# Component Analysis Methods for Signal Processing

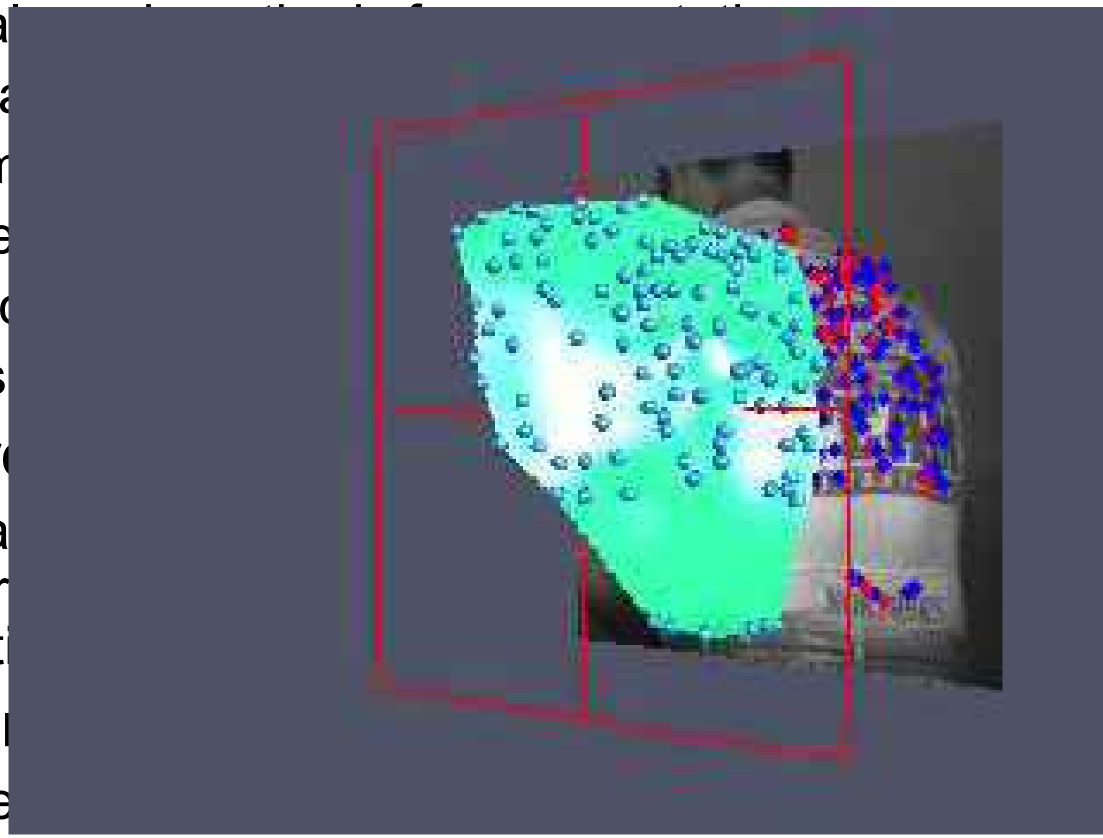
Fernando De la Torre
ftorre@cs.cmu.edu
Carnegie Mellon
THE ROBOTICS INSTITUTE

# Component Analysis for PR

- Computer Vision & Image Processing
  - Structure from motion.
  - Spectral graph methods for segmentation.
  - Appearance and shape models.
  - Fundamental matrix estimation and calibration.
  - Compression.
  - Classification.
  - Dimensionality reduction and visualization.
- Signal Processing
  - Spectral estimation, system identification (e.g. Kalman filter), sensor array processing (e.g. cocktail problem, eco cancellation), blind source separation, …
- Computer Graphics
  - Compression (BRDF), synthesis,…
- Speech, bioinformatics, combinatorial problems.

# Component Analysis for PR

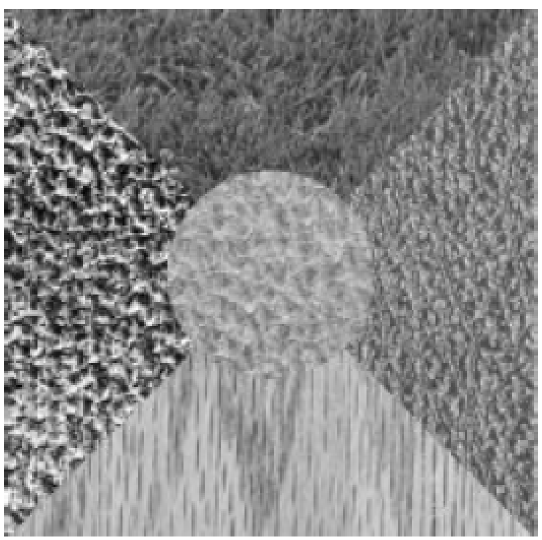
- Computer Vision & Image Processing
  - Structure from motion.
  - Spectra
  - Appeara
  - Fundam
  - Compre
  - Classifi
  - Dimens
- Signal Pr
  - Spectra … filter), sensor array pr … n), blind source separat
- Compute
  - Compre
- Speech, bioinformatics, combinatorial problems.

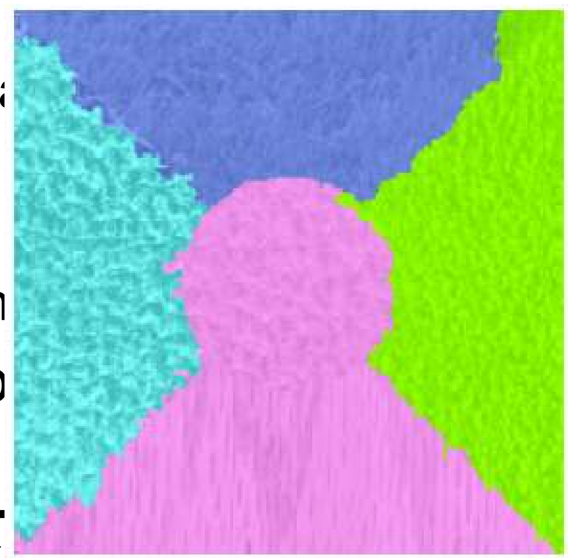


# Component Analysis for PR

- Computer Vision & Image Processing
  - Structure from motion.
  - Spectral graph methods for segmentation.
  - ode
  - atio
  - 
  - 
  - and

- Si
  - Sp n an filter), sensor arr ta tion), blind source sep
- Comp
  - Co h
- Spee o s.





# Component Analysis for PR

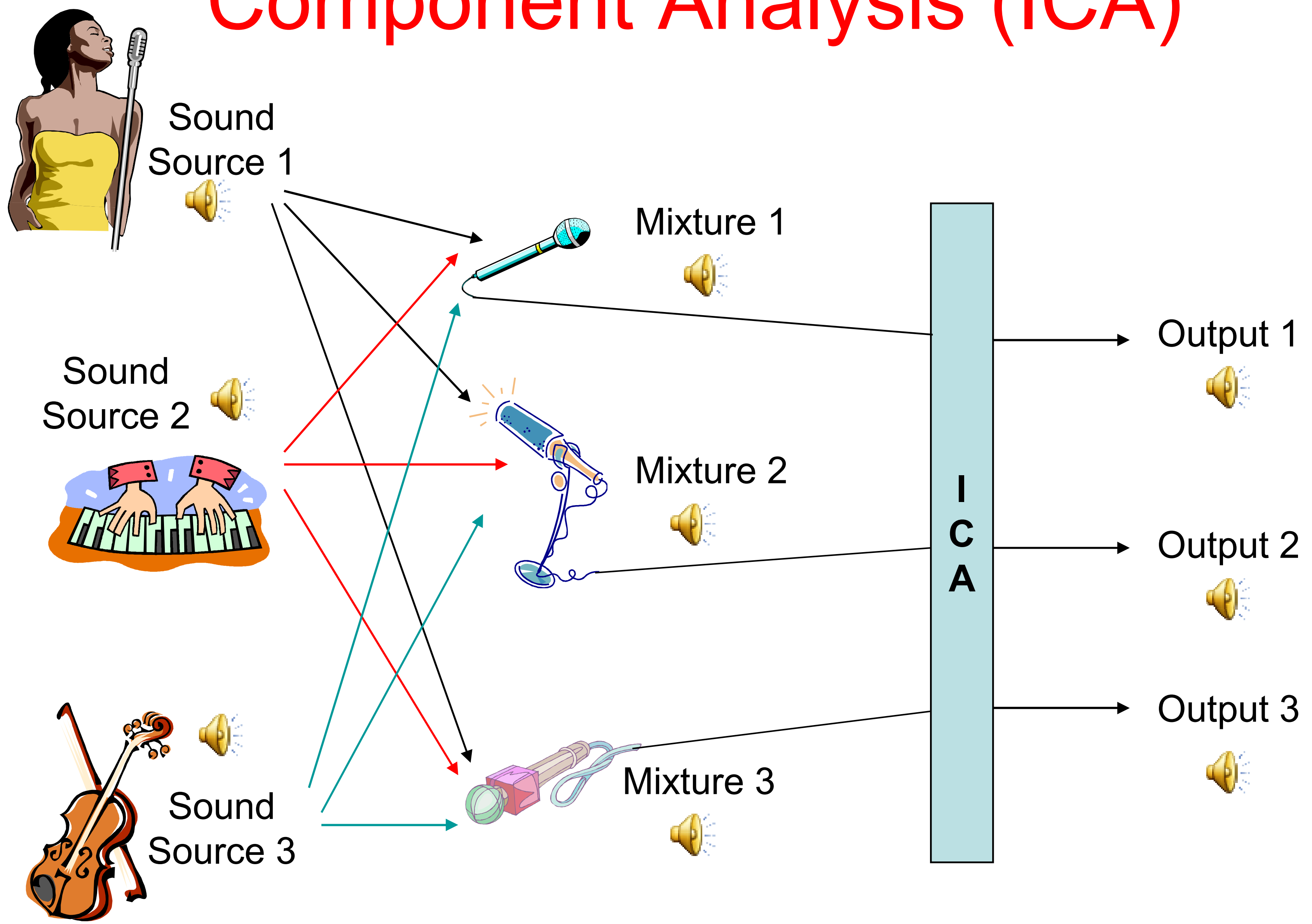
- Computer Vision & Image Processing
  - Structure from motion.
  - Spectral graph methods for segmentation.
  - Appearance and shape models.
  - Fund ration.
  - Com
  - Clas
  - Dime n.

- Signal
  - Spec (e.g. Kalman filter), sensor array eco cancellation), blind source sepa
- Compu
  - Compression (BRDF), synthesis,…
- Speech, bioinformatics, combinatorial problems.

# Component Analysis for PR

- Computer Vision & Image Processing
  - Structure from motion.
  - Spectral graph methods for segmentation.
  - Appearance and shape models.
  - Fundamental matrix estimation and calibration.
  - Compression.
  - Classification.
  - Dimensionality reduction and visualization.
- Signal Processing
  - Spectral estimation, system identification (e.g. Kalman filter), sensor array processing (e.g. cocktail problem, eco cancellation), blind source separation, …
- Computer Graphics
  - Compression (BRDF), synthesis,…
- Speech, bioinformatics, combinatorial problems.

# Component Analysis (ICA)



# Component Analysis for PR

- Computer Vision & Image Processing
  - Structure from motion.
  - Spectral graph methods for segmentation.
  - Appearance and shape models.
  - Fundamental matrix estimation and calibration.
  - Compression.
  - Classification.
  - Dimensionality reduction and visualization.
- Signal Processing
  - Spectral estimation, system identification (e.g. Kalman filter), sensor array processing (e.g. cocktail problem, eco cancellation), blind source separation, …
- Computer Graphics
  - Compression (BRDF), synthesis,…
- Speech, bioinformatics, combinatorial problems.

# Why Component Analysis for PR?

- Learn from high dimensional data and few samples.
  - Useful for dimensionality reduction.

- Easy to incorporate
  - Robustness to noise, missing data, outliers (de la Torre & Black, 2003a)
  - Invariance to geometric transformations (de la Torre & Black, 2003b; de la Torre & Nguyen,2007)
  - Non-linearities (Kernel methods) (Scholkopf & Smola,2002; Shawe-Taylor & Cristianini,2004)
  - Probabilistic (latent variable models) (Everitt,1984)
  - Multi-factorial (tensors) (Paatero & Tapper, 1994 ;O'Leary & Peleg,1983; Vasilescu & Terzopoulos,2002; Vasilescu & Terzopoulos,2003)
  - Exponential family PCA (Gordon,2002; Collins et al. 01)

- Efficient methods $O( d \quad n< <n^2 )$

  d: features

  $n< <n^2$: samples

# Are CA Methods Popular/Useful/Used?

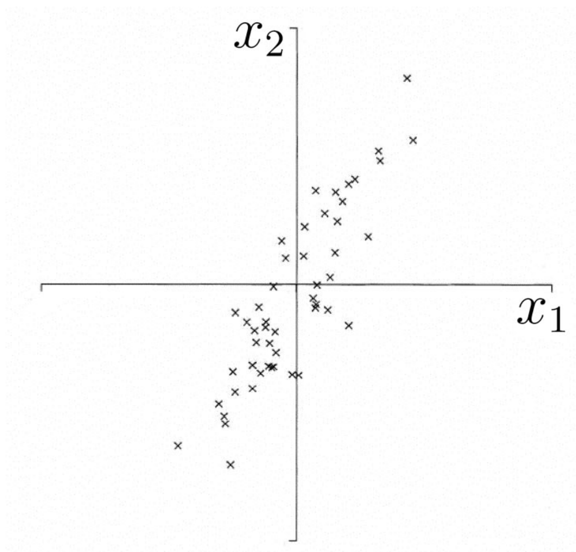
- About 20% of CVPR-06 papers use CA.
- Google:
  - Results **1** - **10** of about **1,870,000** for **"principal component analysis"**.
  - Results **1** - **10** of about **506,000** for **"independent component analysis"**.
  - Results **1** - **10** of about **273,000** for **"linear discriminant analysis"**.
  - Results **1** - **10** of about **46,100** for **"negative matrix factorization"**.
  - Results **1** - **10** of about **491,000** for **"kernel methods"**.
- Still work to do
  - Results **1** - **10** of about **65,300,000** for **"Britney Spears"**.

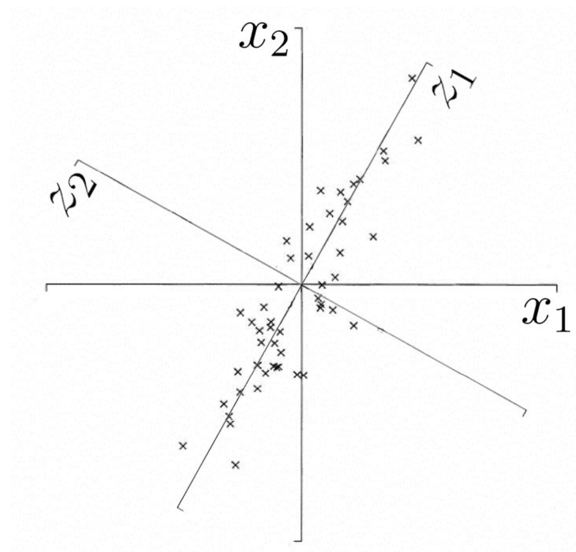
# Outline

- Introduction
- Generative models
  - Principal Component Analysis (PCA)
  - Non-negative Matrix Factorization (NMF)
  - Independent Component Analysis (ICA)
  - Multidimensional Scaling (MDS)
- Discriminative models
  - Linear Discriminant Analysis (LDA).
  - Oriented Component Analysis (OCA).
  - Canonical Correlation Analysis (CCA).
- Standard extensions of linear models
  - Kernel methods.
  - Latent variable models.
  - Tensor factorization

# Principal Component Analysis (PCA)

(Pearson, 1901; Hotelling, 1933;Mardia et al., 1979; Jolliffe, 1986; Diamantaras, 1996)



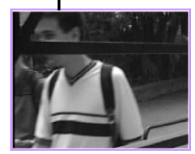

- PCA finds the directions of maximum variation of the data based on linear correlation.
- PCA decorrelates the original variables.

# PCA

d = pixels

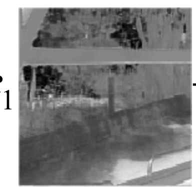
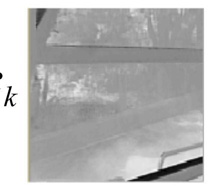
n = images

$$\mathbf{D} = \left[\mathbf{d}_1 \;\; \mathbf{d}_2 \;\; \ldots \;\; \mathbf{d}_n\right] \approx \mathbf{B}\mathbf{C} + \boldsymbol{\mu}\mathbf{1}_n^T$$



$$\mathbf{D} \in \Re^{d\times n} \quad \mathbf{B} \in \Re^{d\times k}$$

$$\mathbf{C} \in \Re^{k\times n} \quad \boldsymbol{\mu} \in \Re^{d\times 1}$$

 $\approx \mu$  $+ c_1$  $+ c_2$  $+ \ldots\ldots + c_k$ 

- Assuming 0 mean data, the basis **B** that preserve the maximum variation of the signal is given by the eigenvectors of **DD**$^T$.

$$\underbrace{\mathbf{D}\mathbf{D}^T}_{d \times d}\mathbf{B} = \mathbf{B}\boldsymbol{\Lambda}$$

# Snap-shot Method & SVD

- If d>>n (e.g. images 100*100 vs. 300 samples) no $\mathbf{DD}^T$.
- $\mathbf{DD}^T$ and $\mathbf{D}^T\mathbf{D}$ have the same eigenvalues (energy) and related eigenvectors (by **D**).
- **B** is a linear combination of the data! (Sirovich, 1987)

$$\mathbf{DD}^T\mathbf{B} = \mathbf{B}\boldsymbol{\Lambda} \quad \mathbf{B} = \mathbf{D}\boldsymbol{\alpha} \quad \mathbf{D}^T\mathbf{DD}^T\mathbf{D}\boldsymbol{\alpha} = \mathbf{D}^T\mathbf{D}\boldsymbol{\alpha}\boldsymbol{\Lambda}$$

- [**α**,**L**]=eig($\mathbf{D}^T\mathbf{D}$) **B**=**D** **α**(diag(diag(**L**))) $^{-0.5}$

- SVD factorizes the data matrix **D** as:

(Beltrami, 1873; Schmidt, 1907; Golub & Loan, 1989)

$$\mathbf{DD}^T = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$$

$$\mathbf{D}^T\mathbf{D} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$$

| PCA | SVD |
| --- | --- |
| $\mathbf{D} = \mathbf{BC}$ | $\mathbf{D} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ |
| $\mathbf{B} \in \Re^{d \times k}$ $\quad \mathbf{C} \in \Re^{k \times n}$ | $\mathbf{U} \in \Re^{d \times k}$ $\quad \boldsymbol{\Sigma} \in \Re^{k \times n}$ $\quad \mathbf{V} \in \Re^{n \times n}$ |
| $\mathbf{B}^T\mathbf{B} = \mathbf{I}$ $\quad \mathbf{CC}^T = \boldsymbol{\Lambda}$ | $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ $\quad \mathbf{V}^T\mathbf{V} = \mathbf{I}$ $\quad \boldsymbol{\Sigma}$ *diagonal* |

# Error Function for PCA

- PCA minimizes the following **CONVEX** function.

(Eckardt & Young, 1936; Gabriel & Zamir, 1979; Baldi & Hornik, 1989; Shum et al., 1995; de la Torre & Black, 2003a)

$$E_1(\mathbf{B},\mathbf{C}) = \sum_{i=1}^{n} \left\| \mathbf{d}_i - \mathbf{B}\mathbf{c}_i \right\|_2^2 = \left\| \mathbf{D} - \mathbf{B}\mathbf{C} \right\|_F$$

- Not unique solution: $\mathbf{B}\mathbf{R}\mathbf{R}^{-1}\mathbf{C} = \mathbf{B}\mathbf{C}$ $\qquad \mathbf{R} \in \Re^{k \times k}$
- To obtain same PCA solution **R** has to satisfy:

$$\hat{\mathbf{B}} = \mathbf{B}\mathbf{R} \qquad \hat{\mathbf{C}} = \mathbf{R}^{-1}\mathbf{C}$$

$$\hat{\mathbf{B}}^T\hat{\mathbf{B}} = \mathbf{I} \qquad \hat{\mathbf{C}}\hat{\mathbf{C}}^T = \Lambda$$

- **R** is computed as a generalized k×k eigenvalue problem.

(de la Torre, 2006)

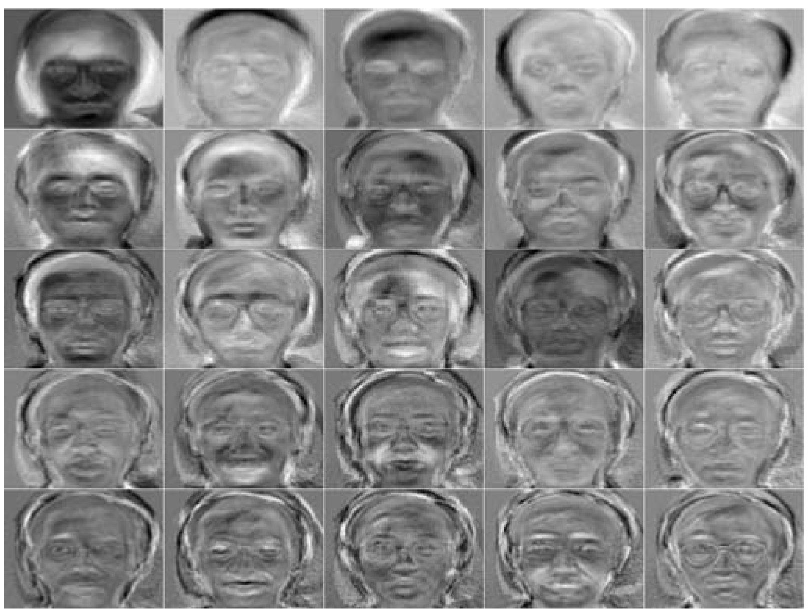
$$\left(\mathbf{C}\mathbf{C}^T\right)^{-1}\mathbf{R} = \mathbf{B}^T\mathbf{B}\mathbf{R}\Lambda^{-1}$$

# PCA/SVD in Computer Vision

- PCA/SVD has been applied to:
  - Recognition (eigenfaces:Turk & Pentland, 1991; Sirovich & Kirby, 1987; Leonardis & Bischof, 2000; Gong et al., 2000; McKenna et al., 1997a)
  - Parameterized motion models (Yacoob & Black, 1999; Black et al., 2000; Black, 1999; Black & Jepson, 1998)
  - Appearance/shape models (Cootes & Taylor, 2001; Cootes et al., 1998; Pentland et al., 1994; Jones & Poggio, 1998; Casia & Sclaroff, 1999; Black & Jepson, 1998; Blanz & Vetter, 1999; Cootes et al., 1995; McKenna et al., 1997; de la Torre et al., 1998b; de la Torre et al., 1998b)
  - Dynamic appearance models (Soatto et al., 2001; Rao, 1997; Orriols & Binefa, 2001; Gong et al., 2000)
  - Structure from Motion (Tomasi & Kanade, 1992; Bregler et al., 2000; Sturm & Triggs, 1996; Brand, 2001)
  - Illumination based reconstruction (Hayakawa, 1994)
  - Visual servoing (Murase & Nayar, 1995; Murase & Nayar, 1994)
  - Visual correspondence (Zhang et al., 1995; Jones & Malik, 1992)
  - Camera motion estimation (Hartley, 1992; Hartley & Zisserman, 2000)
  - Image watermarking (Liu & Tan, 2000)
  - Signal processing (Moonen & de Moor, 1995)
  - Neural approaches (Oja, 1982; Sanger, 1989; Xu, 1993)
  - Bilinear models (Tenenbaum & Freeman, 2000; Marimont & Wandell, 1992)
  - Direct extensions (Welling et al., 2003; Penev & Atick, 1996)

# “Intercorrelations among variables are the bane of the multivariate researcher’s struggle for meaning”

**Cooley and Lohnes, 1971**



# Part-based Representation

- The firing rates of neurons are never negative.
- Independent representations.

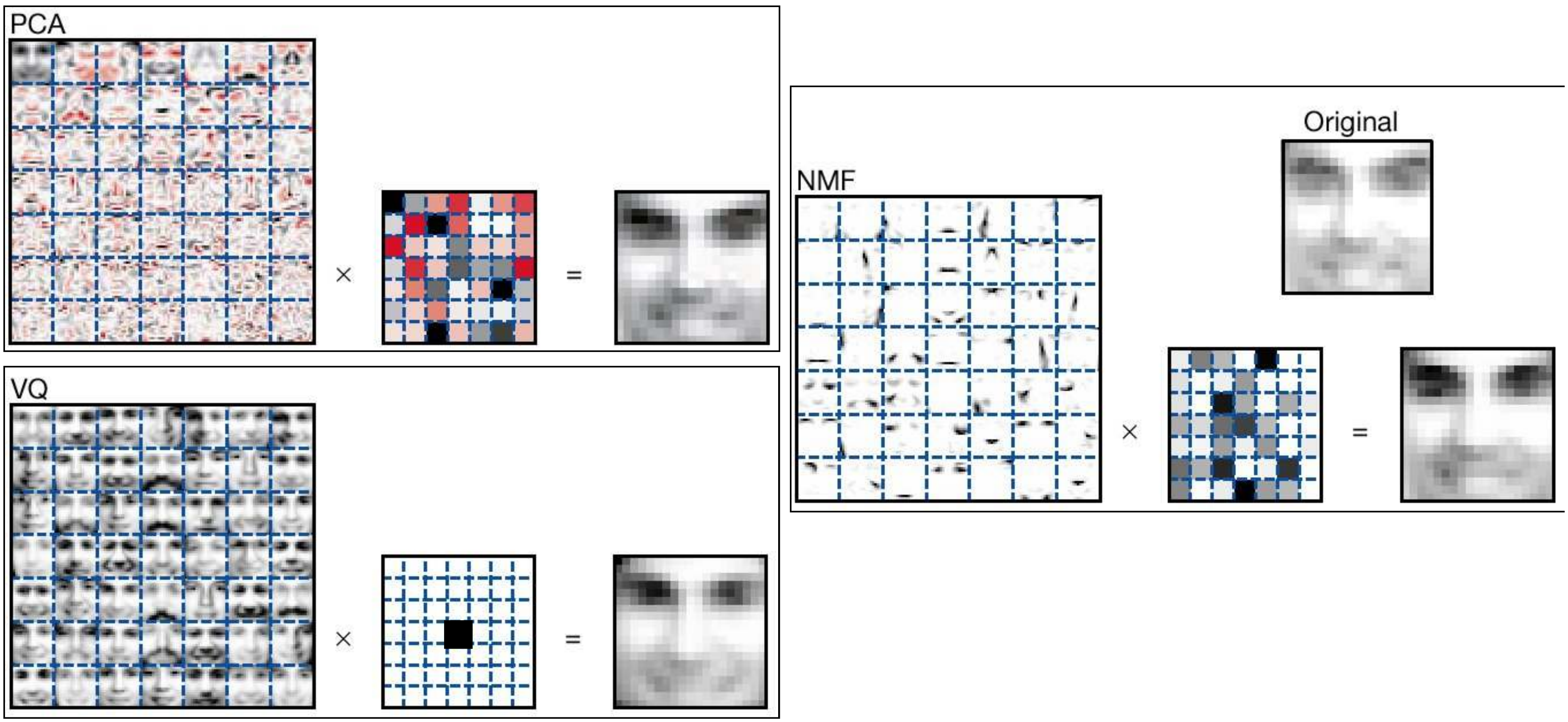
NMF & ICA

# Non-negative Matrix Factorization

- Positive factorization.

$$E(\mathbf{B},\mathbf{C}) = \|\mathbf{D}-\mathbf{BC}\|_F \quad \mathbf{B},\mathbf{C} \geq 0$$

- Leads to part-based representation.



# Nonnegative Factorization

(Lee & Seung, 1999;Lee & Seung, 2000)

$$\min_{\mathbf{B}\geq 0,\mathbf{C}\geq 0} F = \sum_{ij} \left| d_{ij} - (\mathbf{BC})_{ij} \right|^2$$

Inference:

$$\mathbf{C}_{ij} \leftarrow \mathbf{C}_{ij} \frac{(\mathbf{B}^{\mathrm{T}}\mathbf{D})_{ij}}{(\mathbf{B}^{\mathrm{T}}\mathbf{BV})_{ij}}$$

Derivatives:

$$\frac{\partial F}{\partial \mathbf{C}_{ij}} = (\mathbf{B}^{\mathrm{T}}\mathbf{BC})_{ij} - (\mathbf{B}^{\mathrm{T}}\mathbf{C})_{ij}$$

Learning:

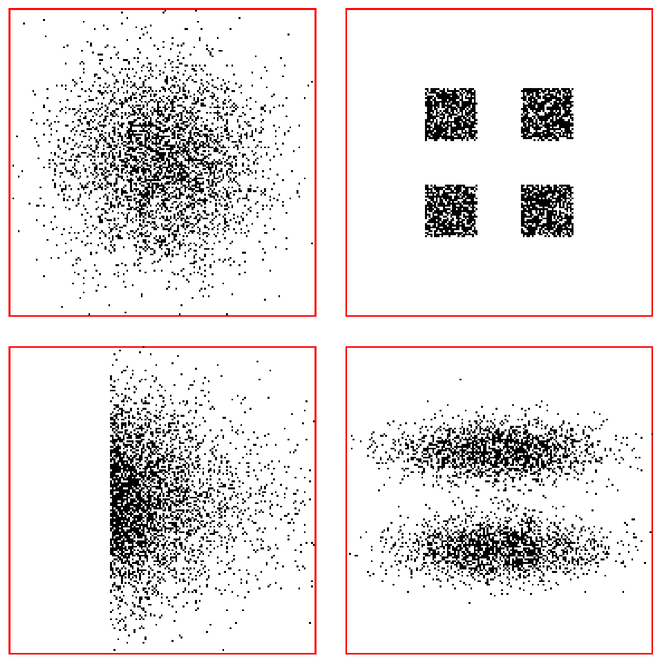
$$\mathbf{B}_{ij} \leftarrow \mathbf{B}_{ij} \frac{(\mathbf{DC}^{T})_{ij}}{(\mathbf{BC}\mathbf{C}^{T})_{ij}}$$

$$\frac{\partial F}{\partial \mathbf{B}_{ij}} = (\mathbf{BCC}^{T})_{ij} - (\mathbf{DC}^{T})_{ij}$$

- Multiplicative algorithm can be interpreted as diagonally rescaled gradient descent.

# Independent Component Analysis

- We need more than second order statistics to represent the signal.

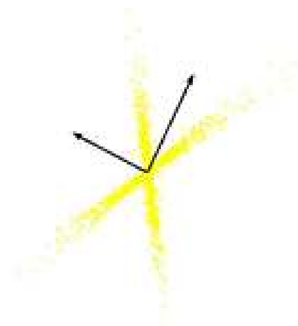


# ICA

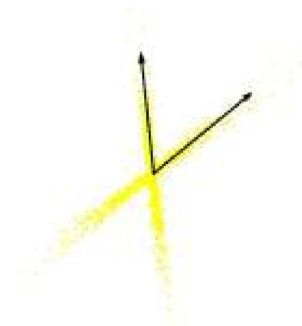
(Hyvrinen et al., 2001)

$$\mathbf{D} = \mathbf{BC} \quad \mathbf{C} \approx \mathbf{S} = \mathbf{WD} \quad \mathbf{W} \approx \mathbf{B}^{-1}$$

- Look for $s_i$ that are independent.
- PCA finds uncorrelated variables, the independent components have non Gaussian distributions.
- Uncorrelated $E(s_i s_j) = E(s_i)E(s_j)$
- Independent $E(g(s_i)f(s_j)) = E(g(s_i))E(f(s_j))$ for any non-linear f,g

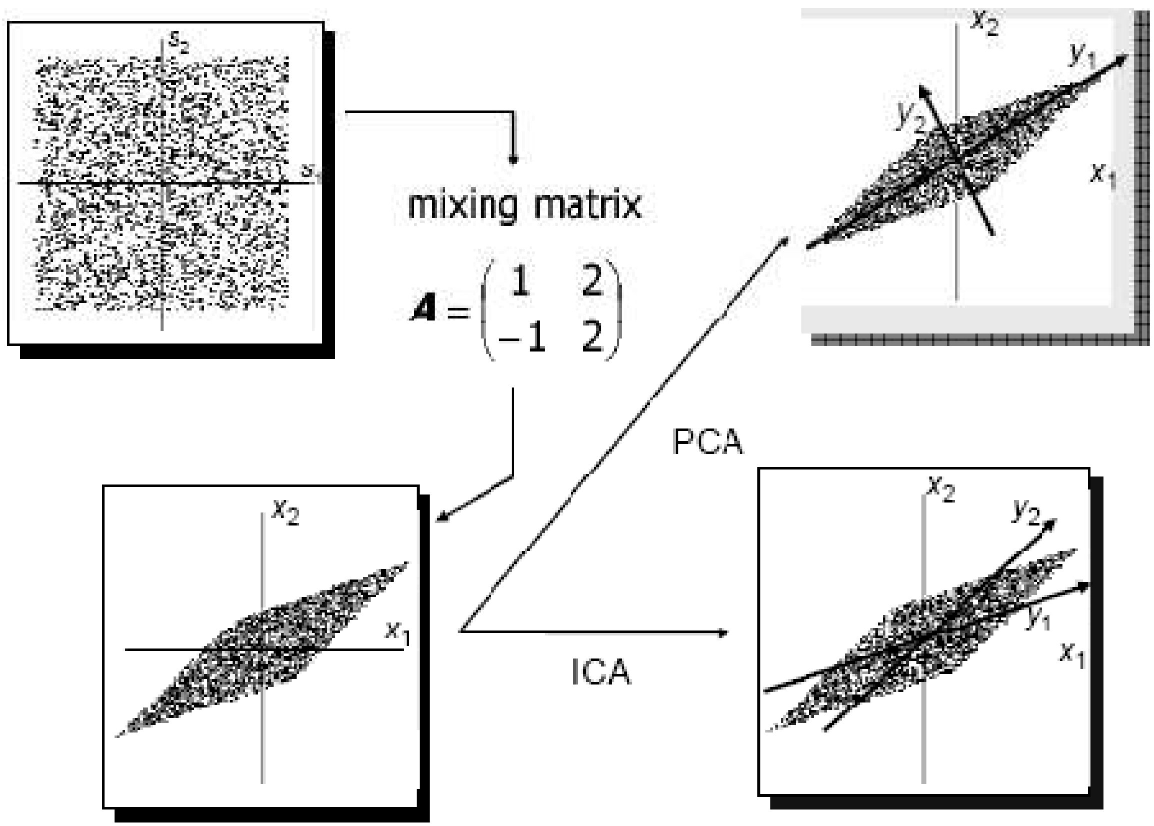


PCA



ICA

# ICA vs PCA



$s_2$
$s_1$
mixing matrix
$A = \begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix}$
$x_2$
$x_1$
PCA
$x_2$
$y_1$
$y_2$
$x_1$
ICA
$x_2$
$y_2$
$y_1$
$x_1$

# Many optimization criteria

- Minimize high order moments: e.g. kurtosis

  kurt($\boldsymbol{W}$) = $E\{s^4\} - 3(E\{s^2\})^2$
- Many other information criteria.
- Also an error function: (Olhausen & Field, 1996)

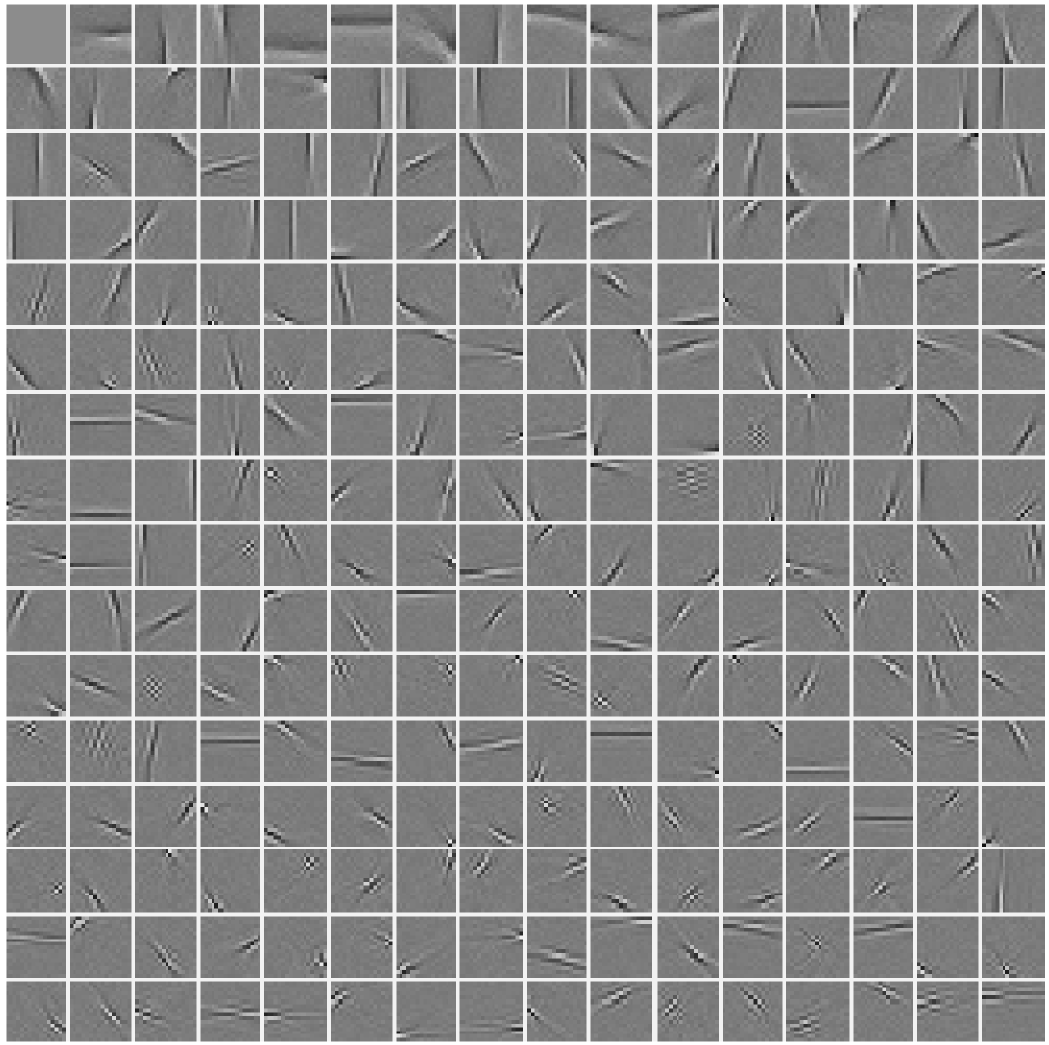
$$\sum_{i=1}^{n} \left\| \mathbf{d}_i - \mathbf{B}\mathbf{c}_i \right\| + \sum_{i=1}^{n} S(\mathbf{c}_i)$$

Sparseness (e.g. S=| |)

- Other sparse PCA.

  (Chennubhotla & Jepson, 2001b; Zou et al., 2005; dAspremont et al., 2004;)

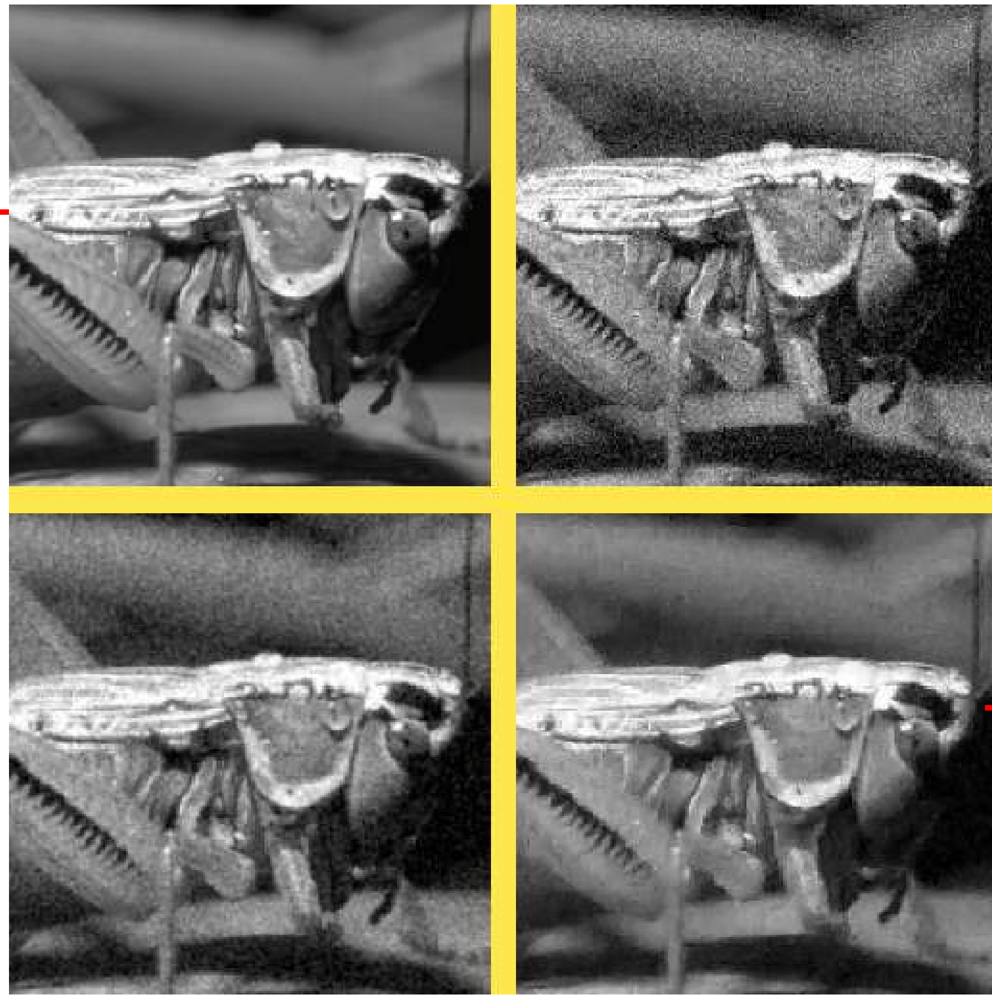
# Basis of natural images



# Denoising

Original image

Noisy Image (30% noise)



Denoise (Wiener filter)

ICA

# Multidimensional Scaling (MDS)

- MDS takes a matrix of pair-wise distances and finds an embedding that preserves the interpoint distances.

## An example: map of the US

| | Chicago | Raleigh | Boston | Seattle | S.F. | Austin | Orlando |
|---|---|---|---|---|---|---|---|
| Chicago | 0 | | | | | | |
| Raleigh | 641 | 0 | | | | | |
| Boston | 851 | 608 | 0 | | | | |
| Seattle | 1733 | 2363 | 2488 | 0 | | | |
| S.F. | 1855 | 2406 | 2696 | 684 | 0 | | |
| Austin | 972 | 1167 | 1691 | 1764 | 1495 | 0 | |
| Orlando | 994 | 520 | 1105 | 2565 | 2458 | 1015 | 0 |

# MDS(II)

Optimize w.r.t $y_i$

$$\sum_i \sum_j (\delta_{ij} - d_{ij})^2$$

Observed distance between points i and j in p-space

Distance between the points in two-dimensional space

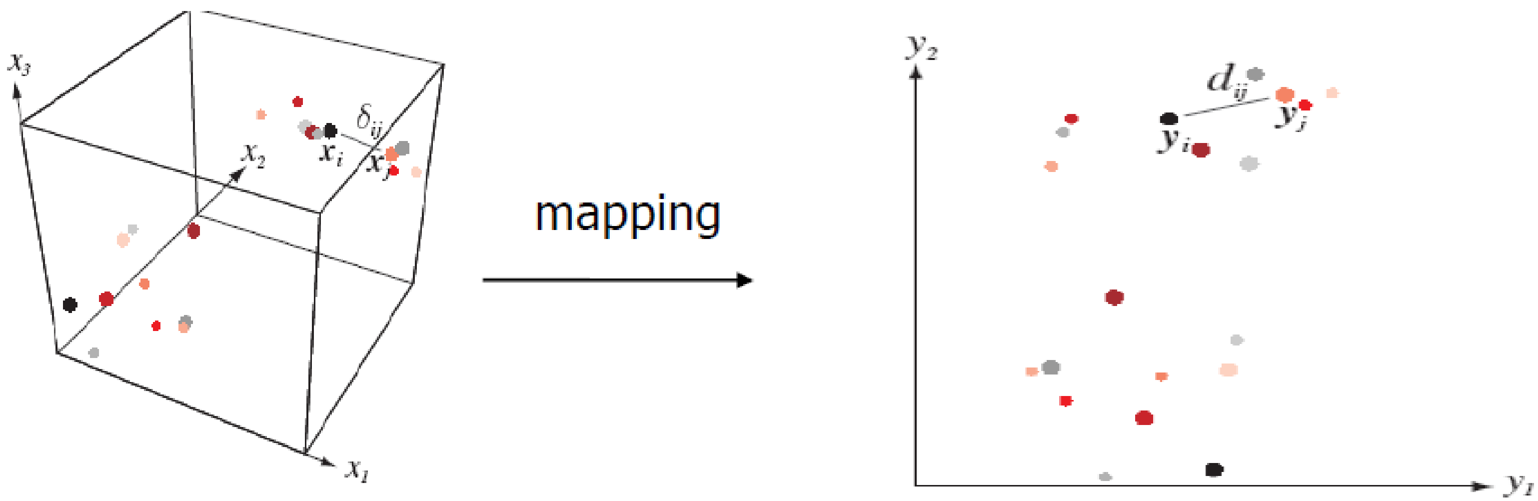
Criterion is invariant wrt rotations and translations.
However it is not invariant to scaling
Better criterion is

$$\frac{\sum_i \sum_j (\delta_{ij} - d_{ij})^2}{\sum_{i,j} d_{ij}^{\ 2}}$$

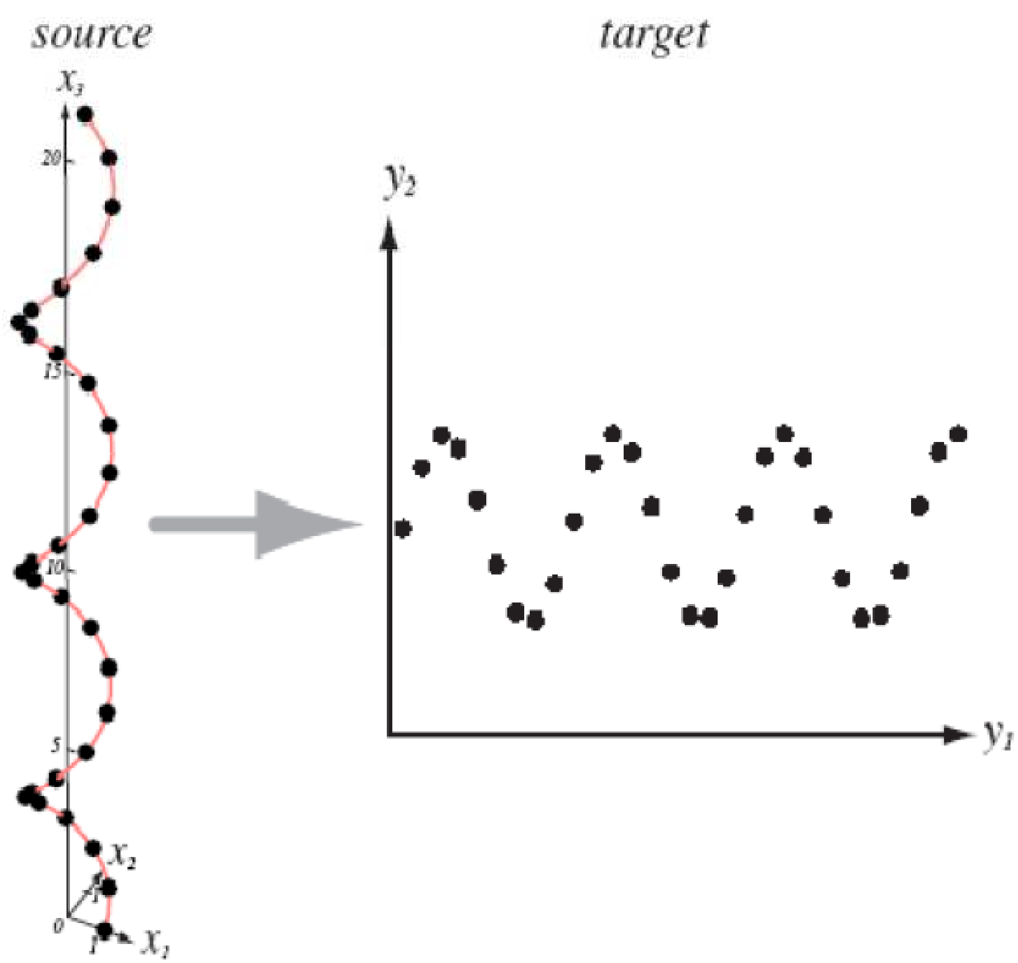
or

$$\sqrt{\frac{\sum_i \sum_j (\delta_{ij} - d_{ij})^2}{\sum_{i,j} d_{ij}^{\ 2}}}$$

Called stress



# MDS (III)

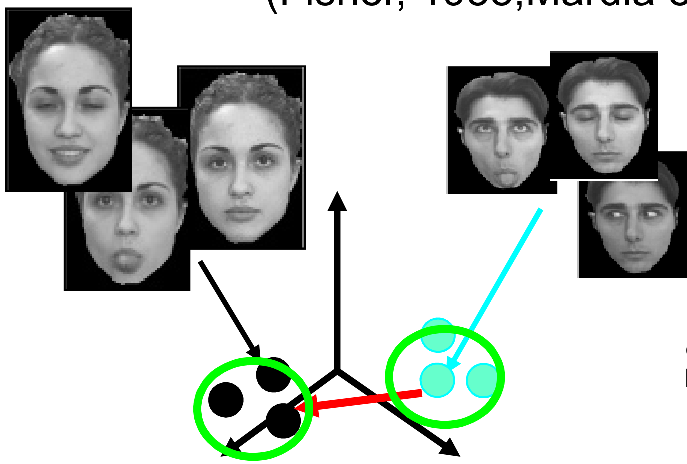


# Outline

- Introduction
- Generative models
  - Principal Component Analysis (PCA)
  - Non-negative Matrix Factorization (NMF)
  - Independent Component Analysis (ICA)
  - Multidimensional Scaling (MDS)
- Discriminative models
  - Linear Discriminant Analysis (LDA).
  - Oriented Component Analysis (OCA).
  - Canonical Correlation Analysis (CCA).
- Standard extensions of linear models
  - Kernel methods.
  - Latent variable models.
  - Tensor factorization

# Linear Discriminant Analysis (LDA)

(Fisher, 1938;Mardia et al., 1979; Bishop, 1995)



$$\mathbf{S}_b = \sum_{i=1}^{C}\sum_{j=1}^{C}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T$$

$$\mathbf{S}_t = \mathbf{D}\mathbf{D}^T = \sum_{i=1}^{n}\mathbf{d}_i\mathbf{d}_i^T$$

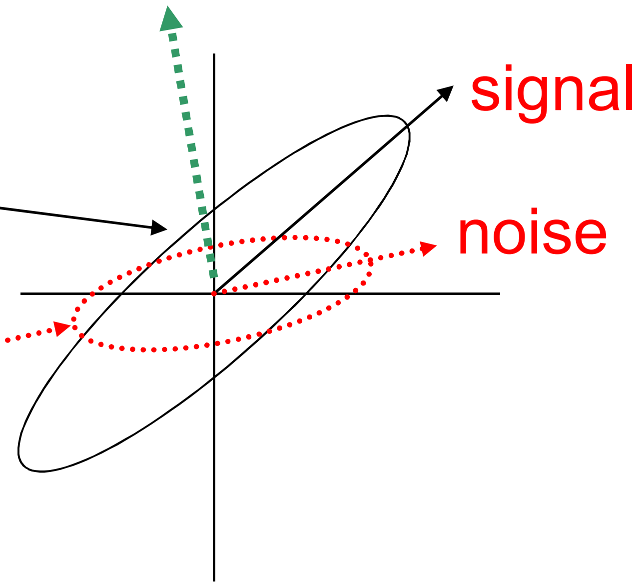
$$J(\mathbf{B}) = \frac{|\mathbf{B}^T\mathbf{S}_\mathbf{b}\mathbf{B}|}{|\mathbf{B}^T\mathbf{S}_t\mathbf{B}|} \qquad \mathbf{S}_\mathbf{b}\mathbf{B} = \mathbf{S}_t\mathbf{B}\boldsymbol{\Lambda}$$

$$\mathbf{S}_w = \sum_{j=1}^{c}\sum_{i=1}^{C_i}(\mathbf{d}_i - \boldsymbol{\mu}_j)(\mathbf{d}_i - \boldsymbol{\mu}_j)^T$$

- Optimal linear dimensionality reduction if classes are Gaussian with equal covariance matrix.

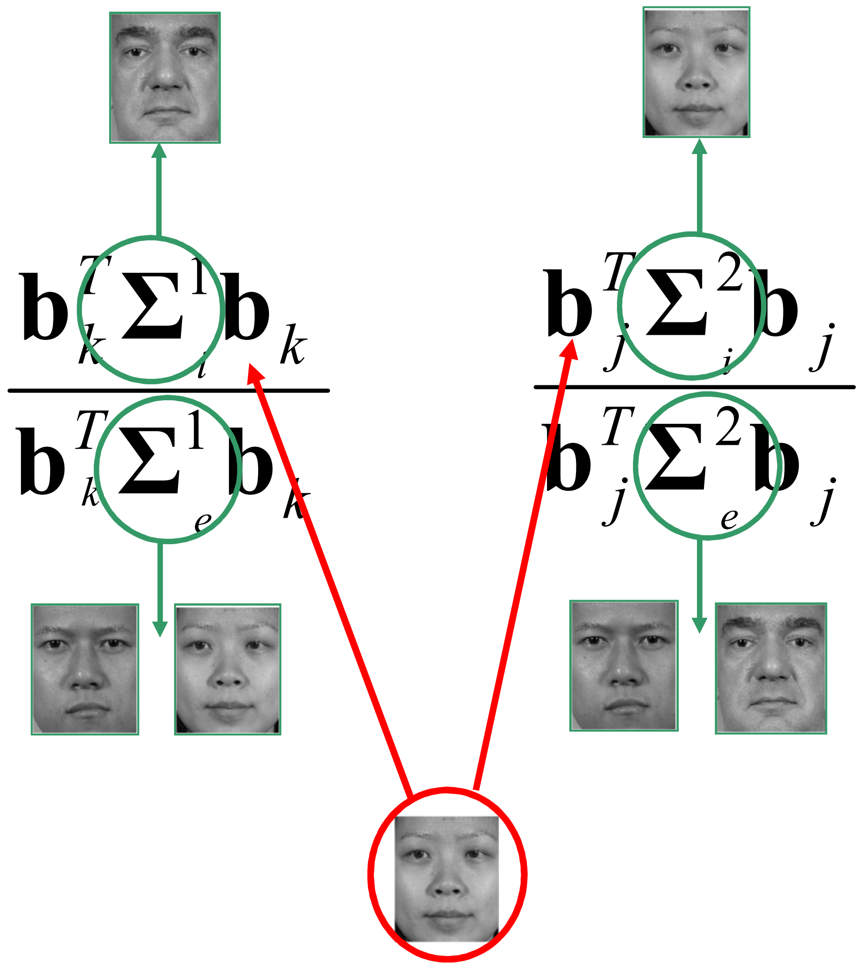
# Oriented Component Analysis (OCA)

$$\frac{\mathbf{b}_{OCA}^{T}\,\boldsymbol{\Sigma}_{signal}\,\mathbf{b}_{OCA}}{\mathbf{b}_{OCA}^{T}\,\boldsymbol{\Sigma}_{noise}\,\mathbf{b}_{OCA}}$$



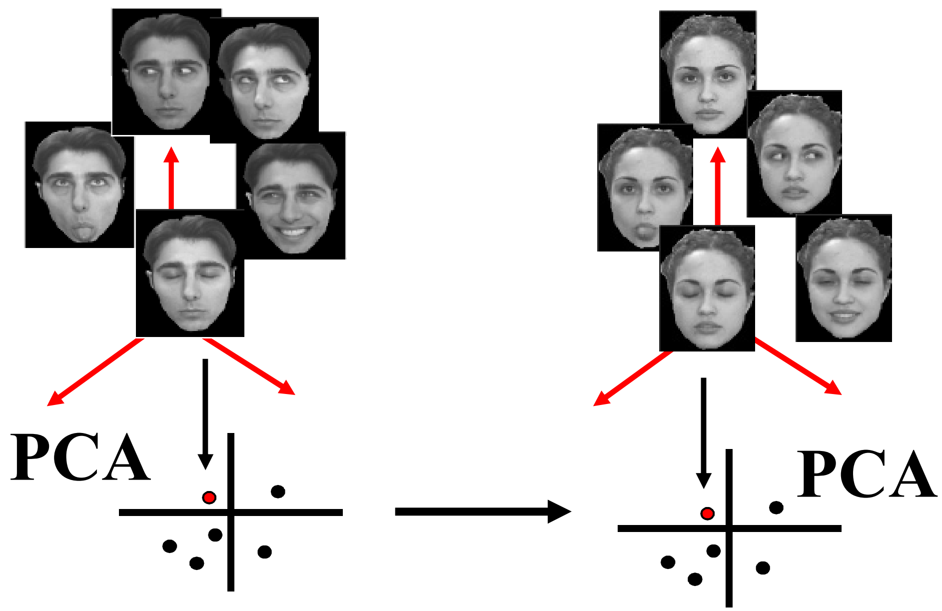

- Generalized eigenvalue problem: $\boldsymbol{\Sigma}_i \mathbf{b}_k = \boldsymbol{\Sigma}_e \mathbf{b}_k \lambda$
- $\mathbf{b}_{oca}$ is steered by the distribution of noise.

# OCA for face recognition



# Canonical Correlation Analysis (CCA)

- PCA independently and general mapping




- Signals dependent signals with small energy can be lost.

# Canonical Correlation Analysis (CCA)

(Mardia et al., 1979; Borga)

- Learn relations between multiple data sets? (e.g. find features in one set related to another data set)
- Given two sets $\mathbf{X} \in \Re^{d_1 \times n}$ *and* $\mathbf{Y} \in \Re^{d_2 \times n}$, CCA finds the pair of directions $\mathbf{w_x}$ and $\mathbf{w_y}$ that maximize the correlation between the projections (assume zero mean data)

$$\rho = \frac{\mathbf{w}_x^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{X}^T \mathbf{X} \mathbf{w}_x^T \mathbf{w}_y^T \mathbf{Y}^T \mathbf{Y} \mathbf{w}_y^T}}$$

- Several ways of optimizing it:

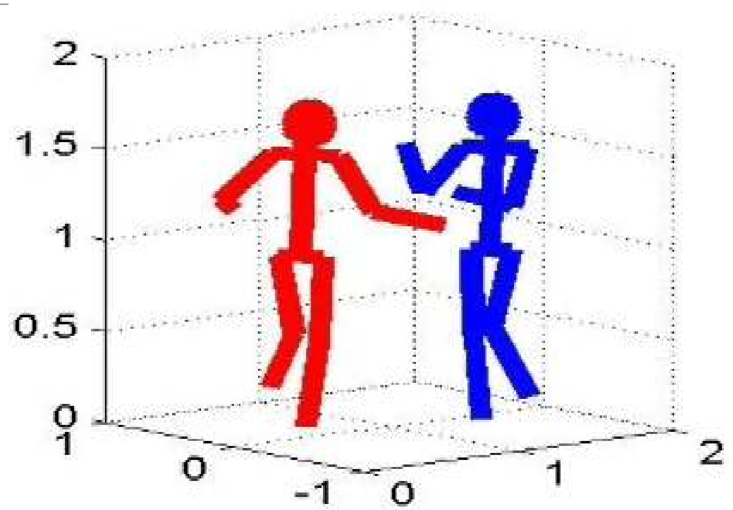
$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{X}^T\mathbf{Y} \\ \mathbf{X}^T\mathbf{Y} & \mathbf{0} \end{bmatrix} \in \Re^{(d_1+d_2)\times(d_1+d_2)}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}^T\mathbf{Y} \end{bmatrix} \in \Re^{(d_1+d_2)\times(d_1+d_2)} \quad \mathbf{w} = \begin{bmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{bmatrix}$$

- An stationary point of r is the solution to CCA.

$$\mathbf{Aw} = \lambda \mathbf{Bw}$$

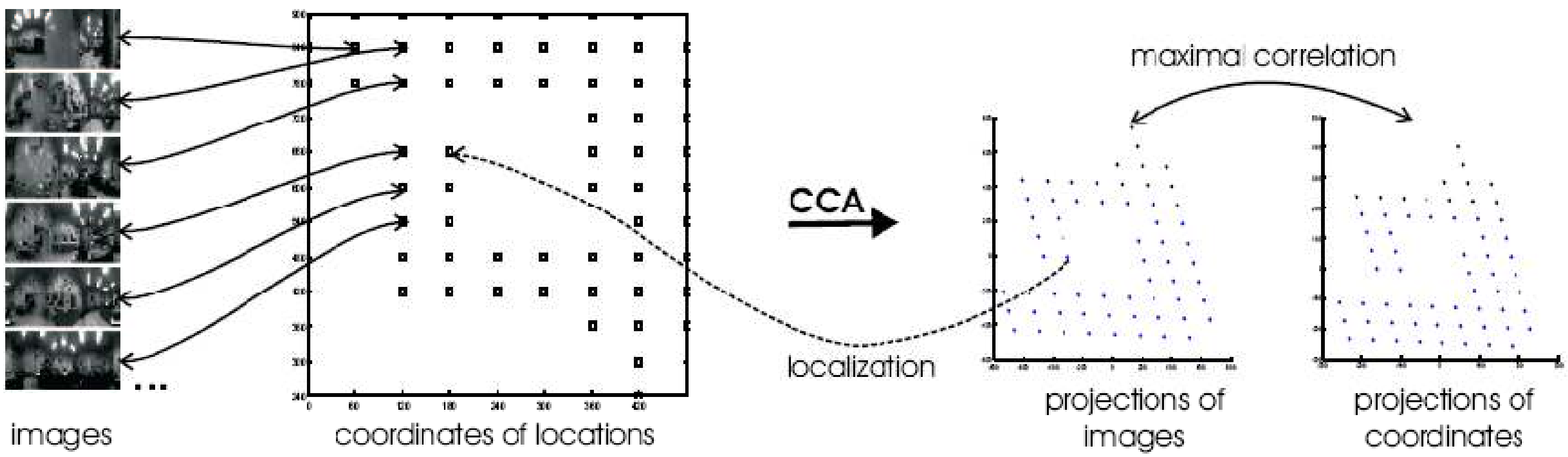
# Dynamic Coupled Component Analysis





# Robot localization with Canonical Correlation Analysis

(Skocaj & Leonardis, 2000)



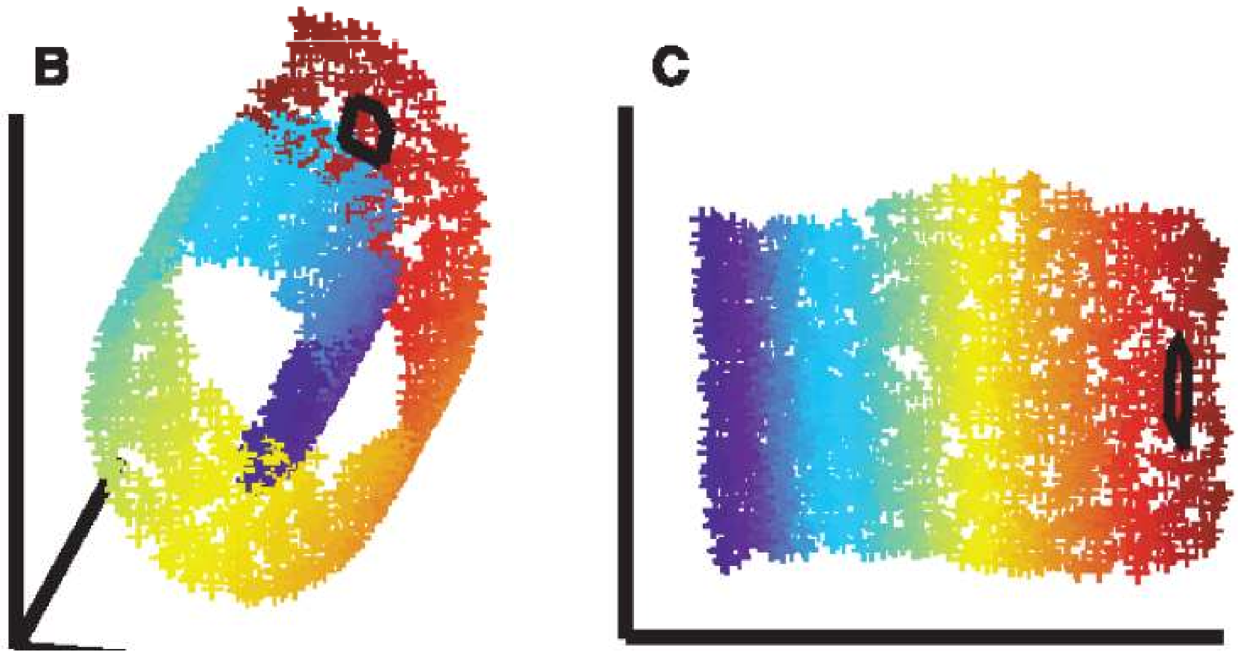
# Outline

- Introduction
- Generative models
  - Principal Component Analysis (PCA)
  - Non-negative Matrix Factorization (NMF)
  - Independent Component Analysis (ICA)
  - Multidimensional Scaling (MDS)
- Discriminative models
  - Linear Discriminant Analysis (LDA).
  - Oriented Component Analysis (OCA).
  - Canonical Correlation Analysis (CCA).
- Standard extensions of linear models
  - Kernel methods.
  - Latent variable models.
  - Tensor factorization

# Kernel Methods

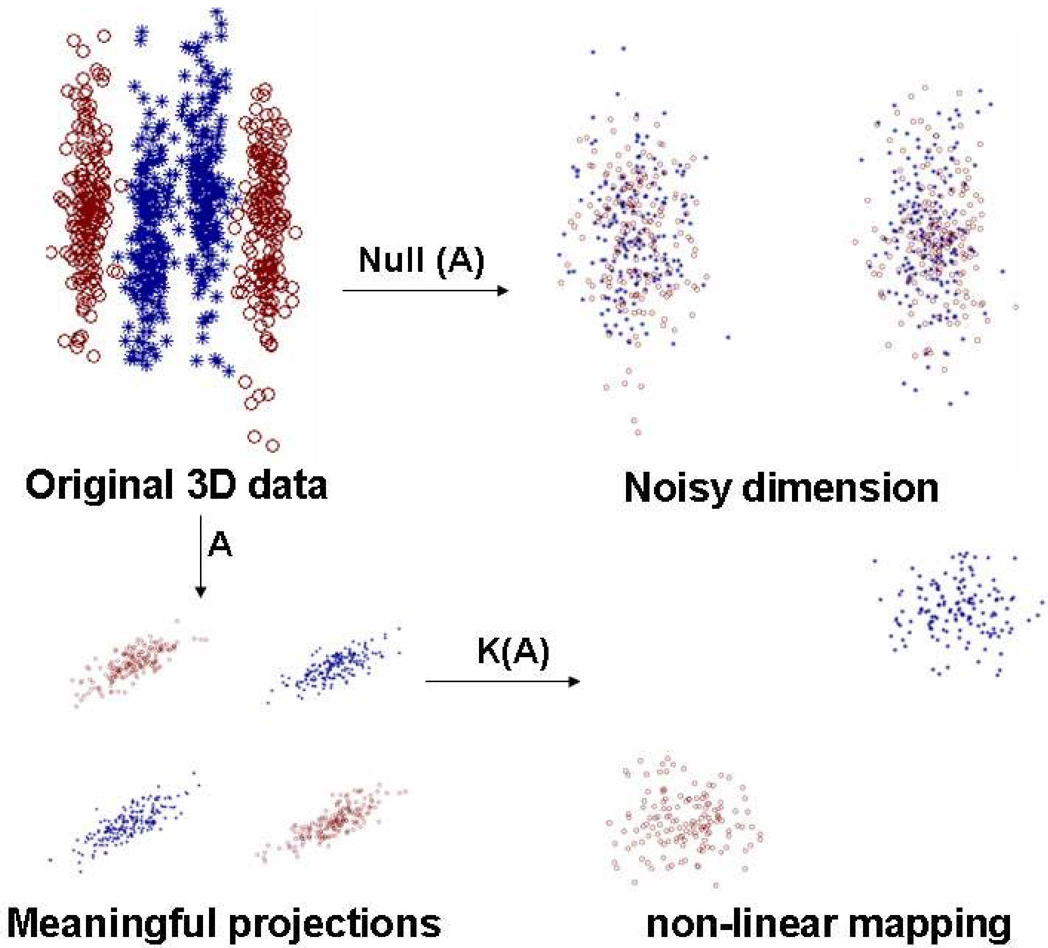
# Linear methods fail

- When data points sit on a non-linear manifold
  - We won't find a good **linear** mapping from the data points to a plane, because there isn't any
  - In the end, linear methods do nothing more than rotate/translate/scale data

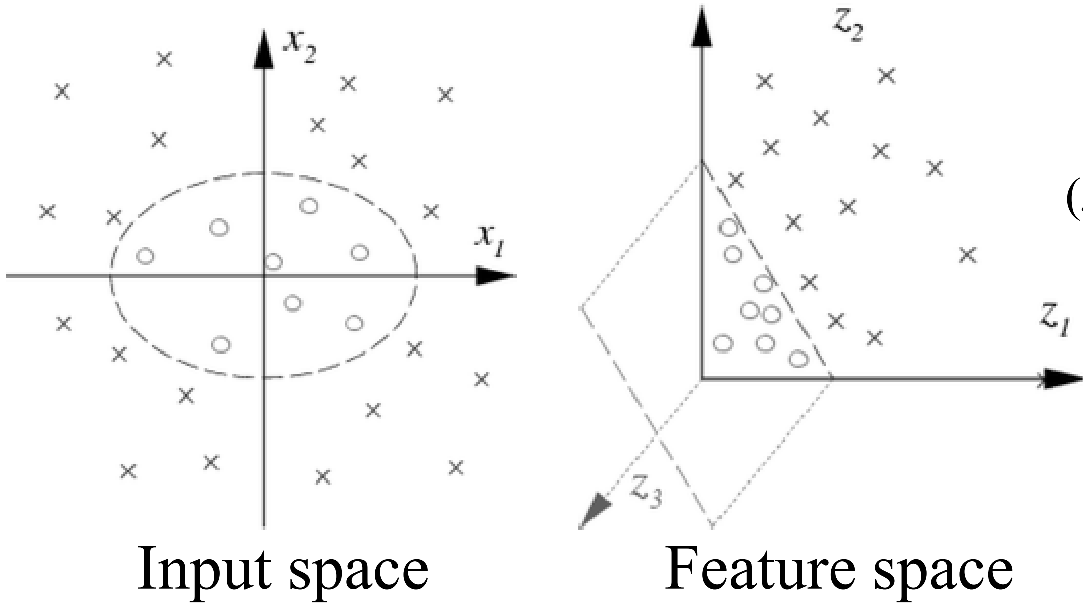


# Linear methods fail

- Learning a non-linear representation for classification



# Kernel Methods for Classification




$$(x_1, x_2) \rightarrow (x_1, x_2, x_1^2 + x_2^2) = (z_1, z_2, z_3)$$

- The kernel defines an implicit mapping (usually high dimensional and non-linear) from input to feature space, so the data becomes linearly separable.
- Computation in the feature space can be costly because it is (usually) high dimensional
  - The feature space is typically infinite-dimensional!

# Kernel Methods

- Suppose $\phi(.)$ is given as follows

$$\phi(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

- An inner product in the feature space is

$$\langle \phi(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}), \phi(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}) \rangle = (1 + x_1y_1 + x_2y_2)^2$$

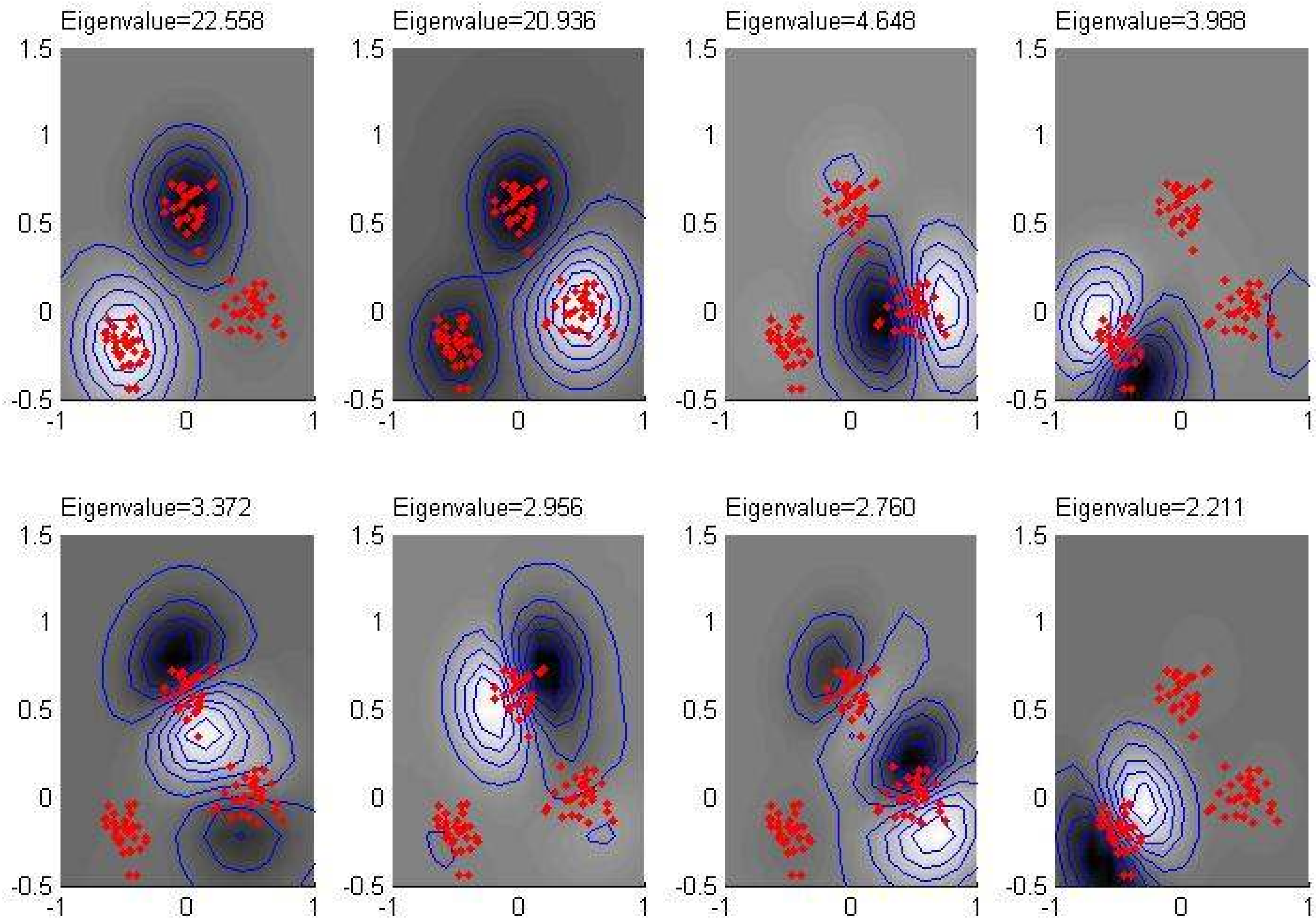
- So, if we define the kernel function as follows, there is no need to carry out $\phi(.)$ explicitly

$$K(\mathbf{x}, \mathbf{y}) = (1 + x_1y_1 + x_2y_2)^2$$

- This use of kernel function to avoid carrying out $\phi(.)$ explicitly is known as the kernel trick. In any linear algorithm that can be expressed by inner products can be made nonlinear by going to the feature space

# Kernel PCA

(Scholkopf et al., 1998)



# Kernel PCA

(Scholkopf et al., 1998)

- Eigenvectors of the cov. Matrix in feature space.

$$\overline{\mathbf{C}} = \frac{1}{n}\sum_{i=1}^{n}\Phi(\mathbf{d}_i)\Phi(\mathbf{d}_i)^{\mathrm{T}} \qquad \overline{\mathbf{C}}\,\mathbf{b}_1 = \mathbf{b}_1\lambda$$

- Eigenvectors lie in the span of data in feature space.

$$\mathbf{b}_1 = \sum_{i=1}^{n}\alpha_i\Phi(\mathbf{d}_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\Phi(\mathbf{d}_i)K(\mathbf{d}_i,\mathbf{d}_j) = [\sum_{i=1}^{n}\alpha_i\Phi(\mathbf{d}_i)]\lambda$$

$$\mathbf{K}\boldsymbol{\alpha} = \boldsymbol{\alpha}\lambda$$

# Latent Variable Models

# Factor Analysis

- A Gaussian distribution on the coefficients and noise is added to PCA➔ Factor Analysis. (Mardia et al., 1979)

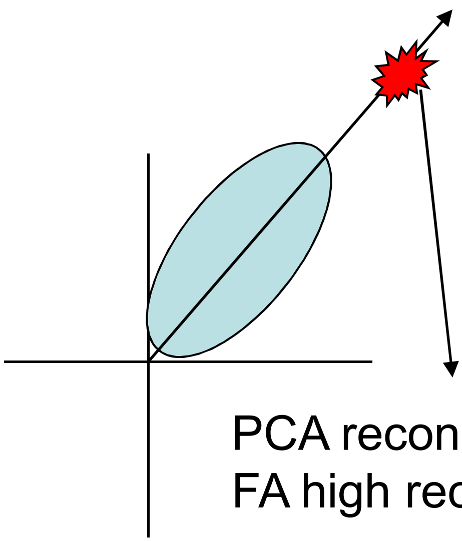
$$\mathbf{d} = \boldsymbol{\mu} + \mathbf{B}\mathbf{c} + \boldsymbol{\eta}$$

$$p(\mathbf{c}) = N(\mathbf{c} \mid \mathbf{0}, \mathbf{I}_k) \qquad p(\mathbf{d} \mid \mathbf{c}, \mathbf{B}) = N(\mathbf{d} \mid \boldsymbol{\mu} + \mathbf{B}\mathbf{c}, \Psi)$$

$$p(\boldsymbol{\eta}) = N(\mathbf{c} \mid \mathbf{0}, \Psi) \qquad \Psi = diag(\eta_1, \eta_2, \ldots, \eta_d)$$

$$\mathrm{cov}(\mathbf{d}) = E((\mathbf{d} - \boldsymbol{\mu})(\mathbf{d} - \boldsymbol{\mu})^T) = \mathbf{B}\mathbf{B}^T + \Psi$$

- Inference (Roweis & Ghahramani, 1999;Tipping & Bishop, 1999a)



$p(\mathbf{c}, \mathbf{d})$ Jointly Gaussian

$$p(\mathbf{c} \mid \mathbf{d}) = N(\mathbf{c} \mid \mathbf{m}, \mathbf{V})$$

$$\mathbf{m} = \mathbf{B}^T (\mathbf{B}\mathbf{B}^T + \Psi)^{-1} (\mathbf{d} - \boldsymbol{\mu})$$

$$\mathbf{V} = (\mathbf{I} + \mathbf{B}^T \Psi^{-1} \mathbf{B})^{-1}$$

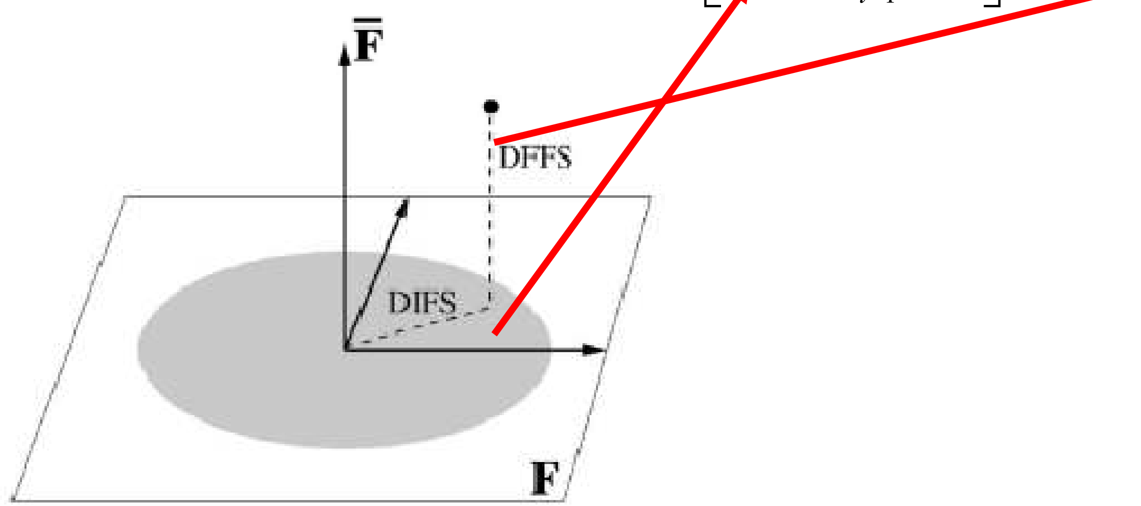
PCA reconstruction low error.
FA high reconstruction error (low likelihood).

# Ppca

- If $\Psi = E(\boldsymbol{\eta}\boldsymbol{\eta}^T) = \varepsilon \mathbf{I}_d$ PPCA.
- If $\varepsilon \to 0$ is equivalent to PCA. $\varepsilon \to 0 \quad \mathbf{B}^T(\mathbf{B}\mathbf{B}^T + \Psi)^{-1} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$
- Probabilistic visual learning (Moghaddam & Pentland, 1997;)

$$p(\mathbf{d}) = \int p(\mathbf{d}\,|\,\mathbf{c})\,p(\mathbf{c})\,d\mathbf{c} = \frac{e^{-\frac{1}{2}(\mathbf{d}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{d}-\boldsymbol{\mu})}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}(\mathbf{d}-\boldsymbol{\mu})^T (\mathbf{B}\mathbf{B}^\mathbf{T}+\varepsilon\mathbf{I})^{-1}(\mathbf{d}-\boldsymbol{\mu})}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} = \left[\frac{e^{-\frac{1}{2}\sum_{i=1}^{k}\frac{c_i^2}{\lambda_i}}}{(2\pi)^{\frac{d}{2}}\prod_{i=1}^{k}\lambda_i^{1/2}}\right]\left[\frac{e^{-\frac{\varepsilon^2(\mathbf{d})}{2\rho}}}{(2\pi\rho)^{\frac{(d-k)}{2}}}\right]$$

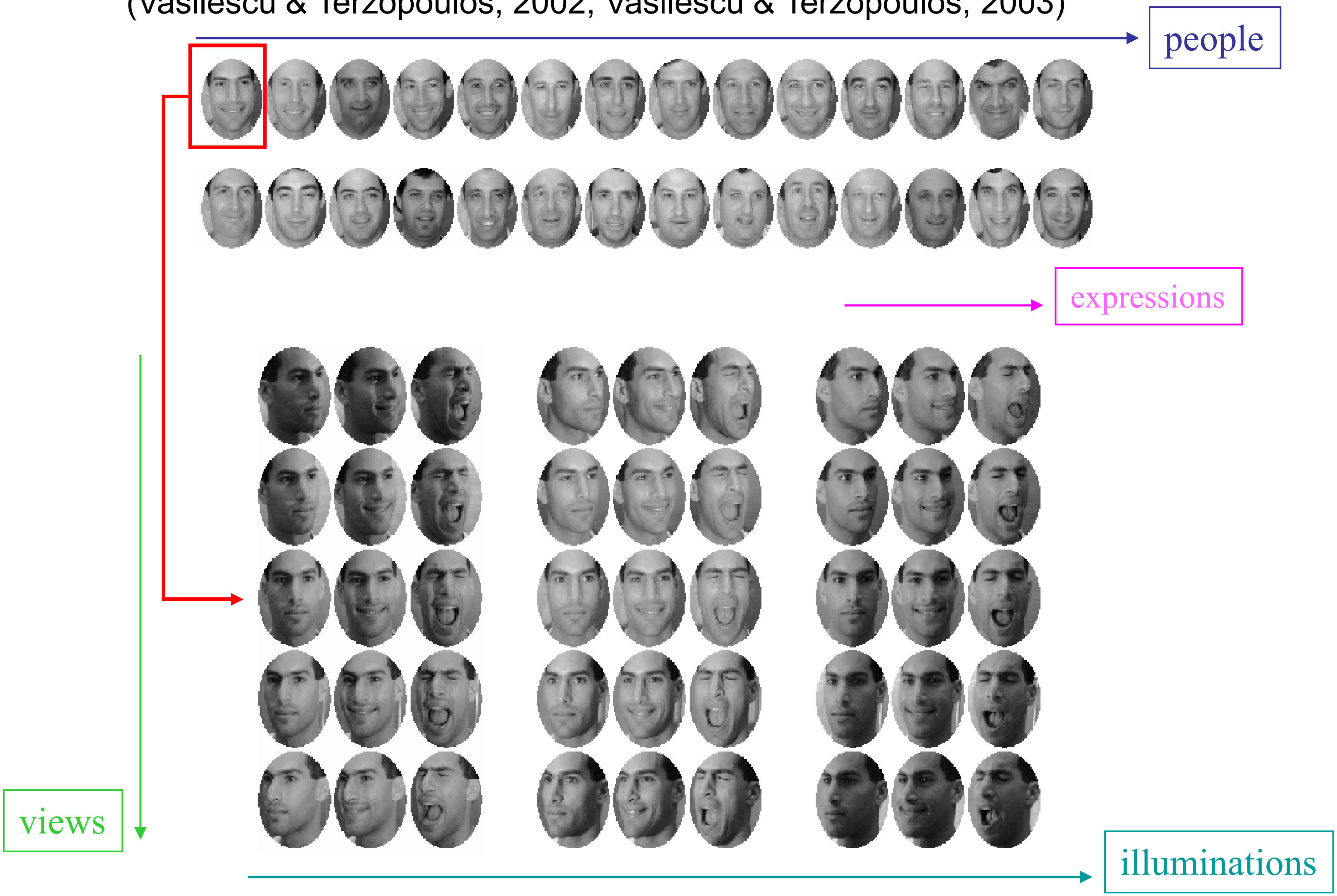
$$\mathbf{c}_i = \mathbf{B}^T\mathbf{d}_i$$



# Tensor Factorization

# Tensor faces

(Vasilescu & Terzopoulos, 2002; Vasilescu & Terzopoulos, 2003)



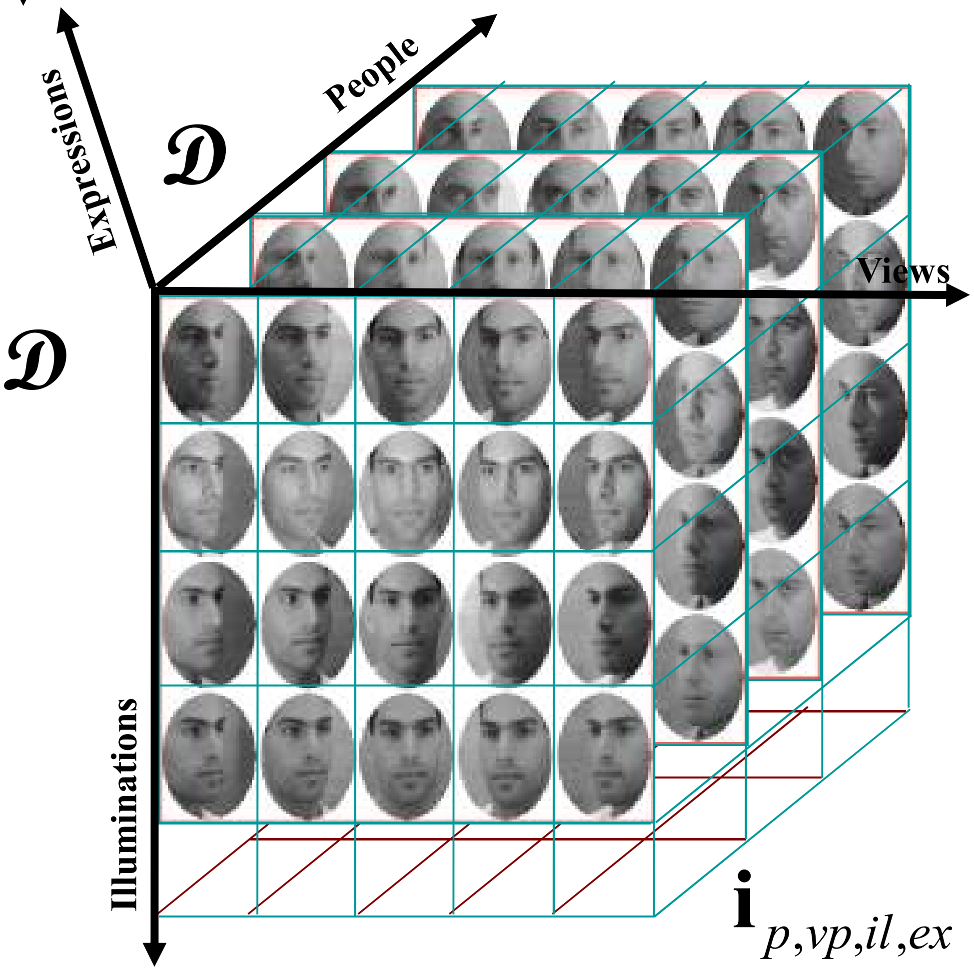
# Eigenfaces

- Facial images (identity change)

- Eigenfaces bases vectors capture the variability in facial appearance (do not decouple pose, illumination, …)

# Data Organization

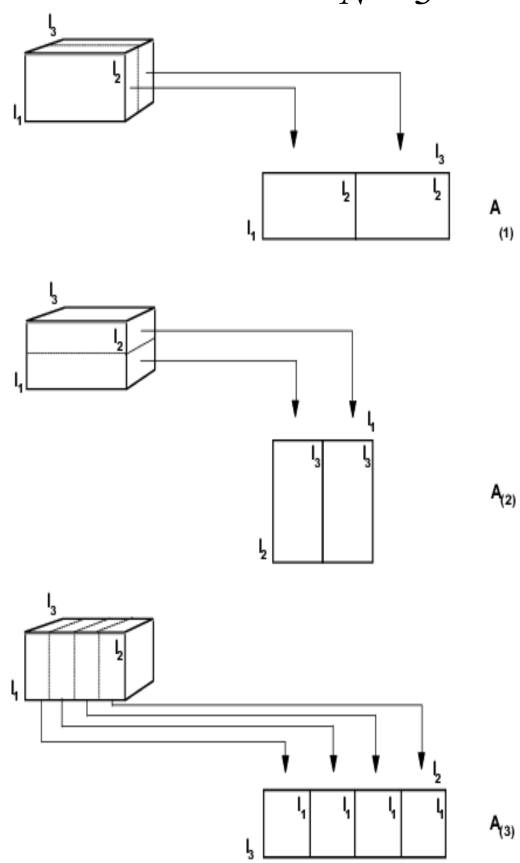
- Linear/PCA: **Data Matrix**
  - $\mathbb{R}^{\text{pixels x images}}$
  - a matrix of image vectors
- Multilinear: *Data Tensor* $\mathcal{D}$
  - $\mathbb{R}^{\text{people x views x illums x express x pixels}}$
  - N-dimensional matrix
  - 28 people, 45 images/person
  - 5 views, 3 illuminations, 3 expressions per person

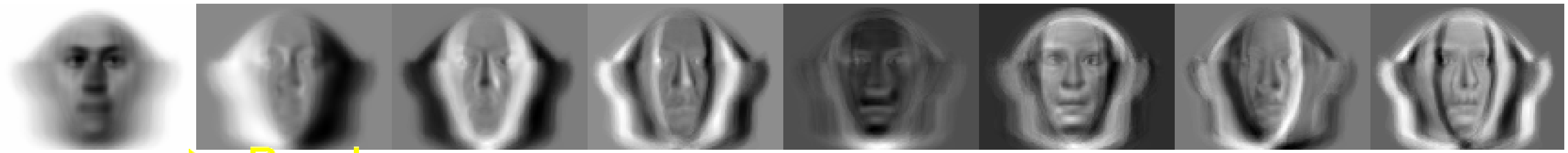
# N-Mode SVD Algorithm

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

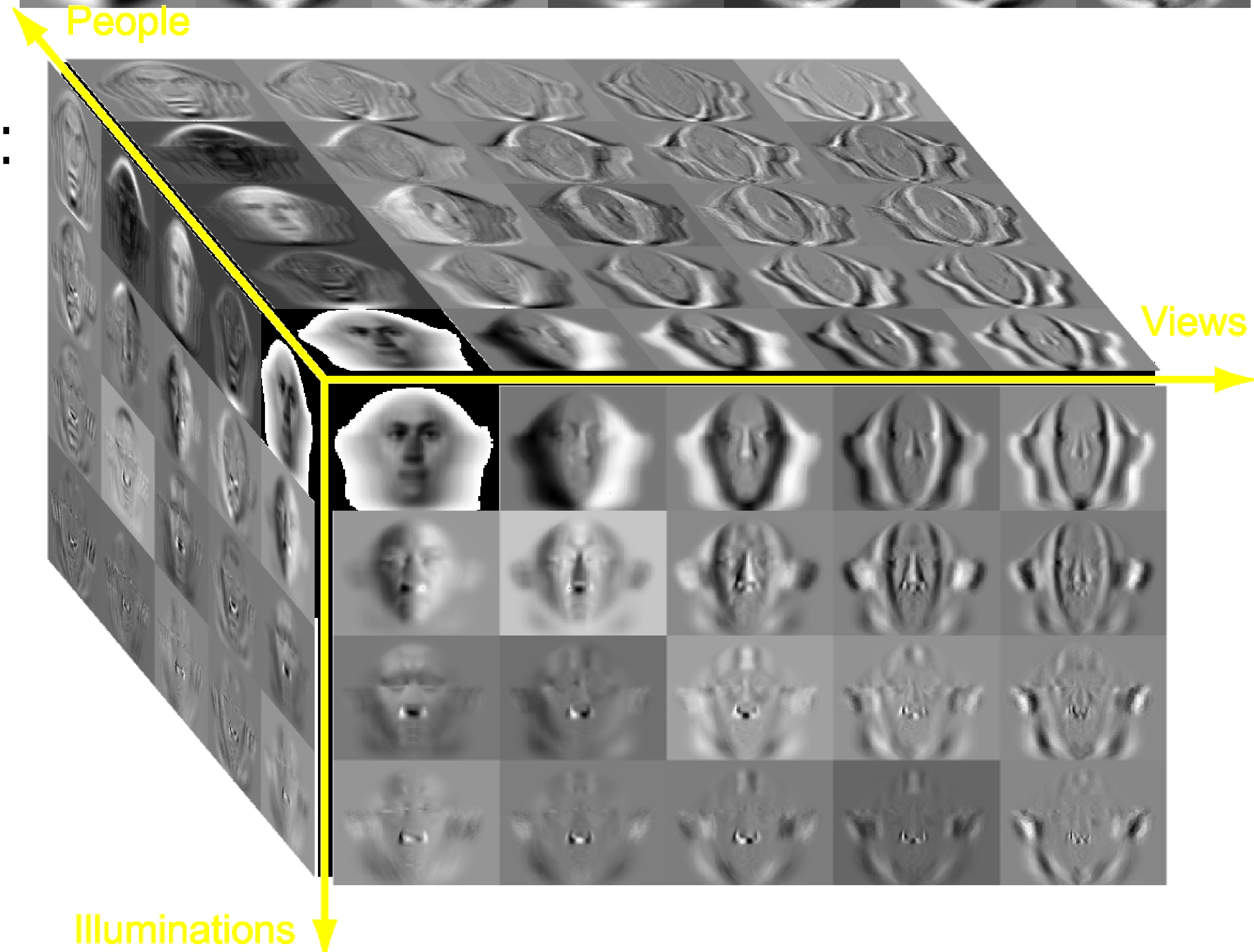
$N = 3$



PCA:



TensorFaces:



# Strategic Data Compression = Perceptual Quality

- TensorFaces data reduction in illumination space primarily degrades illumination effects (cast shadows, highlights)
- PCA has *lower mean square error* but *higher perceptual error*

**Original**

176 basis vectors

**TensorFaces**

6 illum + 11 people param.

66 basis vectors

**TensorFaces**

Mean Sq. Err. = **409.15**

3 illum + 11 people param.

33 basis vectors

**PCA**

Mean Sq. Err. = **85.75**

33 parameters

33 basis vectors

# Acknowledgments


- The content of some of the slides has been taken from previous presentations/papers of:
  - Ales Leonardis.
  - Horst Bischof.
  - Michael Black.
  - Rene Vidal.
  - Anat Levin.
  - Aleix Martinez.
  - Juha Karhunen.
  - Andrew Fitzgibbon.
  - Daniel Lee.
  - Chris Ding.
  - M. Alex Vasilescu.
  - Sam Roweis.
  - Daoqiang Zhang.
  - Ammon Shashua.


CA Can Do It!
WAR PRODUCTION CO-ORDINATING COMMITTEE

# Bibliography


Aans, H., Fisker, R., Aastrom, K., & Carstensen, J. M. (2002). Robust factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *9*, 1215–1225.

Aguiar, P., & Moura, J. (1999). Factorization as a rank 1 problem. *Conference on Computer Vision and Pattern Recognition* (pp. 178–184).

Avidan, S. (2001). Support vector tracking. *Conference on Computer Vision and Pattern Recognition* (pp. 184–191).

Baker, S., Matthews, I., & Schneider, J. (2004). Automatic construction of active appearance models as an image coding problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*, 1380 – 1384.

Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, *2*, 53–58.

Bartlett, M., & Sejnowski, T. (1997). Independent components of face images: a representation for face recognition. *Procs. of the 4th Annual Joint Symposium on Neural Computation* (pp. 523–530).

Belhumeur, P., Hespanha, J., & Kriegman, D. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 711–720.

Beltrami, E. (1873). Sulle funzioni bilineari. *Giornale di Matematiche ad Uso degli Studenti Delle Universita.An English translation by D. Boley is available as University of Minnesota, Department of Computer Science, Technical Report 90–37, 1990.*, *11*.

Bergen, J. R., Anandan, P., Hanna, K. J., & Hingorani, R. (1992). Hierarchical model-based motion estimation. *European Conference on Computer Vision*, 237–252.

Bischof, H., Wildenauer, H., & Leonardis, A. (2004). Illumination insensitive recognition using eigenspaces. *Computer Vision and Image Understanding*, *1*, 86 – 104.

Black, M. (1999). Explaining optical flow events with parameterized spatio-temporal models. *Conference on Computer Vision and Pattern Recognition* (pp. 326–332).

Black, M. J., & Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, *63*, 75–104.

Black, M. J., Fleet, D. J., & Yacoob, Y. (2000). Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, *78*, 8–31.

Black, M. J., & Jepson, A. D. (1998). Eigentracking: Robust matching and tracking of objects using view-based representation. *International Journal of Computer Vision*, *26*, 63–84.

Black, M. J., Yacoob, Y., & Fleet, D. (1998). Modeling appearance change in image sequences. *Advances in Visual Form Analysis* (pp. 11–20).

Blake, A., & Zisserman, A. (1987). *Visual reconstruction*. Massachusetts: MIT Press series.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3d faces. *SIGGRAPH*.

Borga, M. Tutorial on cca. *http://www.imt.liu.se/ magnus/cca/tutorial/node2.html*.

Brand, M. (2001). 3d morphable models from video. *Conference on Computer Vision and Pattern Recognition*.

Brand, M. (2002). Incremental singular value decomposition of uncertain data with missing values. *European Conference on Computer Vision* (pp. 707–720).

Bregler, C., Hertzmann, A., & Biermann, H. (2000). Recovering non-rigid 3D shape from image streams. *CVPR* (pp. 690–696).

Buchanan, A., & Fitzgibbon., A. W. (2005). Damped newton algorithms for matrix factorization with missing data. *Computer Vision and Pattern Recognition*.

Campbell, N., & Tomenson, J. (1983). Canonical variate analysis for several sets of data. *Biometrics*, *39*, 425–435.

Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics*, *29*, 231–2137.

Casia, M. L., & Sclaroff, S. (1999). Fast, reliable tracking under varying illumination. *Conference on Computer Vision and Pattern Recognition* (pp. 604–609).

Champagne, B., & Liu., Q. (1998). Plane rotation-based evd updating schemes for efficient subspaces tracking. *IEEE Transactions on Signal Processing*, 1886–1900.

Chen, L., Liao, H., Ko, M., Lin, J., & Yu, G. (2000). A new ldabased face recognition system which can solve the small sample size problem. *Pattern Recognition*, *33*, 1713–1726.

Chennubhotla, C., & Jepson, A. (2001). Sparse pca: Extracting multi-scale structure from data. *International Conference on Computer Vision* (pp. 641–

Cootes, T., & Taylor, C. (2001). Statistical models of appearance for computer vision. tech. report. university of manchester. .

Cootes, T., Twining, C., V.Petrovic, R.Schestowitz, & Taylor, C. (2005). Groupwise construction of appearance models using piece-wise affine deformations. *British Machine Vision Conference.*

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (1998). Active appearance models. *European Conference Computer Vision* (pp. 484–498).

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models- their training and application. *Computer Vision and Image Understanding, 61*, 38–59.

Croux, C., & Filzmoser., P. (1981). Robust factorization of data matrix. *Proc. in Computational Statistics* (pp. 245–249).

dAspremont, A., Jordan, L. E. G. M., & Lanckriet, G. (2004). A direct formulation for sparse pca using semidefinite programming. *Neural Information Processing Systems.*

de la Torre, F. (2006). Coordinating component analysis. *tech. report CMU-RI-TR-06-08, Robotics Institute, Carnegie Mellon University.*

de la Torre, F., & Black, M. J. (2001a). Dynamic coupled component analysis. *Computer Vision and Pattern Recognition* (pp. 643–650).

de la Torre, F., & Black, M. J. (2001b). Robust principal component analysis for computer vision. *International Conference on Computer Vision* (pp. 362–369).

de la Torre, F., & Black, M. J. (2002). Robust parameterized component analysis: Theory and applications to 2d facial modeling. *European Conf. on Computer Vision* (pp. 653–669).

de la Torre, F., & Black, M. J. (2003a). A framework for robust subspace learning. *International Journal of Computer Vision., 54*, 117–142.

de la Torre, F., & Black, M. J. (2003b). Robust parameterized component analysis: theory and applications to 2d facial appearance models. *Computer Vision and Image Understanding, 91*, 53 – 71.

de la Torre, F., Campoy, J., & Cohn, J. (2007a). Simultaneous registration and clustering for temporal segmentation of facial gestures from video. *2nd International Conference on Computer Vision Theory and Applications,.*

de la Torre, F., Collet, A., Cohn, J., & Kanade, T. (2007b). Filtered component analysis to increase robustness to local minima in appearance models. *submitted to International Conference on Computer Vision and Pattern Recognition.*

de la Torre, F., Gong, S., & McKenna, S. (1998). View-based adaptive affine alignment. *European Conference on Computer Vision* (pp. 828–842).

de la Torre, F., Gross, R., Baker, S., & Kumar, V. (2005a). Representational oriented component analysis (roca) for face recognition with one sample image per training class. *Computer Vision and Pattern Recognition.*

de la Torre, F., & Kanade, T. (2005). Multimodal oriented discriminant analysis. *International Conference on Machine Learning* (pp. 177–184).

de la Torre, F., & Kanade, T. (2006). Discriminative cluster analysis. *International Conference on Machine Learning.*

de la Torre, F., & Nguyen, M. (2007). Kernel appearance models (kams). *submitted to Neural Information Processing.*

de la Torre, F., Vallespi, C., Rybski, P. E., Veloso, M., & Kanade, T. (2005b). Omnidirectional video capturing, multiple people tracking and recognition for meeting monitoring. *tech. report CMU-RI-TR-05-04, Robotics Institute, Carnegie Mellon University, January 2005.*

de la Torre, F., & Vinayls, O. (2007). Learning kernel expansions for image classification. *Accepted for publication in International Conference on Computer Vision and Pattern Recognition.*

de la Torre, F., Vitrià, J., Radeva, P., & Melenchón, J. (2000a). Eigenfiltering for flexible eigentracking. *International Conference on Pattern Recognition* (pp. 1118–1121).

de la Torre, F., Yacoob, Y., & Davis, L. (2000b). A probabilisitc framework for rigid and non-rigid appearance based tracking and recognition. *Int. Conf. on Automatic Face and Gesture Recognition* (pp. 491–498).

Dhillon, I. S., Guan, Y., & Kulis, B. (2004). A unified view of kernel k-means, spectral clustering and graph partitioning. *UTCS Technical Report TR-04-25.*

Diamantaras, K. I. (1996). *Principal component neural networks (therory and applications).* John Wiley & Sons.

Ding, C., & He, X. (2004). K-means clustering via principal component analysis. *International Conference on Machine Learning.*

Ding, C., He, X., & Simon, H. (2005). On the equivalence of nonnegative matrix factorization and spectral clustering. *Siam International Conference on Data Mining (SDM).*

Ding, C., & Ye., J. (2005). Two-dimensional singular value decomposition (2dsvd) for 2d maps and images. .

Eckardt, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1*, 211–218.

Everingham, M., & Zisserman, A. (2006). Regression and classification approaches to eye localization in face images. *Proceedings of the International Conference on Automatic Face and Gesture Recognition.*

Everitt, B. S. (1984). *An introduction to latent variable models.* Chapman and Hall.

Fidler, S., Skocaj, D., & Leonardis, A. (2006). Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*, 337– 350.

Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics, 8*, 376–386.

Fitzgibbon, A., & Zisserman, A. (2003). Joint manifold distance: a new approach to appearance based clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 26–33).

Frey, B. J., & Jojic, N. (1999a). Estimating mixture models of images and inferring spatial transformations using the em algorithm. *Conference on Computer Vision and Pattern Recognition* (pp. 416–422).

Frey, B. J., & Jojic, N. (1999b). Transformed component analysis: Joint estimation of spatial transformations and image components. *International Conference on Computer Vision.*

Fukunaga, K. (1990). *Introduction to statistical pattern recognition, second edition.* Academic Press.Boston, MA.

Gabriel, K. R., & Odoroff, C. L. (1984). Resistant lower rank approximation of matrices. *Data Analysis and Informatics, III* (pp. 23–30).

Gabriel, K. R., & Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics, Vol. 21, pp., 21*, 489–498.

Gallinari, P., Thiria, S., Badran, F., & Fogelman-Soulie, F. (1991). On the relations between discriminant analysis and multilayer perceptrons. *Neural Networks, 4*, 349–360.

Golub, G., & Loan, C. F. V. (1989). *Matrix computations.* 2nd ed. The Johns Hopkins University Press.

Gong, S., Mckenna, S., & Psarrou, A. (2000). *Dynamic vision: From images to face recognition.* Imperial College Press.

Gordon, G. (2002). Generalized$^2$ linear$^2$ models. *Neural Information Processing.*

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Guerreiro, R. F. C., & Aguiar, P. M. Q. (2002). 3d structure from video streams with partially overlapping images. *IEEE International Conference on Image Processing.*

Hartley, R., & Schaffalitzky, F. (2003). Powerfactorization: an approach to affine reconstruction with missing and uncertain data. *Australia-Japan Advance Workshop on Computer Vision.*

Hartley, R. I. (1992). Estimation of relative camera positions for uncalibrated cameras. *European Conference on Computer Vision.*

Hartley, R. I., & Zisserman, A. (2000). *Multiple view geometry in computer vision.* Cambridge University Press.

Hastie, T., Tibshirani, R., & Buja, A. (1995). Flexible discriminant and mixture models. *Neural Networks and Statistics. J. Kay and D. Titterington, Eds.*

Hayakawa, H. (1994). Photometric stereo under a light-source with arbitrary motion. *JOSA-A, 11*, 3079–3089.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24.*

Hyvrinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis.* John Wiley and Sons.

Irani, M., & Anandan, P. (2000). Factorization with uncertainty. *European Conference on Computer Vision* (pp. 539–553).

Jain, A., Murty, M., & Flynn, P. (1999). Data clustering: A review. *ACM Computing Surveys.*

Jebara, T., Russell, K., & Pentland, A. (1998). Mixtures of eigenfeatures for real-time structure from texture. *International Conference on Computer Vision* (pp. 128–135).

Jieping Ye, Ravi Janardan, Q. L. (2005). Two-dimensional linear discriminant analysis. *Neural Information Processing Systems 2005*, 1569–1576.

Jogan, M., agar, E., & Leonardis, A. (2003). Karhunen-loeve transform of a set of rotated templates. *IEEE Trans. on Image Processing, 12*, 817–825.

Jolliffe, I. T. (1986). *Principal component analysis.* New York: Springer-Verlag.

Jones, D. G., & Malik, J. (1992). Computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing, 10*, 699–708.

Jones, M. J., & Poggio, T. (1998). Multidimensional morphable models. *International Conference on Computer Vision* (pp. 683–688).

Ke, Q., & Kanade, T. (2004). A robust subspace approach to layer extraction. *CVPR*.

Kiers, H. A. L. (1995). Maximization of sums of quotients of quadratic forms and some generalizations. *Psychometrika*, *60*, 221–245.

Kong, H., & Wang, L. (2005). Generalized 2d principal component analysis for face image representation and recognition. *Neural Networks*, *18*, 585–94.

Kumar, N., & Andreou, A. (1998). Heteroscedastic discriminant analysis and reduced rank hmms for improved speech recognition. *Speech Communication*, *26*, 283 – 297.

Lanitis, A., Hill, A., Cootes, T. F., & Taylor, C. J. (1995). Locating facial features using genetic algorithms. *International Conference on Digital Signal Processing* (pp. 520–525).

Lee, C., & Elgammal, A. (2005). Facial expression analysis using nonlinear decomposable generative models. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures* (pp. 17–31).

Lee, D., & Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–791.

Leeuw, J. D. (1994). *Block relaxation algorihtms in statistics*. H.H. Bock, W. Lenski, M. Ritcher eds. Information Systems and Data Analysis. Springer-Verlag.

Leonardis, A., & Bischof, H. (2000). Robust recognition using eigenimages. *Computer Vision and Image Understanding*, *1*, 99–118.

Leonardis, A., Bischof, H., & Maver, J. (2002). Multiple eigenspaces. *Pattern Recognition*, *35*, 2613–2627.

Levin, A., & Shashua, A. (2002). Principal component analysis over continuous subspaces and intersection of half-spaces. *European Conference on Computer Vision*.

Levy, A., & Lindenbaum., M. (2000). Sequential karhunen-loeve basis extraction and its application to images. *IEEE Transactions on Image Processing*, 1371–1374.

Liu, K., Cheng, Y., & Yang, J. (1993). Algebraic feature extraction for image recognition based on an optimal discriminant criterion. *Pattern Recognition*, *6*, 903–911.

Liu, R., & Tan, T. (2000). A new svd based image watermarking method. *4th Asian Conference on Computer Vision*.

Lowe, D. G., & Webb, A. (1991). Optimized feature extraction and the bayes decision in feed-forward classifier networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 355–364.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *5-th Berkeley Symposium on Mathematical Statistics and Probability.Berkeley, University of California Press.* (pp. 1:281–297).

Mardia, K., Kent, J., & Bibby, J. (1979). *Multivariate analysis*. Academic Press, London.

Marimont, D. H., & Wandell, B. (1992). Linear models of surface and illuminant spectra. *Journal of the Optical Society of America A*, *9*, 1905–1913.

Martinez, A., & Kak, A. (2003). Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 228–233.

Martinez, A., & Zhu., M. (2005). Where are linear feature extraction methods applicable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*, 1934–1944.

McKenna, S., Gong, S., & Raja, Y. (1997a). Face recognition in dynamic scenes. *British Machine Vision Conference*.

McKenna, S. J., Gong, S., Würtz, R. P., Tanner, J., & Banin, D. (1997b). Tracking facial feature points with Gabor wavelets and shape models. *Proceedings of the First International Conference on Audio- and Video-based Biometric Person Authentication Crans-Montana, Switzerland* (pp. 35–42).

Melzer, T., Reiter, M., & H.Bischof (2001). Kernel cca: A nonlinear extension of canonical correlation analysis. *ICANN*.

Moghaddam, B. (1999). Principal manifolds and Bayesian subspaces for visual recognition. *Seventh International Conference on Computer Vision* (pp. 1131–1136).

Moghaddam, B., Jebara, T., & Pentland, A. (2000). Bayesian face recogntion. *Pattern Recogntion*, *11*, 1771–1782.

Moghaddam, B., & Pentland, A. (1997). Probabilistic visual learning for object representation. *Pattern Analysis and Machine Intelligence*, *19*, 137–143.

Moonen, M., & de Moor, B. (1995). *Svd and signal processing iii: Algorithms, applications and architectures*. Elsevier Science Publishers.

Morris, D., & Kanade, T. (1998). A unified factorization algorithm for points, line segments and planes with uncertainty models. *International Conference on Computer Vision* (pp. 696–702).

Murase, H., & Nayar, S. (1994). Illumination planning for object recognition using parametric eigenspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *16*, 1219–1227.

Murase, H., & Nayar, S. K. (1995). Visual learning and recognition of 3D objects from appearance. *International Journal of Computer vision*, *1*, 5–24.

Oja, E. (1982). A simplified neuron model as principal component analyzer. *Journal of Mathematical Biology*, *15*, 267–273.

O'Leary, D., & Peleg, S. (1983). Digital image compression by outer product expansion. *IEEE Trans. on Communications*, *31*, 441–444.

Olhausen, B., & Field, D. (1996). Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, *7*, 333–339.

Olshausen, B., & Field, D. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 3311–3325.

Orriols, X., & Binefa, X. (2001). An EM algorithm for video summarization, generative model approach. *International Conference on Computer Vision* (pp. 335–342).

Paatero, P., & Tapper, U. (1994). Positive matrix factorization:a non-negative factor model with optimal utilization of error estimates of data values. *Envirometrics*, *5*, 111–126.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal*, *6*, 559–572.

Penev, P. S., & Atick, J. J. (1996). Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, *7*, 477–500.

Pentland, A., Moghaddam, B., & Starner, T. (1994). View-based and modular eigenspaces for face recognition. *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 84–91).

Rao, R., & Miao, X. (1999). Learning lie groups for invariant visual perception. *Neural Information Processing Systems* (pp. 810–816).

Rao, R. P. N. (1997). Dynamic appearance-based vision. *Conference on Computer Vision and Pattern Recognition* (pp. 540–546).

Romdhani, S., Gong, S., & Psarrou, A. (1999). Multi-view nonlinear active shape model using kernel PCA. *In British Machine Vision Conference* (pp. 483–492).

Ross, D., Lim, J., & Yang, M. (2004). Adaptive probabilistic visual tracking with incremental subspace update. *Eighth European Conference on Computer Vision*.

Roweis, S. (1997). EM algorithms for PCA and SPCA. *Neural Information Processing Systems* (pp. 626–632).

Roweis, S., & Ghahramani, Z. (1999). A unifying review of linear gaussian models. *Neural Computation*, *11*, 305–345.

Ruymagaart, F. H. (1981). A robust principal component analysis. *Journal of Multivariate Analysis*, *11*, 485–497.

Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, *2*, 459–473.

Saon, G., Padmanabhan, M., Gopinath, R., & Chen, S. (2000). Maximum likelihood discriminant feature spaces. *ICASSP*.

Schewitzer, H. (1999). Optimal eigenfeature selection by optimal image registration. *Conference on Computer Vision and Pattern Recognition* (pp. 219–224).

Schmidt, E. (1907). Zur theorie der linearen und nichtlinearen integralgleichungen. *Math. Ann.*, *63*.

Scholkopf, B., & Smola, A. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond.* MIT Press.

Scholkopf, B., Smola, A., & Muller, K. (1998). Detecting faces in images: a survey. *Neural Computation*, *10*, 1299–1319.

Shakhnarovich, G., Fisher, J. W., & Darrell., T. (2002). Face recognition from long-term observations. *European Conference on Computer Vision*.

Shakhnarovich, G., & Moghaddam, B. (2004). Face recognition in subspaces. *Handbook of Face Recognition. Eds. Stan Z. Li and Anil K. Jain, Springer-Verlag. Also MERL TR2004-041.*

Shashua, A., & Hazan, T. (2005). Non-negative tensor factorization with applications to statistics and computer vision. .

Shashua, A., & Levin, A. (2001). Linear image coding for regression and classification using the tensor-rank principle. *IEEE Conf. on Computer Vision and Pattern Recognition*.

Shashua, A., Levin, A., & Avidan, S. (2002). Manifold pursuit: A new approach to appearance based recognition. *ICPR*.

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis.* Cambridge University Press.

Shental, N., Hertz, T., Weinshall, D., & Pavel, M. (2002). Adjustment learning and relevant component analysis. *European Conferenceon Computer Vision* (pp. 776–790).

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*.

Shum, H., Ikeuchi, K., & Reddy, R. (1995). Principal component analysis with missing data and its application to polyhedral object modeling. *Pattern Analysis and Machine Intelligence*, *17*, 855–867.

Sidenbladh, H., de la Torre, F., & Black, M. J. (2000). A framework for modeling the appearance of 3D articulated figures. *Face and Gesture Recognition* (pp. 368–375).

Simoncelli, E. P., Freeman, W. T., Adelson, E. H., & Heeger, D. J. (1992). Shiftable multi-scale transforms. *IEEE Trans. Information Theory, Special Issue on Wavelets*, *38*, 587–607.

Sirovich, L. (1987). Turbulence and the dynamics of coherent structure. *Quart. Applied Mathematics*, *XLV*, 561–590.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, *4*, 519–524.

Skocaj, D., Bischof, H., & Leonardis, A. (2002). A robust pca algorithm for building representations from panoramic images. *European Conference on Computer Vision*.

Skocaj, D., & Leonardis, A. (2000). Appearance-based localization using cca. *Computer Vision Winter Workshop*.

Skocaj, D., & Leonardis, A. (2003). Weighted and robust incremental method for subspace learning. *International Conference on Computer Vision ICCV* (pp. 1494–1501).

Soatto, S., Doretto, G., & Wu, Y. N. (2001). Dynamic textures. *International Conference Computer Vision* (pp. 439–446).

Sozou, P., Cootes, T. F., Taylor, C. J., & DiMauro, E. (1995). A non-linear generalisation of point distribution models using polynomial regression. *Image and Vision computing*, *13*, 451–457.

Sturm, P., & Triggs, B. (1996). A factorization based algorithm for multi-image projective structure and motion. *European Conference on Computer Vision* (pp. 709–720).

Tenenbaum, J. B., & Freeman, W. T. (2000). Separating style and context with bilinear models. *Neural Computation*, *12*, 1247–1283.

Tipping, M., & Bishop, C. M. (1999a). Mixtures of probabilistic principal component analyzers. *Neural Computation*, *11*, 443–482.

Tipping, M., & Bishop, C. M. (1999b). Probabilistic principal component analysis. *Journal of the Royal Statistical Society B*, *61*, 611–622.

Tomasi, C., & Kanade, T. (1992). Shape and motion from image streams under orthography: a factorization method. *Int. Jorunal of Computer Vision.*, *9*, 137–154.

Torresani, L., & Bregler, C. (2004). Automatic non-rigid 3d modeling from video. *European Conference on Computer Vision*.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal Cognitive Neuroscience*, *3*, 71–86.

Uenohara, M., & Kanade, T. (1998). Optimal approximation of uniformly rotated images: Relationships between karhunenloeve expansion and discrete cosine transform. *IEEE Transactions on Image Processing*, *7*, 116–119.

Vasilescu, M., & Terzopoulos, D. (2002). Multilinear analysis of image ensembles: Tensorfaces. *Proc. European Conf. on Computer Vision*.

Vasilescu, M., & Terzopoulos, D. (2003). Multilinear subspace analysis of image ensembles. *Computer Vision and Pattern Recognition*.

Verbeek, J. (2006). Learning non-linear image manifolds by combining local linear models. *IEEE transactions on Pattern Analysis and Machine Intelligence*.

Vidal, R., Ma, Y., & Sastry, S. (2003). Generalized principal component analysis. *Computer Vision and Pattern Recognition*.

Walker, K., Cootes, T., & Taylor, C. (2000). Determining correspondences for statistical models of appearance. *European Conference on Computer Vision* (pp. 829–843).

Welling, M., Agakov, F., & Williams, C. (2003). Extreme components analysis. *Neural Information Processing Systems*.

Wiberg, T. (1976). Computation of principal components when data are missing. *Proc. Second Symp. Computational Statistics* (pp. 229–236).

Wildenauer, H., Melzer, T., & Bischof, H. (2002). A gradient-based eigenspace approach to dealing with occlusions and non-gaussian noise. *International Conference on Pattern Recognitoin*.

Williams, C., & Titsias, M. (2004). Greedy learning of multiple objects in images using robust statistics and factorial learning. *Neural Computation*, *4*, 1039–1062.

# Bibliography


Xu, L. (1993). Least mean square error recosntruction for self-organizing nerual nets. *Neural Networks*, *6*, 627–648.

Xu, L., & Yuille., A. (1995). Robust principal component analysis by self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks*, *6*, 131–143.

Yacoob, Y., & Black, M. J. (1999). Parameterized modeling and recognition of activities. *CVIU*, *2*, 232–247.

Yang, J., Zhang, D., Frangi, A., & Yang, J. (2004a). Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*, 131–137.

Yang, J., Zhang, D., Frangi, A., & Yang, J. (2004b). Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*, 131–137.

Yang, M., Ahuja, N., & Kriegman., D. (2000a). Face detection using mixtures of linear subspaces. *IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 70–76).

Yang, M., Ahuja, N., & Kriegman, D. (2000b). Face recognition using kernel eigenfaces. .

Yang, M.-H., Ahuja, N., & Kriegman, D. (1999). Face detection using a mixture of factor analyzers. .

Ye, J. (2004). Generalized low rank approximations of matrices. *International Conference on machine Learning* (pp. 887–894).

Ye, J., Li, Q., Xiong, H., Park, H., Janardan, R., & Kumar, V. (2005). Idr/qr: An incremental dimension reduction algorithm via qr decomposition. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, *17*, 1208–1222.

Yu, H., & Yang, J. (2001). A direct lda algorith for high-dimensional data– with applicatins to face recognition. *Pattern Recognition*, *34*, 2067–2070.

Yu, S., & Shi, J. (2003). Multiclass spectral clustering. *International Conference on Computer Vision*.

Zass, R., & Shashua, A. (2005). A unifying approach to hard and probabilistic clustering. *International Conference on Computer Vision*.

Zha, H., Ding, C., Gu, M., He, X., & Simon., H. (2001). Spectral relaxation for k-means clustering. *Neural Information Processing Systems* (pp. 1057–1064).

Zhang, D., & Zhou, Z. (2005). (2d)2pca: 2-directional 2-dimensional pca for efficient face representation and recognition. *Neurocomputing*, 224–231.

Zhang, S., Zhou, Z., & Chen, S. (2006). Diagonal principal component analysis for face recognition. *Pattern Recognition Letters*, *39*, 133–135.

Zhang, Z., Deriche, R., Faugeras, O., & Luong, Q. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, *78*, 87–119.

Zhao, W. (2000). Discriminant component analysis for face recognition. *ICPR* (pp. 818–821).

Zhu, M., & Martinez, A. (2006). Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. *Accepted for publication*.

Zou, H., Hastie, T., & Tibshirani, R. (2005). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*.

Zhou F., De la Torre F. and Hodgins J. (2008) **"Aligned Cluster Analysis for Temporal Segmentation of Human Motion"** *IEEE Conference on Automatic Face and Gestures Recognition, September, 2008.*

De la Torre, F. and Nguyen, M. (2008) **"Parameterized Kernel Principal Component Analysis: Theory and Applications to Supervised and Unsupervised Image Alignment"** *IEEE Conference on Computer Vision and Pattern Recognition, June, 2008.*