A Multilinear Algebraic Framework for Analysis, Recognition & Synthesis

M. Alex O. Vasilescu

MIT Media Laboratory

www.media.mit.edu/~maov

Visual Computing

Image Science

- Image synthesis
  - *Mathematical forward problem*
  - *Computer graphics*
- Image analysis
  - *Mathematical inverse problem*
  - *Computer vision*
- Image recognition
  - *Pattern recognition*
  - *Statistical machine learning*

Why is Face Recognition Difficult?

Viewpoint changes

Why is Face Recognition Difficult?

Illumination Changes

Appearance-Based Recognition

Recognition of 3D objects (faces) directly from their appearance in ordinary images

- PCA / Eigenimages:
  - *[Sirovich & Kirby 1987]*
    - *"Low Dimensional Procedure for the Characterization of Human Faces"*
  - *[Turk & Pentland 1991]*
    - *"Face Recognition Using Eigenfaces"*
  - *[Murase & Nayar 1995]*
    - *"Visual learning and recognition of 3D objects from appearance"*

Linear Algebra

The algebra of vectors and matrices

- Traditionally of great value in image science
  - *Fourier transform*
  - *Karhunen-Loeve transform (PCA)*
  - *Eigenfaces*
- Linear methods model:
  - *Linear operators over a vector space*
  - *Single-factor linear variation in image formation*
  - *The linear combination of multiple sources (ICA)*

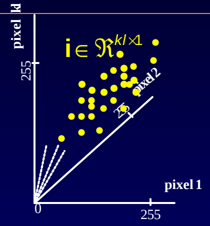
Multilinear Algebra

The algebra of higher-order (>2) tensors

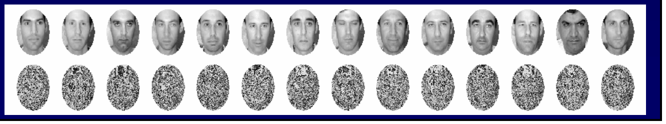
- A unifying mathematical framework for image science
- Natural images result from the interaction of multiple factors related to
  - *scene geometry*
  - *Illumination*
  - *Imaging*
- Multilinear algebra can explicitly represent multiple factors
  - *Multilinear operators over a* ***set*** *of vector spaces*
- Multilinear algebra subsumes linear algebra as a special case

Images

$I \in \Re^{k \times l}$

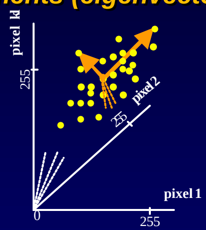


An image is a point in $\Re^{kl \times 1}$ dimensional space



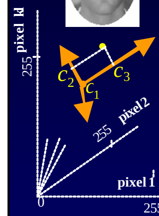
Eigenimages

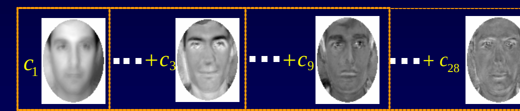
Principal components (eigenvectors) of image ensemble



- Typically computed using the SVD Algorithm

Linear Representation





Running Sum:

1 term | 3 terms | 9 terms | 28 terms

Eigenfaces

- Facial images

- Eigenfaces basis vectors capture the variability in facial appearance

- Eigenfaces have been successful in simple facial recognition problem
  - *i.e., frontal images with fixed illumination*

The Problem with Linear (PCA) Appearance-Based Recognition Methods

Eigenimages work best for recognition when only a single factor – e.g., object identity – is allowed to vary

- However, natural images are the consequence of ***multiple factors*** (or modes) related to scene structure, illumination and imaging

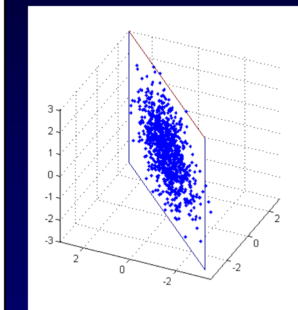
Our Approach

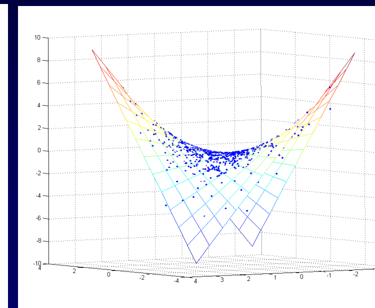
[Vasilescu & Terzopoulos, ECCV 02, ICPR 02, CVPR 03, CVPR 05]

A nonlinear appearance-based technique

- Our appearance-based model ***explicitly accounts*** for each of the multiple factors inherent in image formation
- Multilinear algebra, the algebra of higher order tensors
- Applied to facial images, we call our tensor technique "TensorFaces"

Linear vs Multilinear Manifolds



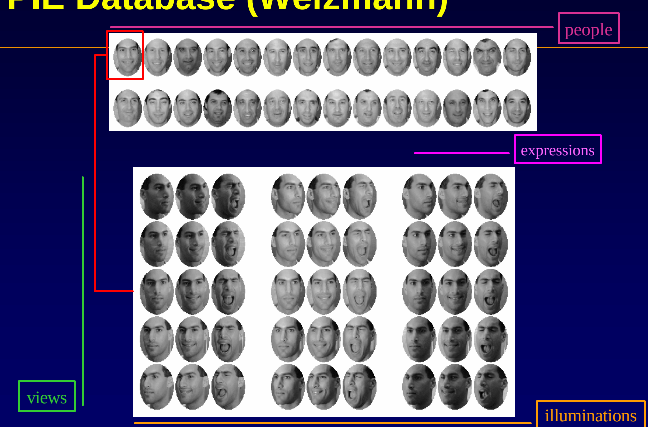


Preliminary Recognition Results

[Vasilescu & Terzopoulos, ICPR'02]

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

PIE Database (Weizmann)



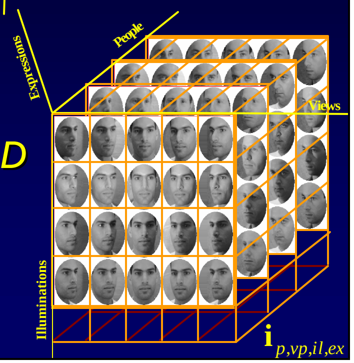
Data Organization

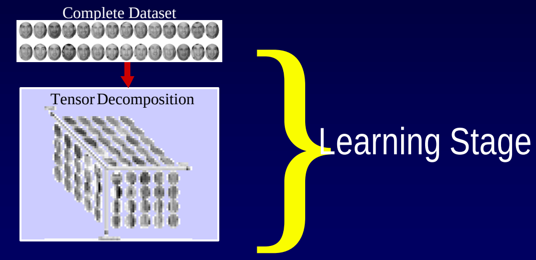
Linear/PCA: *Data Matrix* **D**

- $R^{\text{pixels x images}}$
- a matrix of image vectors

Multilinear: **Data Tensor** $\mathcal{D}$

- $R^{\text{people x views x illums x express x pixels}}$
- N-dimensional matrix
- 28 people, 45 images/person
- 5 views, 3 illuminations, 3 expressions per person





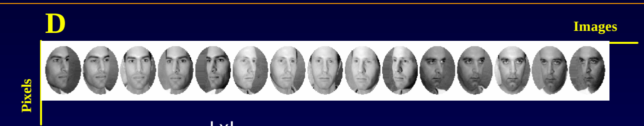
Background on Tensor Decomposition

- Factor Analysis:
  - *Psychometrics, Econometrics, Chemometrics,…*
- SVD:
  - *[Eckart and Young, 1936]* (Psychometrika)

    "The approximation of one matrix by another of lower rank"
- 3-Way Factor Analysis:
  - *[Tucker, 1966]* (Psychometrika)

    "Some mathematical notes on three mode factor analysis"
- N-Way Factor Analysis:
  - *[Harshman, 1970] – Parafac*
  - *[Carrol and Chang, 1970] – Candecomp*
  - *[Kruskal, 1977]*
  - *[Kroonenberg and De Leeuw, 1980]*
  - *[Kapteyn, Neudecker, and Wansbeek, 1986]*
  - *[Franc, 1992]*
  - *[de Lathauwer, 1997]*

Matrix Decomposition - SVD



- A matrix $\mathbf{D} \in \Re^{I_1 \times I_2}$ has a column and row space
- SVD orthogonalizes these spaces and decomposes **D**

$$\mathbf{D} = \mathbf{U}_1 \mathbf{S} \mathbf{U}_2^T$$

($\mathbf{U}_1$ contains the eigenfaces)

- Rewrite in terms of *mode-n products*

$$\mathbf{D} = \mathbf{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2$$

Tensor Decomposition

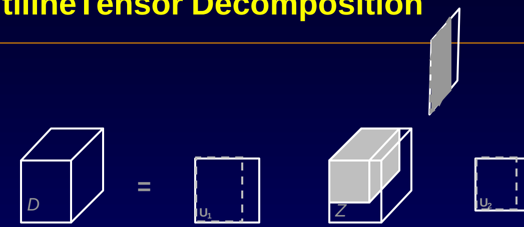
D is a N-dimensional “matrix”, with N spaces

- N-mode SVD is the natural generalization of SVD
- N-mode SVD orthogonalizes these spaces and decomposes D as the mode-n product of N-orthogonal spaces

$$D = Z \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n \cdots \times_N \mathbf{U}_N$$

- Core tensor Z governs interaction between mode matrices
- Mode-n matrix $\mathbf{U}_n$ spans the column space of $\mathbf{D}_{(n)}$

MultilineTensor Decomposition

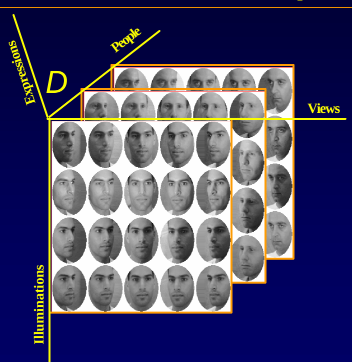


$$D = Z \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$$

$$= \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} s_{r_1 r_2 r_3} \mathbf{u}_{1,r_1} \circ \mathbf{u}_{2,r_2} \circ \mathbf{u}_{3,r_3}$$

$$vec(D) = (\mathbf{U}_3 \otimes \mathbf{U}_2 \otimes \mathbf{U}_1)\, vec(Z)$$

Facial Data Tensor Decomposition



$$D = Z \times_1 \mathbf{U}_{people} \times_2 \mathbf{U}_{views} \times_3 \mathbf{U}_{illums.} \times_4 \mathbf{U}_{express.} \times_5 \mathbf{U}_{pixels}$$

N-Mode SVD Algorithm

Two steps:

1. For n = 1,…,N, compute matrix $\mathbf{U}_n$ 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 as follows

$$Z = D \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_n \mathbf{U}_n^T \cdots \times_N \mathbf{U}_N^T$$

Computing U_{pixels}

- $D_{(pixels)}$ – flatten D along the pixel dimension
- U_{pixels} – orthogonal column space of $D_{(pixels)}$
  – eigenimages

Multilinear (Tensor) Algebra

Nth-order tensor $\mathcal{A} \in \Re^{I_1 \times I_2 \times \cdots I_N}$

matrix (2nd-order tensor) $\mathbf{M} \in \Re^{J_n \times I_n}$

mode-n product:

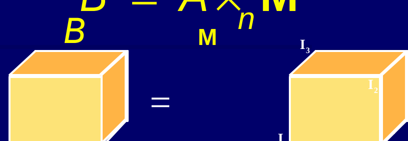
$\mathcal{B} = \mathcal{A} \times_n \mathbf{M}$ where $\mathbf{B}_{(n)} = \mathbf{M}\,\mathbf{A}_{(n)}$

Mode-*n* Product

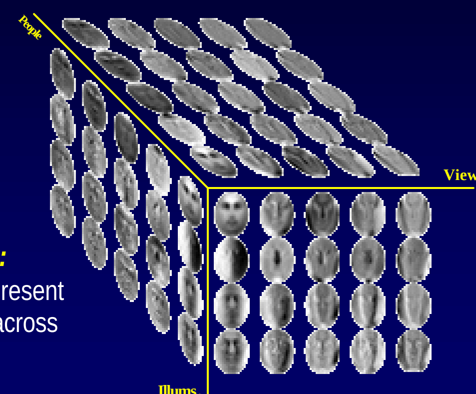
- The mode-n product is a generalization of the product of two matrices
- It is the product of a tensor with a matrix
- Mode-n product of $\mathcal{A} \in \Re^{I_1 \mathrm{x} \ldots \mathrm{x} I_n \mathrm{x} \ldots \mathrm{x} I_N}$ and $\mathbf{M} = \Re^{J_n \mathrm{x} I_n}$
  $\mathcal{B} \in \Re^{I_1 \mathrm{x} \ldots \mathrm{x} I_{n-1} \mathrm{x}\, J_n \mathrm{x} I_{n+1} \mathrm{x} \ldots \mathrm{x} I_N}$

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

$$\mathcal{B} = \mathcal{A} \times_n \mathbf{M}$$

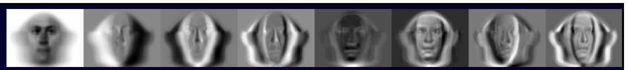


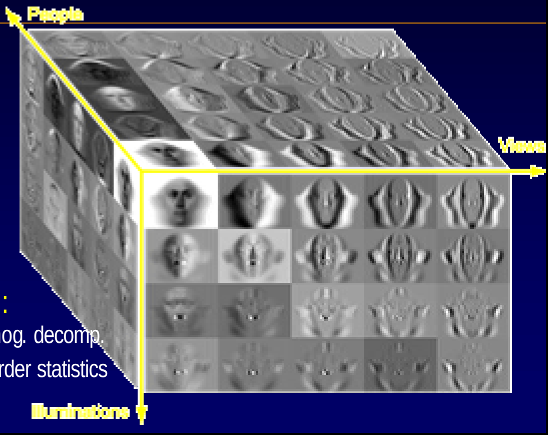
TensorFaces: $\mathcal{B} = \mathcal{Z} \times_5 \mathbf{U}_{pixels}$



TensorFaces:
explicitly represent covariance across factors

PCA:





TensorFaces:

- Multilinear orthog. decomp.
- Encodes 2nd order statistics

TensorFaces Subsume Eigenfaces

Multilinear Analysis / TensorFaces:

$$D = Z \times_1 \mathbf{U}_{\text{people}} \times_2 \mathbf{U}_{\text{views}} \times_3 \mathbf{U}_{\text{illums.}} \times_4 \mathbf{U}_{\text{express.}} \times_5 \mathbf{U}_{\text{pixels}}$$

Linear Analysis / Eigenfaces:

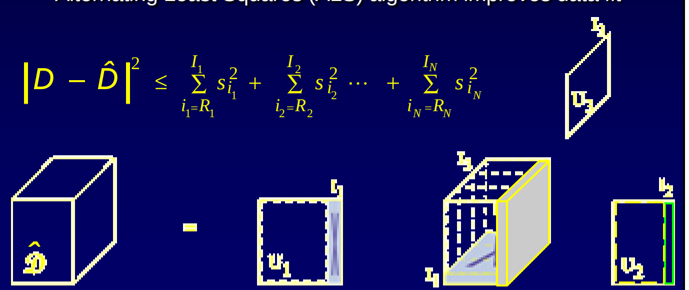
$$\underbrace{\mathbf{D}_{(\text{pixels})}}_{\text{data matrix}} = \underbrace{\mathbf{U}_{\text{pixels}}}_{\text{basis matrix}} \underbrace{\mathbf{Z}_{(\text{pixels})} \left(\mathbf{U}_{\text{express}} \otimes \mathbf{U}_{\text{illums.}} \otimes \mathbf{U}_{\text{views}} \otimes \mathbf{U}_{\text{people}}\right)^T}_{\text{coefficient matrix}}$$

Dimensionality Reduction

Iterative dimensionality reduction approach:

- Optimize mode per mode in an iterative way
- Alternating Least Squares (ALS) algorithm improves data fit

$$\left| D - \hat{D} \right|^2 \leq \sum_{i_1=R_1}^{I_1} s_{i_1}^2 + \sum_{i_2=R_2}^{I_2} s_{i_2}^2 \cdots + \sum_{i_N=R_N}^{I_N} s_{i_N}^2$$

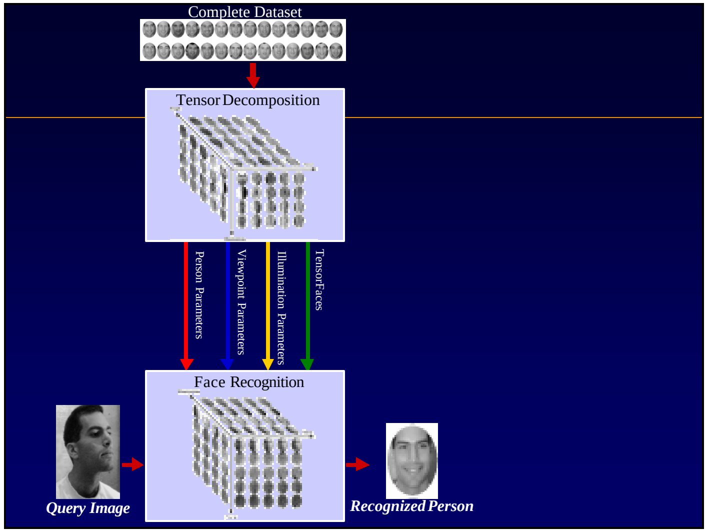


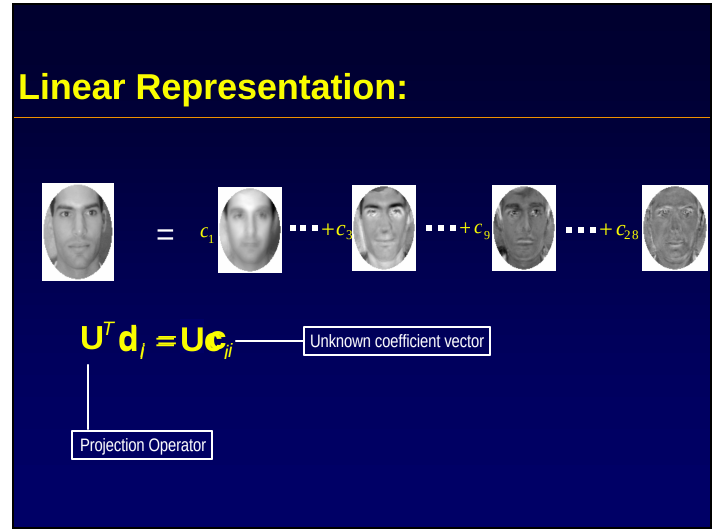
Strategic Data Compression = Perceptual Quality

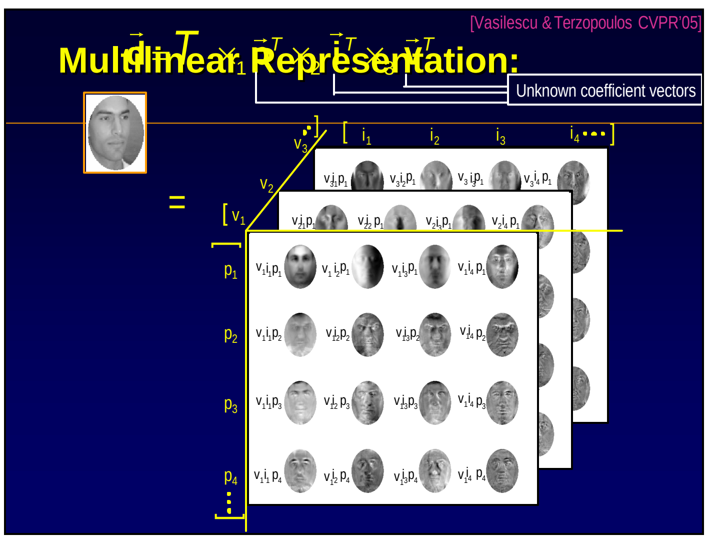
TensorFaces data reduction in illumination space primarily degrades illumination effects (cast shadows, highlights)

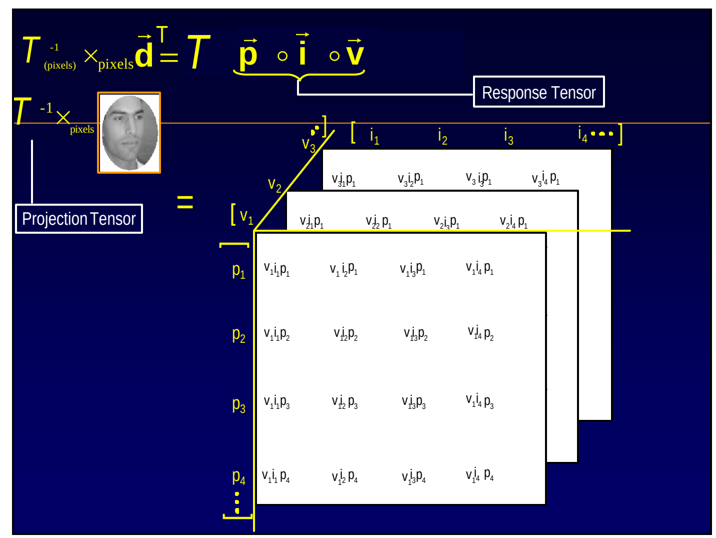
Original	TensorFaces	TensorFaces	PCA
	6 illum + 11 people param.	3 illum + 11 people param.	33 parameters
176 basis vectors	66 basis vectors	33 basis vectors	33 basis vectors

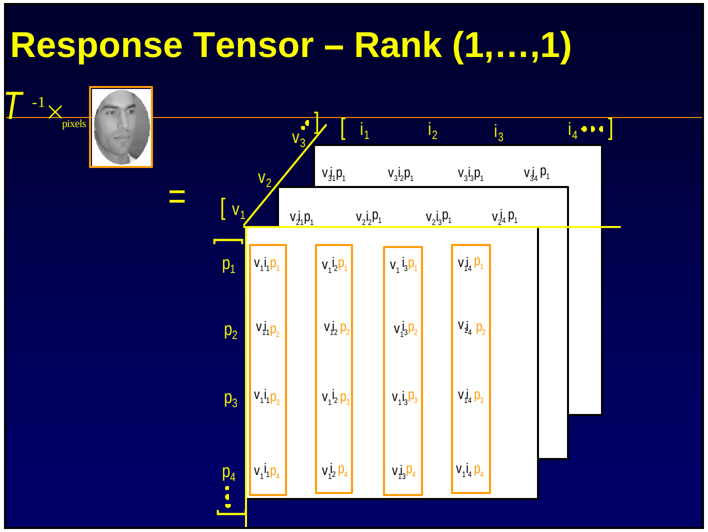
- PCA has *lower mean square error* but *higher perceptual error*

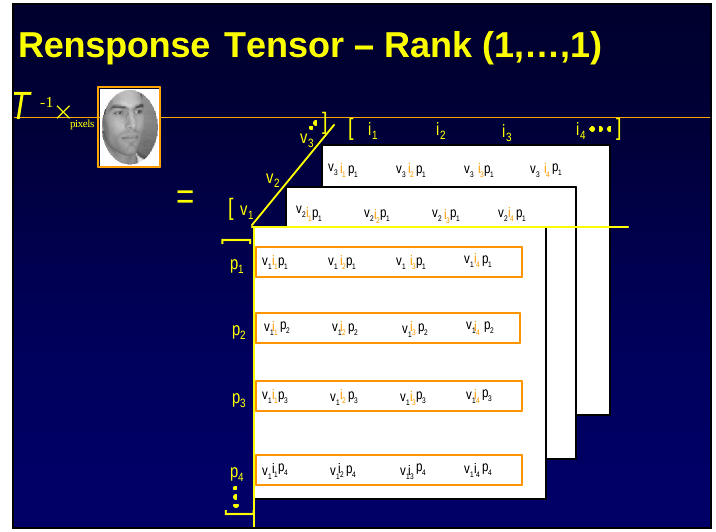

Complete Dataset
Tensor Decomposition
Person Parameters
Viewpoint Parameters
Illumination Parameters
TensorFaces
Face Recognition
Query Image
Recognized Person

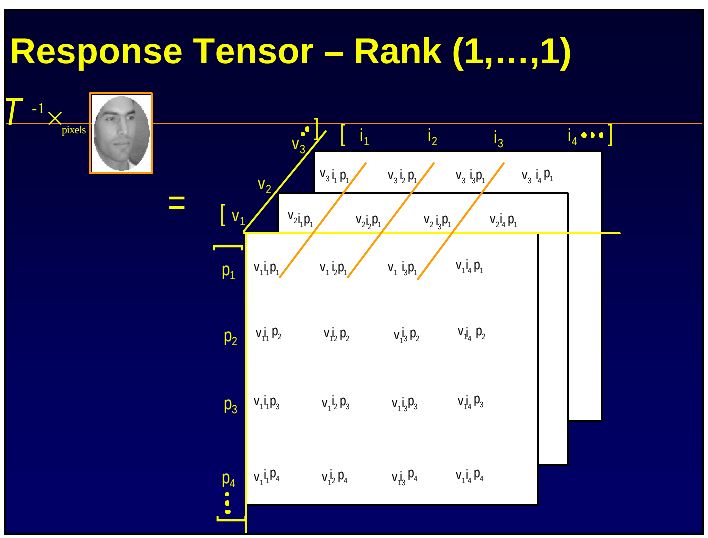

Linear Representation:
$= c_1 \cdots + c_3 \cdots + c_9 \cdots + c_{28}$
$U^T d_i = Uc_i$
Unknown coefficient vector
Projection Operator


[Vasilescu & Terzopoulos CVPR'05]
Multilinear Representation:
Unknown coefficient vectors
=
v1 v2 v3
i1 i2 i3 i4
p1 p2 p3 p4


Response Tensor
Projection Tensor
pixels
=
v1 v2 v3
i1 i2 i3 i4
p1 p2 p3 p4







Multilinear Projection

1. Compute the Projection Tensor or Pseudo-Inverse:

$$\mathbf{P}_{(pixels)} = \mathbf{T}^{+}_{(pixels)}$$

→ observation mode

2. Compute the Response Tensor:

$$\mathcal{P} \times_{pixels} \mathbf{d}^{T} =$$

Mode-Identity Tensor

$$\mathcal{R} = \mathcal{J} \times_{pixels} (\mathbf{I}^{T} \otimes \mathbf{v}^{T} \otimes \mathbf{p}^{T})$$

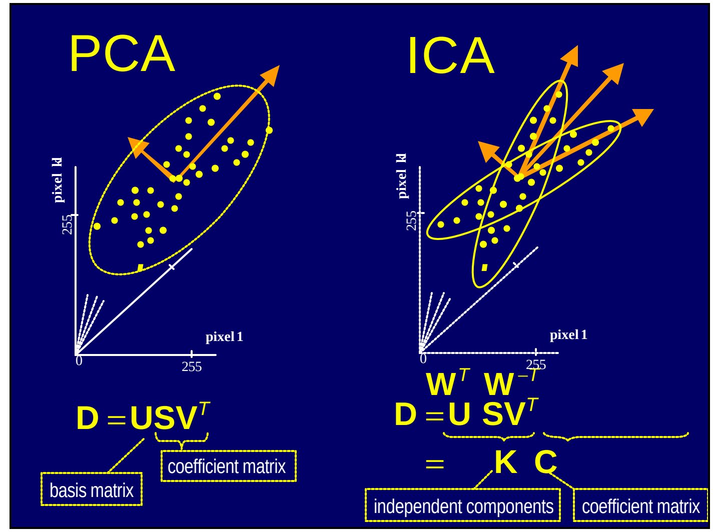
$$= \mathbf{p} \circ \mathbf{v} \circ \mathbf{I}$$

3. Extract the coefficient vectors by factorizing the Response Tensor using the N-mode SVD algorithm

Perspective on Multilinear Models

	Linear Models	*Our Nonlinear (Multilinear) Models*
2nd -Order Statistics ***(covariance)***	**PCA** Eigenfaces	**Multilinear PCA** TensorFaces
Higher-Order Statistics	**ICA**	**Multilinear ICA** Independent TensorFaces

[Vasilescu & Terzopoulos, Learning 2004]



N-Mode ICA

[Vasilescu & Terzopoulos, CVPR 2005]

1. For n=1,…,N, compute matrix $\mathbf{U}_n$ by computing the SVD of the flattened matrix $\mathbf{D}_{(n)}$ and setting $\mathbf{U}_n$ to be the left matrix of the SVD. Compute $\mathbf{W}_n^T$ using ICA. Our new mode matrix is $\mathbf{K}_n$

$$\mathbf{D}_{(n)} = \mathbf{U}_n \mathbf{Z}_{(n)} \mathbf{V}_n^T = \underbrace{\left(\mathbf{U}_n \mathbf{W}_n^T\right)}\mathbf{W}_n^{-T} \mathbf{Z}_{(n)} \mathbf{V}_n^T$$

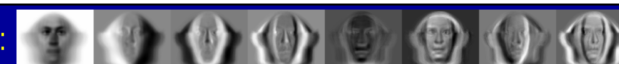
$$= \mathbf{K}_n \quad \mathbf{W}_n^{-T} \mathbf{Z}_{(n)} \mathbf{V}_n^T$$

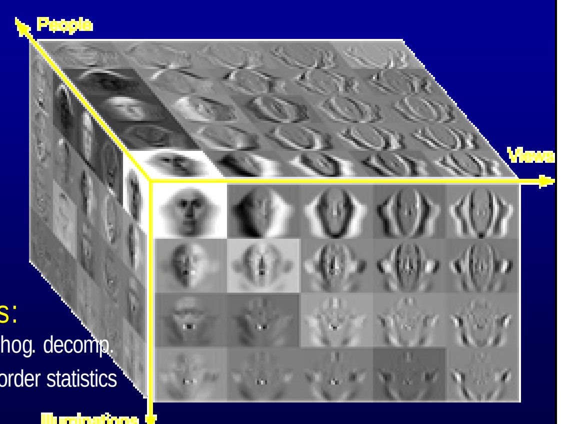
2. Solve for the core tensor as follows

$$\mathcal{S} = \mathcal{D} \times_1 \mathbf{K}_1^{-1} \times_2 \mathbf{K}_2^{-1} \times \cdots \times_n \mathbf{K}_n^{-1} \times \cdots \times_N \mathbf{K}_N^{-1}$$

$$\mathcal{S} = \mathcal{Z} \times_1 \mathbf{W}_1^{-T} \times_2 \mathbf{W}_2^{-T} \times \cdots \times_n \mathbf{W}_n^{-T} \times \cdots \times_N \mathbf{W}_N^{-T}$$

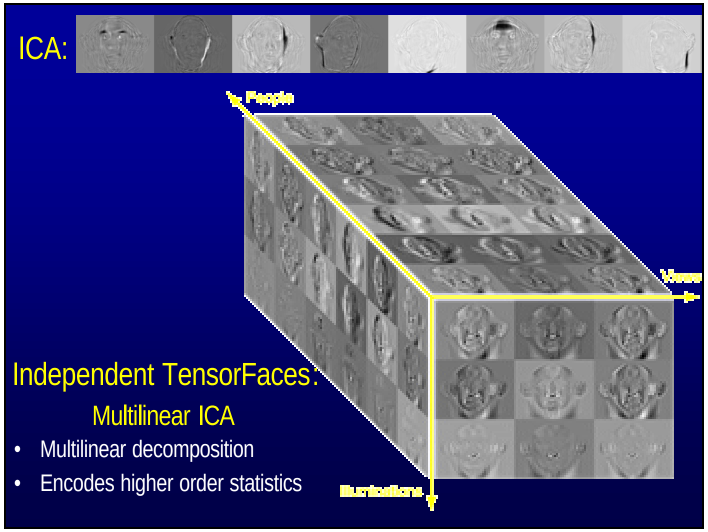
PCA:

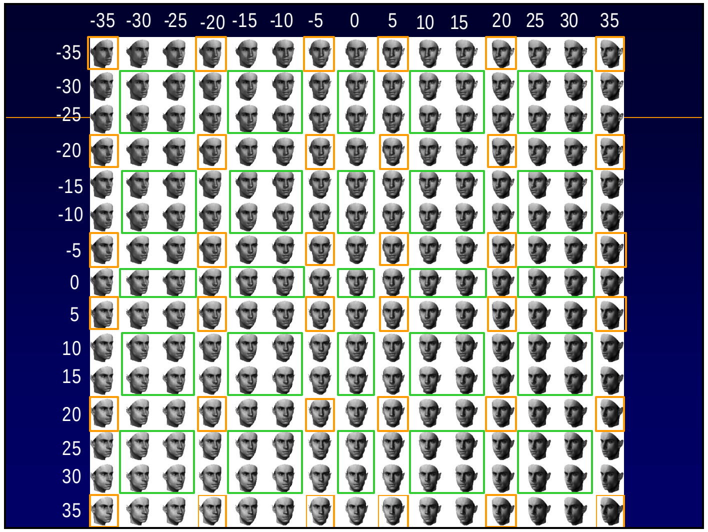




TensorFaces:

- Multilinear orthog. decomp.
- Encodes 2nd order statistics





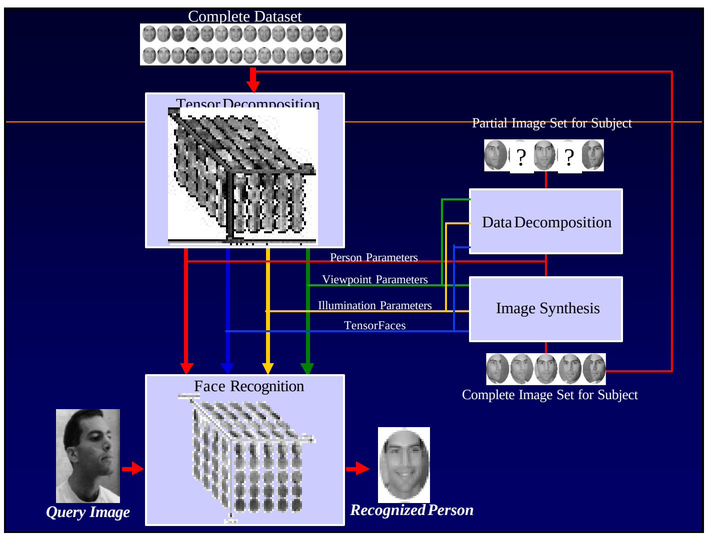
Results

Data Set - 16,875 images
- 75 people
- 15 viewpoints
- 15 illuminations

Training Images - 2,700
- 75 people
- 6 viewpoints
- 6 illuminations

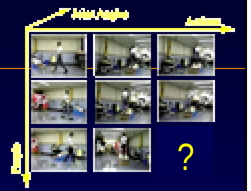
Test Images:
- 75 people
- 9 viewpoints
- 9 illums

Linear Models		***Multilinear Models***	
PCA	***ICA***	***TensorFaces***	***Independent TensorFaces***
83%	***89%***	***93%***	***97%***



Other Applications

- Human Motion Signatures



  - *3-Mode Decomposition, Recognition, & Synthesis*

    [*Vasilescu ICPR 02, CVPR 01, SIGGRAPH 01*]

- Multilinear Image-Based Rendering

  [Vasilescu & Terzopoulos, SIGGRAPH 04]

Conclusion

Multilinear algebraic framework for computer vision and computer graphics

- Tensor approach to the analysis and synthesis of image ensembles
  - *Multilinear PCA and Multilinear ICA*
  - *Multilinear Projection*
  - *Tensor Mode-Identity, Tensor Pseudo-Inverse*
  - *TensorFaces and TensorTextures*
- Applications in many other fields of science