, linear models that optimally encode particular statistical properties of the data have been broadly applied. In particular, the linear, appearance-based face recognition method known as Eigenfaces [9] is founded on the principal components analysis (PCA) of facial image ensembles [7]. PCA encodes pairwise relationships between pixels—the second-order, correlational structure of the training image ensemble—but it ignores all higher-order pixel relationships—the higher-order statistical dependencies. By contrast, a generalization of PCA known as independent components analysis (ICA) [5] learns a set of statistically independent components by analyzing the higher-order dependencies in the training data in addition to the correlations. However, ICA cannot distinguish between higher-order statistics that rise from different factors, or modes, inherent to image formation—factors pertaining to scene structure, illumination, and imaging.

In this paper, we introduce a nonlinear, multifactor model of image ensembles that generalizes conventional ICA.[1] Whereas ICA employs linear (matrix) algebra, our *Multilinear ICA* (MICA) model exploits multilinear (tensor) algebra. Unlike its conventional, linear counterpart, MICA is able to learn the interactions of multiple factors inherent to image formation and separately encode the higher-order statistics of each of these factors. Unlike our recently proposed multilinear generalization of Eigenfaces dubbed TensorFaces [11] which encodes only second order statistics associated with the different factors inherent to image formation, MICA encodes higher order dependencies associated with the different factors.

ICA has been employed in face recognition [1, 2] and, like PCA, it works best when person identity is the only factor that is permitted to vary. If additional factors, such as illumination, viewpoint, and expression can modify facial images, recognition rates decrease dramatically. This problem is addressed by our multilinear framework. In the context of recognition, our second contribution in this paper is a novel, *multilinear projection algorithm*. This algorithm projects an unlabeled test image into the multiple factor representation spaces in order to simultaneously infer the person, viewpoint, illumination, expression, etc., labels associated with the test image. Equipped with this new algorithm, we demonstrate the application of multilinear ICA to the problem of face recognition under varying viewpoint and illumination, obtaining significantly improved results.

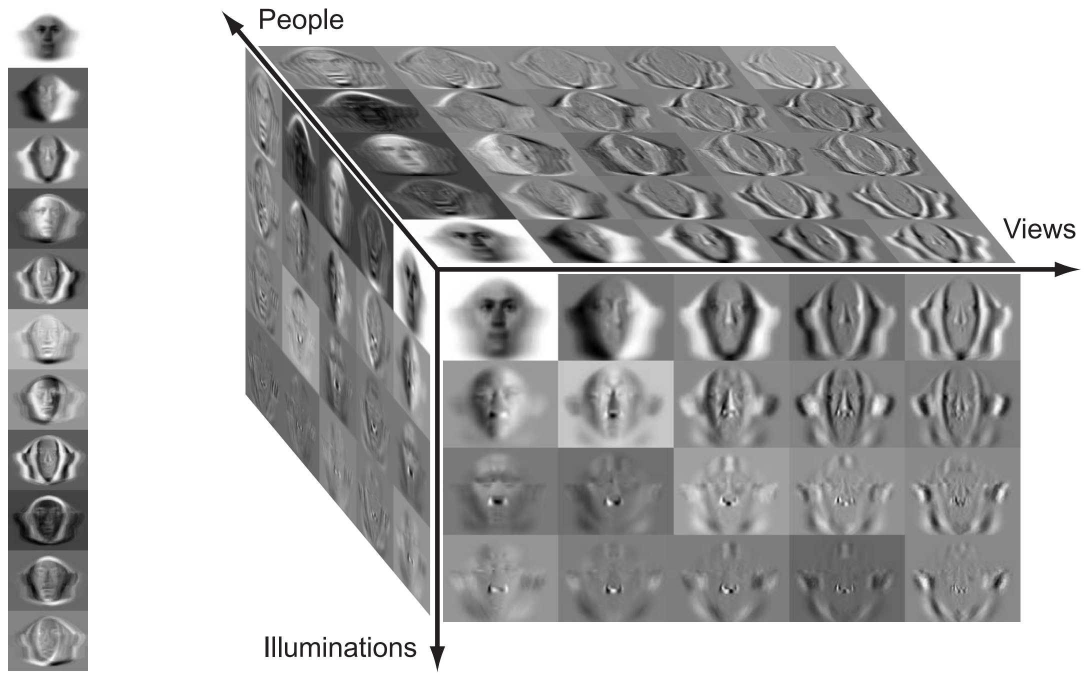
After reviewing the mathematical foundations of our work in the next section, we introduce our multilinear ICA algorithm in Section 3 and develop the associated recognition algorithm in Section 4. Section 5 describes our experiments and presents results. Section 6 concludes the paper.

2 Mathematical Background

In this section, we review the mathematics of PCA, multilinear PCA, and ICA.

[1]An initial description of this work appeared as an extended abstract in the *Learning 2004 Workshop*, Snowbird, UT, April, 2004.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**



(a) (b)

Figure 1: Eigenfaces and TensorFaces bases for an ensemble of 2,700 facial images spanning 75 people, each imaged under 6 viewing and 6 illumination conditions (see Section 5). (a) PCA eigenvectors (eigenfaces), which are the principal axes of variation across all images. (b) A partial visualization of the $75 \times 6 \times 6 \times 8560$ TensorFaces representation of $\mathcal{D}$, obtained as $\mathcal{T} = \mathcal{Z} \times_4 \mathbf{U}_{\text{pixels}}$.

2.1 PCA

The principal components analysis of an ensemble of I_2 images is computed by performing an SVD on a $I_1 \times I_2$ *data matrix* $\mathbf{D}$ whose columns are the "vectored" I_1-pixel "centered" images.[2] The matrix $\mathbf{D} \in \mathbb{R}^{I_1 \times I_2}$ is a two-mode mathematical object that has two associated vector spaces, a row space and a column space. In a factor analysis of $\mathbf{D}$, the SVD orthogonalizes these two spaces and decomposes the matrix as

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{1}$$

the product of an orthogonal column-space represented by the left matrix $\mathbf{U} \in \mathbb{R}^{I_1 \times J_1}$, a diagonal singular value matrix $\mathbf{\Sigma} \in \mathbb{R}^{J_1 \times J_2}$ with diagonal entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$ called the singular values of $\mathbf{D}$, and an orthogonal row space represented by the right matrix $\mathbf{V} \in \mathbb{R}^{I_2 \times J_2}$. The eigenvectors $\mathbf{U}$ are also called the *principal component* (or Karhunen-Loeve) directions of $\mathbf{D}$ (Fig. 1(a)).

Optimal dimensionality reduction in matrix principal components analysis is obtained by truncation of the singular value decomposition (i.e., deleting eigenvectors associated with the smallest singular values).

[2] Each vectored-centered image is obtained by subtracting the mean image of the ensemble from each input image and identically arranging the resulting pixels into a column vector.

2.2 Multilinear PCA

The analysis of an ensemble of images resulting from the confluence of multiple factors, or modes, related to scene structure, illumination, and viewpoint is a problem in multilinear algebra [11]. Within this mathematical framework, the image ensemble is represented as a higher-order tensor. This image data tensor $\mathcal{D}$ must be decomposed in order to separate and parsimoniously represent the constituent factors. To this end, we prescribe the *N-mode SVD* algorithm [11, 10], a multilinear extension of the aforementioned conventional matrix SVD.

Appendix A overviews the mathematics of multilinear analysis. Briefly, 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 (25)), of N-orthogonal spaces, as follows:

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

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 orthonor-

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 IEEE

mal vectors spanning the column space of matrix $\mathbf{D}_{(n)}$ resulting from the *mode-n flattening* of $\mathcal{D}$ (see Appendix A).[3]

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

1. For $n = 1, \ldots, N$, compute matrix $\mathbf{U}_n$ in (2) by computing the SVD of the flattened matrix $\mathbf{D}_{(n)}$ and setting $\mathbf{U}_n$ to be the left matrix of the SVD. [4]

2. Solve for the core tensor as follows:

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

Dimensionality reduction in the linear case does not have a trivial multilinear counterpart. According to [6, 4], 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.

The tensor basis associated with multilinear PCA is displayed in Fig. 1(b).

2.3 ICA

The independent components analysis of multivariate data looks for a sequence of projections such that the projected data look as far from Gaussian as possible. ICA can be applied in two ways [1, 2]: Architecture I applies ICA to $\mathbf{D}^T$, each of whose rows is a different image, and finds a spatially independent basis set that reflects the local properties of faces. On the other hand, Architecture II applies ICA to $\mathbf{D}$ and finds a set of coefficients that are statistically independent while the basis reflects the global properties of faces.

Architecture I: ICA starts essentially from the factor analysis or PCA solution (1), and computes a rotation of the principal components such that they become *independent components* [8]; that is ICA rotates the principal component directions $\mathbf{U}$ in (1) as follows:

$$\begin{aligned} \mathbf{D}^T &= \mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^T & (4)\\ &= \left(\mathbf{V}\boldsymbol{\Sigma}\mathbf{W}^{-1}\right)\left(\mathbf{W}\mathbf{U}^T\right) & (5)\\ &= \mathbf{K}^T\mathbf{C}^T, & (6) \end{aligned}$$

where every column of $\mathbf{D}$ is a different image, $\mathbf{W}$ is an invertible transformation matrix that is computed by the ICA algorithm, $\mathbf{C} = \mathbf{U}\mathbf{W}^T$ are the independent components (Fig. 2(a)), and $\mathbf{K} = \mathbf{W}^{-T}\boldsymbol{\Sigma}\mathbf{V}^T$ are the coefficients. Various objective functions, such as those based on mutual information, negentropy, higher-order cumulants, etc., are presented in the literature for computing the independent components along with different optimization methods for extremizing these objective functions [5]. Dimensionality reduction with ICA is usually performed in the PCA preprocessing stage.

Alternatively, in **Architecture II**, ICA is applied to $\mathbf{D}$ and it rotates the principal components directions such that the coefficients are statistically independent, as follows:

$$\begin{aligned} \mathbf{D} &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T & (7)\\ &= \left(\mathbf{U}\mathbf{W}^{-1}\right)\left(\mathbf{W}\boldsymbol{\Sigma}\mathbf{V}^T\right) & (8)\\ &= \mathbf{C}\mathbf{K}. & (9) \end{aligned}$$

where $\mathbf{C} = \mathbf{U}\mathbf{W}^{-1}$ is the basis matrix and $\mathbf{K} = \mathbf{W}\boldsymbol{\Sigma}\mathbf{V}^T$ are the statistically independent coefficients.

Note, that $\mathbf{C}$, $\mathbf{K}$ and $\mathbf{W}$ are computed differently in the two different architectures. Architecture I yields statistically independent bases, while Architecture II yields a "factorial code".

Like PCA, ICA is a linear analysis method, hence it is not well suited to the representation of multi-factor image ensembles. To address this shortcoming, we next propose a multilinear generalization of ICA.

3 Multilinear ICA

Analogously to (2), multilinear ICA is obtained by decomposing the data tensor $\mathcal{D}$ as the mode-n product of N mode matrices $\mathbf{C}_n$ and a core tensor $\mathcal{S}$, as follows:

$$\mathcal{D} = \mathcal{S} \times_1 \mathbf{C}_1 \times_2 \mathbf{C}_2 \ldots \times_n \mathbf{C}_n \ldots \times_N \mathbf{C}_N. \quad (10)$$

The N**-mode ICA algorithm** is as follows:

1. For $n = 1, \ldots, N$, compute the mode matrix $\mathbf{C}_n$ in (10) in an architecture-dependent way (see below).

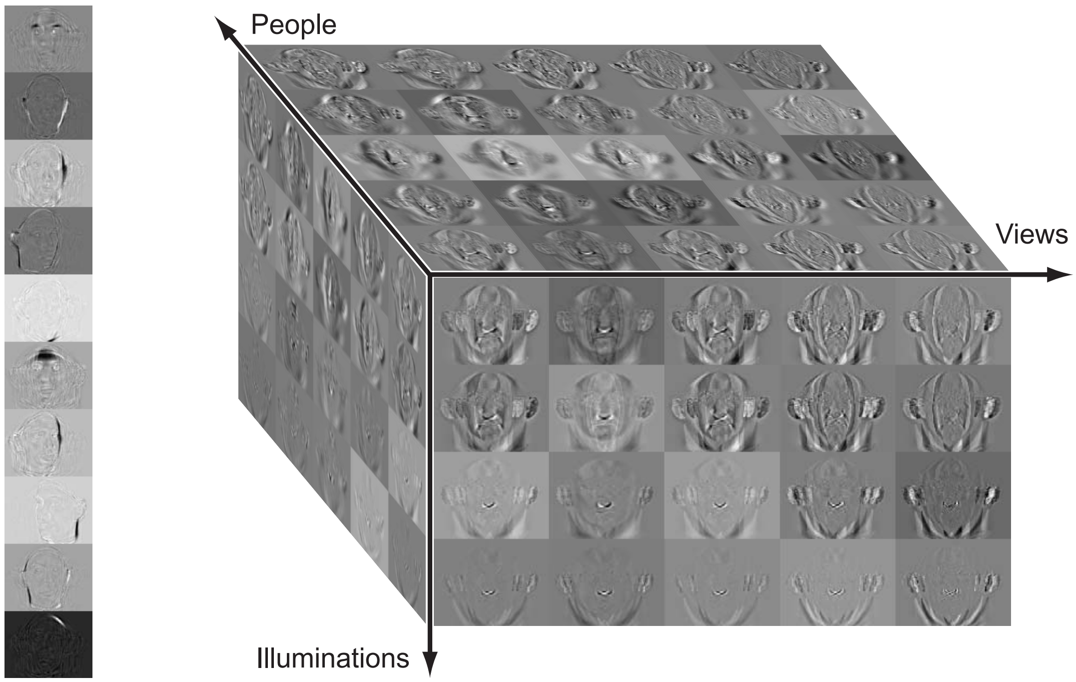
2. Solve for the core tensor as follows:

$$\mathcal{S} = \mathcal{D} \times_1 \mathbf{C}_1^{-1} \times_2 \mathbf{C}_2^{-1} \ldots \times_n \mathbf{C}_n^{-1} \ldots \times_N \mathbf{C}_N^{-1}. \quad (11)$$

As in ICA, there are two strategies for multilinear independent components analysis (MICA). Architecture I results in a factorial code, where each set of coefficients that encodes people, viewpoints, illuminations, etc., is statistically independent, while Architecture II finds a set of independent bases across people, viewpoints, illuminations, etc.

[3] Note that the conventional SVD in (1) can be rewritten as $\mathbf{D} = \boldsymbol{\Sigma} \times_1 \mathbf{U} \times_2 \mathbf{V}$ using mode-n products.

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



(a) (b)

Figure 2: ICA and MICA bases for an ensemble of 2,700 facial images spanning 75 people, each imaged under 6 viewing and 6 illumination conditions (see Section 5). (a) Independent components $\mathbf{C}_{\text{pixels}}$. (b) A partial visualization of the $75 \times 6 \times 6 \times 8560$ MICA representation of $\mathcal{D}$, obtained as $\mathcal{B} = \mathcal{S} \times_4 \mathbf{C}_{\text{pixels}}$.

Architecture I: Transposing the flattened data tensor $\mathcal{D}$ in the nth mode and computing the ICA as in (4)–(6), we obtain:

$$\mathbf{D}_{(n)}^T = \mathbf{V}_n \mathbf{\Sigma}_n^T \mathbf{U}_n^T \tag{12}$$

$$= \left(\mathbf{V}_n \mathbf{\Sigma}_n^T \mathbf{W}_n^{-1}\right)\left(\mathbf{W}_n \mathbf{U}_n^T\right) \tag{13}$$

$$= \mathbf{K}_n^T \mathbf{C}_n^T, \tag{14}$$

where the mode matrices are given by

$$\mathbf{C}_n = \mathbf{U}_n \mathbf{W}_n^T. \tag{15}$$

The columns associated with each of the mode matrices, $\mathbf{C}_n$ are statistically independent. We can derive the relationship between N-mode ICA and N-mode SVD (2) in the context of Architecture I as follows:

$$\begin{aligned}
\mathcal{D} &= \mathcal{Z} \times_1 \mathbf{U}_1 \ldots \times_N \mathbf{U}_N \\
&= \mathcal{Z} \times_1 \mathbf{U}_1 \mathbf{W}_1^T \mathbf{W}_1^{-T} \ldots \times_N \mathbf{U}_N \mathbf{W}_N^T \mathbf{W}_N^{-T} \\
&= \mathcal{Z} \times_1 \mathbf{C}_1 \mathbf{W}_1^{-T} \ldots \times_N \mathbf{C}_N \mathbf{W}_N^{-T} \\
&= (\mathcal{Z} \times_1 \mathbf{W}_1^{-T} \ldots \times_N \mathbf{W}_N^{-T}) \times_1 \mathbf{C}_1 \ldots \times_N \mathbf{C}_N \\
&= \mathcal{S} \times_1 \mathbf{C}_1 \ldots \times_N \mathbf{C}_N,
\end{aligned}$$

where the core tensor $\mathcal{S} = \mathcal{Z} \times_1 \mathbf{W}_1^{-T} \ldots \times_N \mathbf{W}_N^{-T}$.

Architecture II: Flattening the data tensor $\mathcal{D}$ in the nth mode and computing the ICA as in (7)–(9), we obtain:

$$\mathbf{D}_{(n)} = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n^T \tag{16}$$

$$= \left(\mathbf{U}_n \mathbf{W}_n^{-1}\right)\left(\mathbf{W}_n \mathbf{\Sigma}_n \mathbf{V}_n^T\right) \tag{17}$$

$$= \mathbf{C}_n \mathbf{K}_n, \tag{18}$$

where the mode matrices are given by

$$\mathbf{C}_n = \mathbf{U}_n \mathbf{W}_n^{-1}. \tag{19}$$

Architecture II results in a set of basis vectors that are statistically independent across the different modes. Note that the $\mathbf{W}_n$ in (19) differ from those in (15).

We can derive the relationship between N-mode ICA and N-mode SVD (2) in the context of Architecture II as follows:

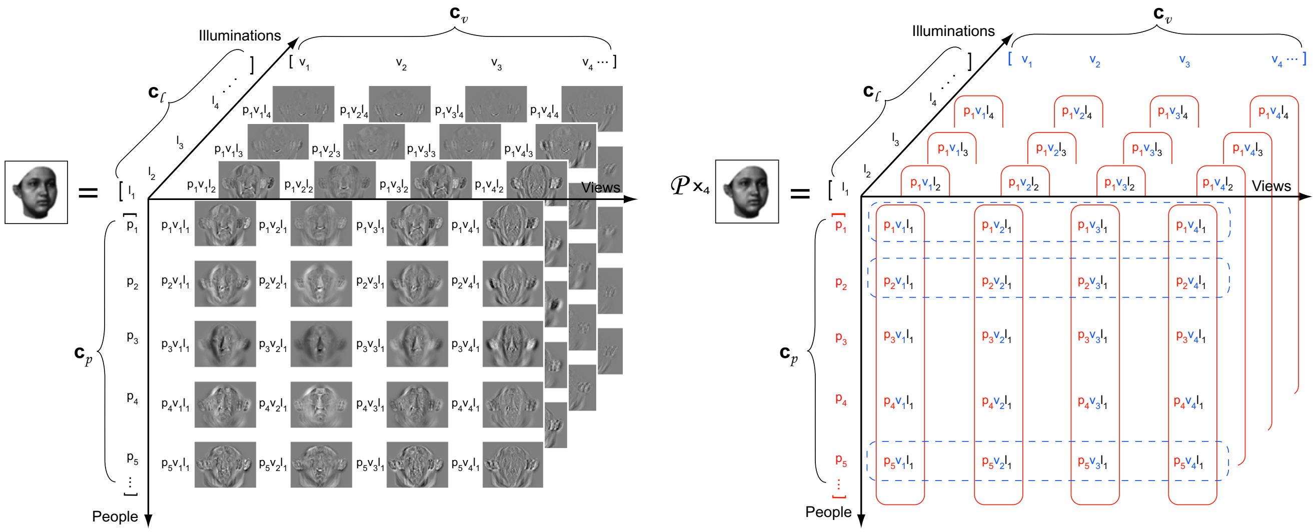
$$\begin{aligned}
\mathcal{D} &= \mathcal{Z} \times_1 \mathbf{U}_1 \ldots \times_N \mathbf{U}_N \\
&= \mathcal{Z} \times_1 \mathbf{U}_1 \mathbf{W}_1^{-1} \mathbf{W}_1 \ldots \times_N \mathbf{U}_N \mathbf{W}_N^{-1} \mathbf{W}_N \\
&= \mathcal{Z} \times_1 \mathbf{C}_1 \mathbf{W}_1 \ldots \times_N \mathbf{C}_N \mathbf{W}_N \\
&= (\mathcal{Z} \times_1 \mathbf{W}_1 \ldots \times_N \mathbf{W}_N) \times_1 \mathbf{C}_1 \ldots \times_N \mathbf{C}_N \\
&= \mathcal{S} \times_1 \mathbf{C}_1 \ldots \times_N \mathbf{C}_N,
\end{aligned}$$

where the core tensor $\mathcal{S} = \mathcal{Z} \times_1 \mathbf{W}_1 \ldots \times_N \mathbf{W}_N$.

4 Recognition

Our approach performs a multilinear ICA decomposition of the tensor $\mathcal{D}$ of vectored training images $\mathbf{d}_{p,v,l}$,

$$\mathcal{D} = \mathcal{B} \times_1 \mathbf{C}_{\text{people}} \times_2 \mathbf{C}_{\text{views}} \times_3 \mathbf{C}_{\text{illums}}, \tag{20}$$



(a) (b)

Figure 3: (a) Image representation $\mathbf{d}^T = \mathcal{B} \times_1 \mathbf{c}_p^T \times_2 \mathbf{c}_v^T \times_3 \mathbf{c}_l^T$. (b) Given an unlabeled test image $\mathbf{d}^T$, the associated coefficient vectors $\mathbf{c}_p, \mathbf{c}_v, \mathbf{c}_l$ are computed by decomposing the response tensor $\mathcal{R} = \mathcal{P} \times_4 \mathbf{d}^T$ using the N-mode SVD algorithm.

extracting a set of mode matrices—the matrix $\mathbf{C}_{\text{people}}$ containing row vectors $\mathbf{c}_p^T$ of coefficients for each person p, the matrix $\mathbf{C}_{\text{views}}$ containing row vectors $\mathbf{c}_v^T$ of coefficients for each view direction v, and the matrix $\mathbf{C}_{\text{illums}}$ containing row vectors $\mathbf{c}_l^T$ of coefficients for each illumination direction l—and a MICA basis tensor

$$\mathcal{B} = \mathcal{S} \times_4 \mathbf{C}_{\text{pixels}} \tag{21}$$

that governs the interaction between the different mode matrices (Fig. 2(b)). Given an unlabeled test image, recognition is performed by inferring the associated coefficient vectors.

The recognition algorithm for TensorFaces proposed in our earlier work [10] was based on a linear projection approach.[5] It computed a set of linear projection operators for each mode, which yielded a set of candidate coefficients per mode.

We will now develop a multilinear method for simultaneously inferring the identity, illumination, viewpoint, etc., coefficient vectors of an unlabeled, test image. Our method maps an image from the pixel space to N different constituent mode spaces. We obtain a new recognition algorithm that is based on the multilinear structure of the tensor framework and the statistical independence properties of ICA.

From (20), MICA represents the unlabeled, test image by a set of coefficient vectors:

$$\mathbf{d}^T = \mathcal{B} \times_1 \mathbf{c}_p^T \times_2 \mathbf{c}_v^T \times_3 \mathbf{c}_l^T, \tag{23}$$

where the coefficient vector $\mathbf{c}_p$ encodes the person, the coefficient vector $\mathbf{c}_v$ encodes the viewpoint, and the coefficient vector $\mathbf{c}_l$ encodes the illumination. The multilinear representation of an image is illustrated in Fig. 3(a).

The response tensor, illustrated in Fig. 3(b), is computed as

$$\mathcal{R} = \mathcal{P} \times_4 \mathbf{d}^T, \tag{24}$$

where the "projection tensor" $\mathcal{P}$ is obtained by re-tensorizing matrix $\mathbf{P}_{\text{(pixels)}} = \mathbf{B}_{\text{(pixels)}}^{-T}$ (the matrix $\mathbf{B}_{\text{(pixels)}}$ is the pixel-mode flattening of tensor $\mathcal{B}$).

The response $\mathcal{R}$ has the structure $(\mathbf{c}_p \circ \mathbf{c}_v \circ \mathbf{c}_l)$.[6] It is a tensor of rank (1,1,1) as it is the outer product of the three coefficient vectors $\mathbf{c}_p$, $\mathbf{c}_v$, and $\mathbf{c}_l$.

The structure of $\mathcal{R}$ enables us to compute the three coefficient vectors via a tensor decomposition using the N-mode SVD algorithm. This is because the fibers (columns, rows, tubes) of $\mathcal{R}$ are multiples of their corresponding coefficient vectors ($\mathbf{c}_p$, $\mathbf{c}_v$, $\mathbf{c}_l$) (cf. the boxed rows/columns in Fig. 3(b)). Thus, flattening $\mathcal{R}$ in each mode yields rank 1 matrices, enabling the modal SVD to compute the corresponding coefficient vector. The N-mode SVD thus maps

[5] In the PCA (eigenfaces) technique, one decomposes a data matrix $\mathbf{D}$ of known, training facial images $\mathbf{d}_d$ into a basis matrix $\mathbf{U}_{\text{pixels}}$ and a matrix of coefficient vectors. In our tensor notation, an unlabeled, test facial image $\mathbf{d}$ can be decomposed and represented as

$$\mathbf{d}^T = \mathbf{\Sigma} \times_1 \mathbf{c}_p^T \times_2 \mathbf{U}_{\text{pixels}}. \tag{22}$$

We obtain the vector of "person coefficients" $\mathbf{p}$ associated with $\mathbf{d}$ as follows: $\mathbf{c}_p^T = \mathbf{P} \times_2 \mathbf{d}^T$, where the linear projection operator $\mathbf{P} = \mathbf{U}_{\text{pixels}} \mathbf{\Sigma}^{-1}$.

[6] We can show that $\mathcal{P} \times_4 \mathbf{d}^T = \mathcal{I} \times_4 (\mathbf{c}_l^T \otimes \mathbf{c}_v^T \otimes \mathbf{c}_p^T)$, where $\otimes$ denotes the Kronecker product and $\mathcal{I}$ is the re-tensorized identity matrix $\mathbf{I}_{\text{(pixels)}} = (\mathbf{B}^{+}_{\text{(pixels)}} \mathbf{B}_{\text{(pixels)}})^T$.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 IEEE

the response tensor $\mathcal{R}$ into N different mode spaces that explicitly account for the contribution of each mode—person, viewpoint, illumination.

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 test image $\mathbf{d}^T$, we apply a normalized nearest neighbor classification scheme by computing normalized scalar products between $\mathbf{c}_p$ and each of the row vectors of the people mode matrix $\mathbf{C}_{\text{people}}$.

5 Experiments

In our experiments, we employ gray-level facial images of 75 subjects. Each subject is imaged from 15 different viewpoints ($\theta = -35°$ to $+35°$ in $5°$ steps on the horizontal plane $\phi = 0°$) under 15 different illuminations ($\theta = -35°$ to $+35°$ in $5°$ steps on an inclined plane $\phi = 45°$). Fig. 4(b) shows the full 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 shown boxed in Fig. 4(a). The 75 scans shown in the figure were recorded using a Cyberware^{TM} 3030PS laser scanner and are part of the 3D morphable faces database created at the University of Freiburg [3].

Recognition Experiment	**PCA**	**ICA**	**MPCA**	**MICA**
Training: 75 people, 6 viewpoints ($\theta = \pm 35, \pm 20, \pm 5, \phi = 0$), 6 illuminations ($\theta = 45, \phi = 90 + \delta, \delta = \pm 35, \pm 20, \pm 5$) **Testing:** 75 people, 9 viewpoints ($\phi = 0 \pm 10, \pm 15, \pm 25, \pm = 30$), 9 illuminations ($\theta = 90 + \delta, \delta = \pm 35, \pm 20, \pm 5, \theta = 0$)	83.9%	89.5%	93.4%	98.14%

As the above table shows, in our experiments with $16,875$ images captured from the University of Freiberg 3D Morphable Faces Database, MICA yields better recognition rates than PCA (eigenfaces) and ICA in scenarios involving the recognition of people imaged in previously unseen viewpoints and illuminations. MICA training employed an ensemble of 2,700 images. Fig. 2(b) illustrates the MICA basis derived from the training ensemble, while Fig. 2(a) illustrates the ICA basis.

6 Conclusion

Motivated by the reported outperformance in the face recognition literature of PCA by ICA in the linear case where only a single factor is allowed to vary, and the outperformance of PCA by TensorFaces when multiple factors are allowed to vary, it is natural to ask whether there a multilinear generalization of ICA and if its performance is better than the other two methods. In this paper, we developed a multilinear generalization of ICA and successfully applied our multilinear ICA (MICA) algorithm to a multimodal face recognition problem involving multiple people imaged under different viewpoints and illuminations. We also introduced a multilinear projection algorithm for recognition, which projects an unlabeled test image into the N constituent mode spaces to infer its mode labels—person, viewpoint, illumination, etc. In our experiments, we obtained improved recognition results relative to the prior methods under consideration, because MICA disentangles the multiple factors inherent to image formation and explicitly represents the higher-order statistics associated with each factor.

Acknowledgements

This research was funded by the Technical Support Working Group (TSWG) of the US Department of Defense. The 3D Morphable faces database was obtained courtesy of Prof. Sudeep Sarkar of the University of South Florida as part of the USF HumanID 3D database.

A Multilinear Math

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 [11]). The *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.

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

$$(\mathcal{A} \times_n \mathbf{M})_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} m_{j_n i_n}. \quad (25)$$

The mode-n product can be expressed in terms of flattened matrices as $\mathbf{B}_{(n)} = \mathbf{M}\mathbf{A}_{(n)}$. 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 $(\mathcal{A} \times_n \mathbf{U}) \times_n \mathbf{V} = \mathcal{A} \times_n (\mathbf{V}\mathbf{U})$.

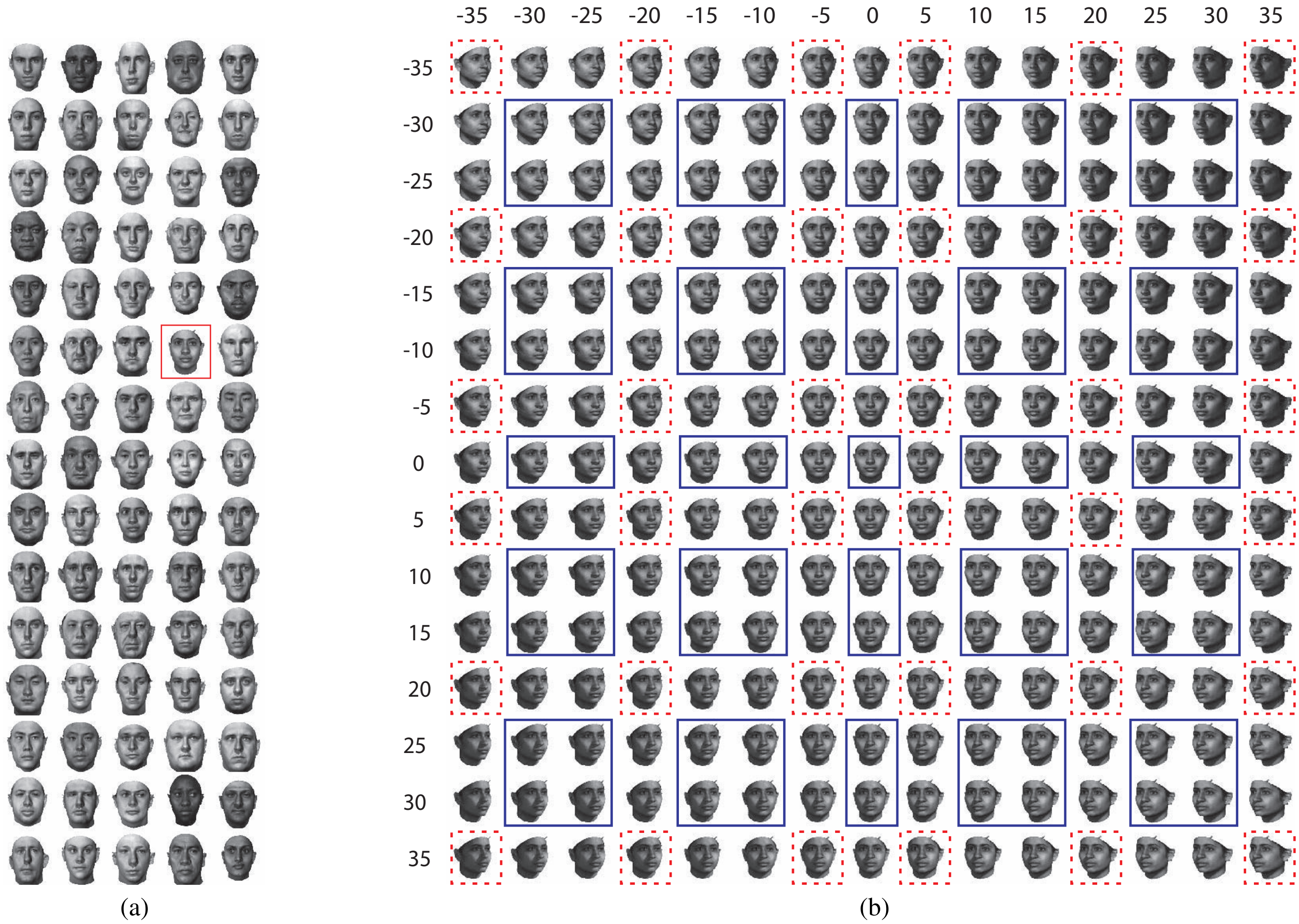


Figure 4: (a) 3D scans of 75 subjects, recorded using a Cyberware^{TM} 3030PS laser scanner as part of the University of Freiburg 3D morphable faces database [3]. (b) Full set of facial images for a subject (boxed head in (a)), viewed from 15 different viewpoints (across) under 15 different illuminations (down). The dashed-boxed images served as training images; the solid-boxed images served as testing images.

References

[1] M.S. Bartlett. *Face Image Analysis by Unsupervised Learning*. Kluwer Academic, Boston, 2001.

[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 *SIGGRAPH 99 Conference Proceedings*, pages 187–194. ACM SIGGRAPH, 1999.

[4] 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.

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

[6] 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.

[7] 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.

[8] J. Friedman T. Hastie, R. Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2001.

[9] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–590, Hawai, 1991.

[10] 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.

[11] M.A.O. 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.