CSE276C - Subspace Methods

Henrik I. Christensen

Computer Science and Engineering
University of California, San Diego
http://cri.ucsd.edu

October 2024

Literature

- Leonardis, A. and Bischof, H., 2000. "Robust recognition using eigenimages". **Computer Vision and Image Understanding**, 78(1), pp.99-118.
- Largely adopted from ECCV tutorial by Leonardis and Bischof

Outline

Recognition of objects in clutter

Recognition of objects in clutter

Typical tasks

- Where can I find a can of coke?
- Check the stove – is it off?
- Put away the groceries in the pantry?

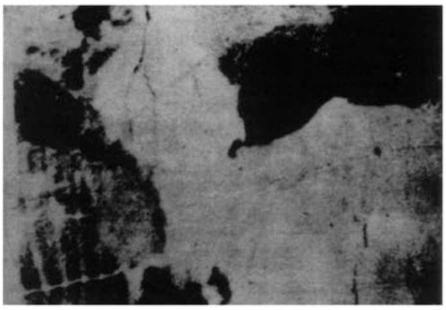
Outline

Object Representation

- High-level Shape Models (e.g., Generalized Cylinders)
  - Idealized images
  - Texture Less
- Mid-level Shape Models (e.g., CAD models, Superquadrics)
  - More complex
  - Well-defined geometry
- Low-level Appearance Based Models (e.g., Eigenspaces)
  - Most complex
  - Complicated shapes

A number of challenges

Segmentation:



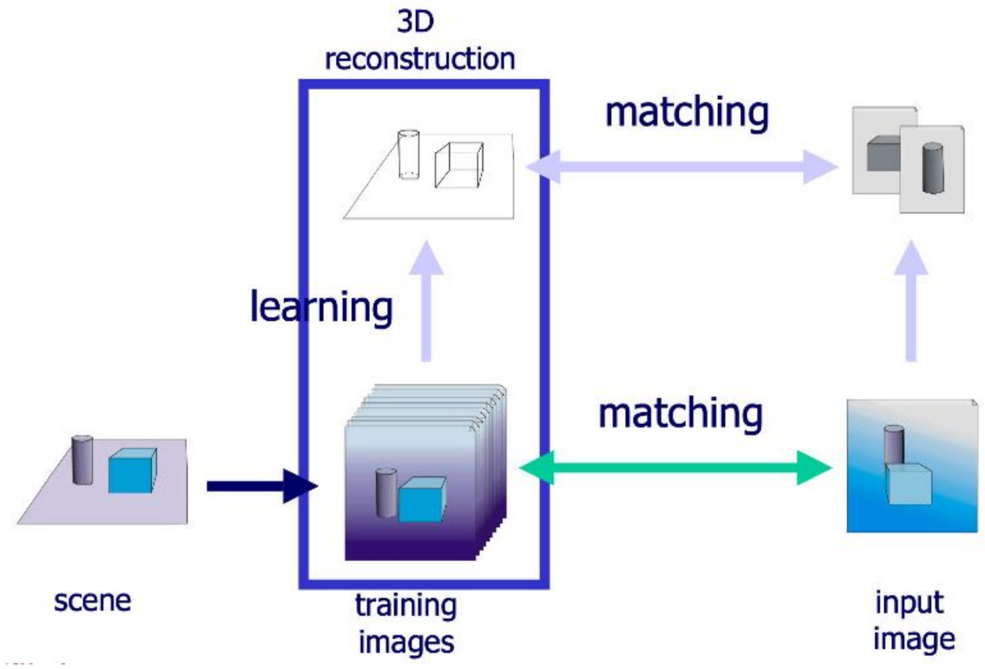
Pose/Shape:

Changes in illumination

The importance of context

The importance of context - see

Learning and recognition



Outline

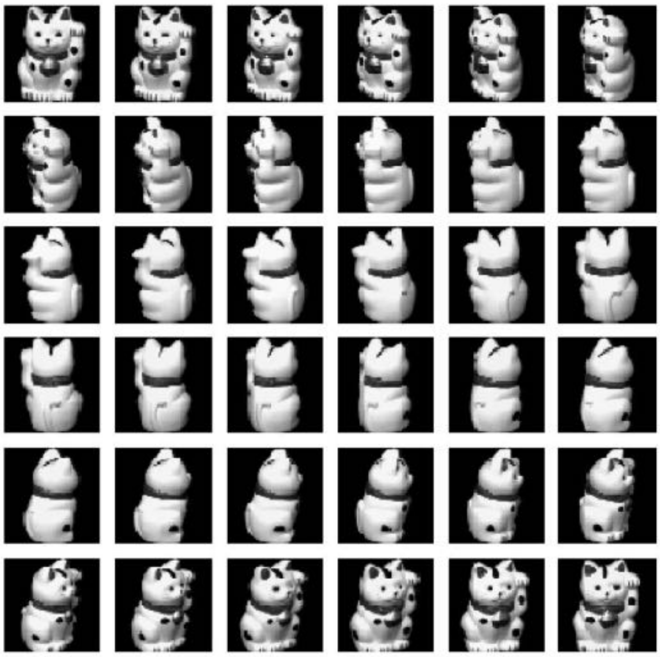
Appearance-based approaches

- The abundance of image data gives a renewed interest in appearance-based approaches
- Combined effort of:
  - Shape
  - Reflectance properties
  - Pose in the scene
  - Illumination conditions / variations
- Acquired through an automatic learning phase
- Well defined error characteristics

Numerous use-cases

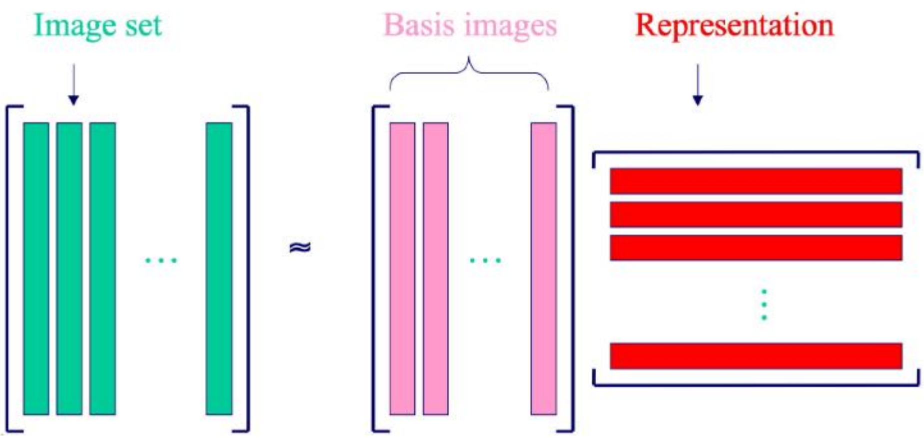
- Face-recognition (eigen faces)
- Visual inspection
- Tracking and pose estimation for robotics
- Basic object tracking
- Planning of illumination
- Image spotting
- Mobile robot localization
- ...

IDEA: Take a large number of image views



IDEA: Subspace Methods

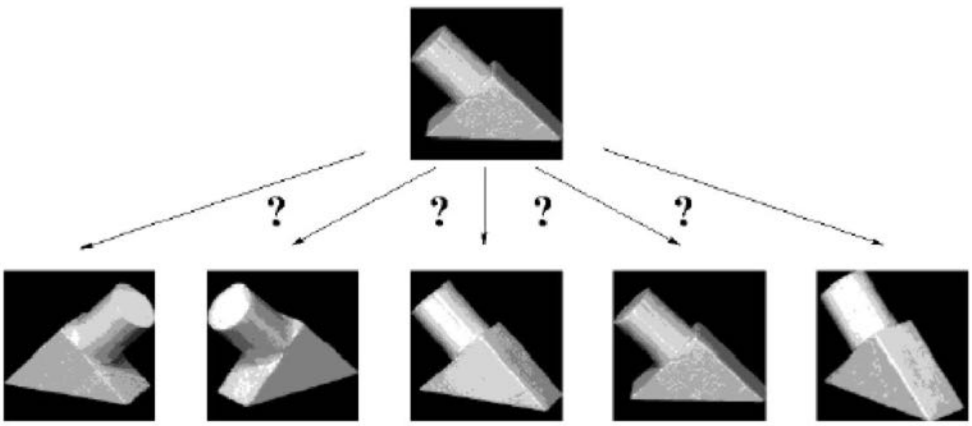
- Images are represented as points in an N-dimensional space
- Images only occupy a small fraction of the hyper-space
- Characterize the subspace / manifold spanned by the images



Multiple subspace methods

- Optimal Reconstruction $\Rightarrow$ PCA
- Optimal Separation $\Rightarrow$ LDA
- Optimal Correlation $\Rightarrow$ CCA
- Independent Factors $\Rightarrow$ ICA
- Non-negative factorization $\Rightarrow$ NMF

Image matching



$$\rho = \frac{x^T y}{||x||||y||} \geq \Theta$$

Or Normalized Images

$$||x - y||^2 \leq \Phi$$

Eigenspace representation

- Image set (normalized, zero mean)

$$X = [x_0 \; x_1 \; \dots \; x_{n-1}]; \; X \in R^{m \times n}$$

- Looking for ortho-normal basis

$$U = [u_0 \; u_1 \; \dots \; u_k]; k \ll n$$

- Individual images are then a linear combination of basis vectors

$$x_i \approx \tilde{x}_i = \sum_{j=0}^{k} q_j(x_i) u_j$$

$$||x - y||^2 \approx || \sum_{j=0}^{k} q_j(x) u_j - \sum_{j=0}^{k} q_j(y) u_j ||^2$$

$$|| \sum_j q_j(x) - q_j(y) ||^2$$

Choosing a basis function?

- The optimization problem

$$\sum_{i=0}^{n-1} ||x_i - \sum_{j=0}^{k} q_j(x_i) u_j||^2 \rightarrow \min$$

- Taking k eigenvectors with the largest eigenvalues

$$C = X \, X^T = [x_0 \; x_1 \; \dots \; x_{n-1}] \begin{bmatrix} x_0^T \\ x_1^T \\ \dots \\ x_{n-1}^T \end{bmatrix}$$

- The PCA or Karhunen-Loeve Transform

$$C u_i = \lambda_i u_i$$

Efficient eigenspace computation

- n ≪ m
- Computing the eigenvectors u'_i i = 0, . . . , n-1 of the inner product matrix

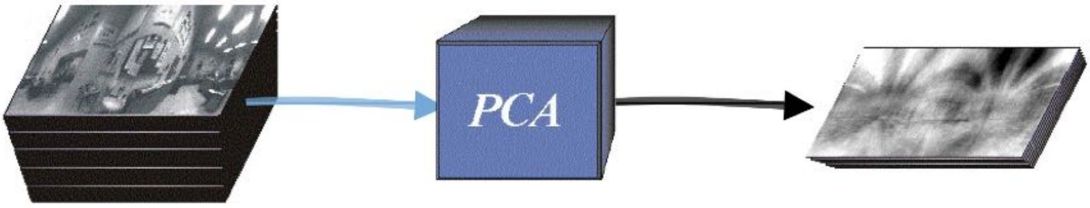
$$Q = X^T X = \begin{bmatrix} x_0^T \\ x_1^T \\ \dots \\ x_{n-1}^T \end{bmatrix} [x_0 \; x_1 \; \dots \; x_{n-1}]; Q \in R^{n \times n}$$

- The eigenvectors of XX^T can be obtained using $XX^T X v'_i = \lambda'_i X v'_i$:

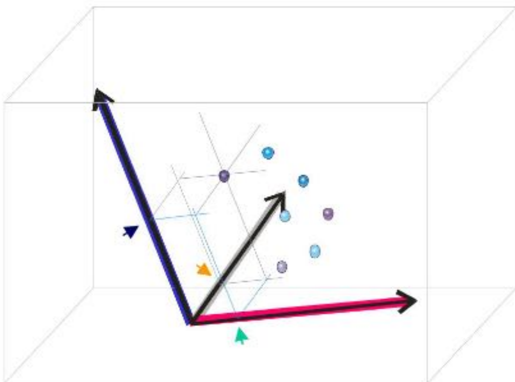
$$u_i = \frac{1}{\sqrt{\lambda'_i}} X u'_i$$

Outline

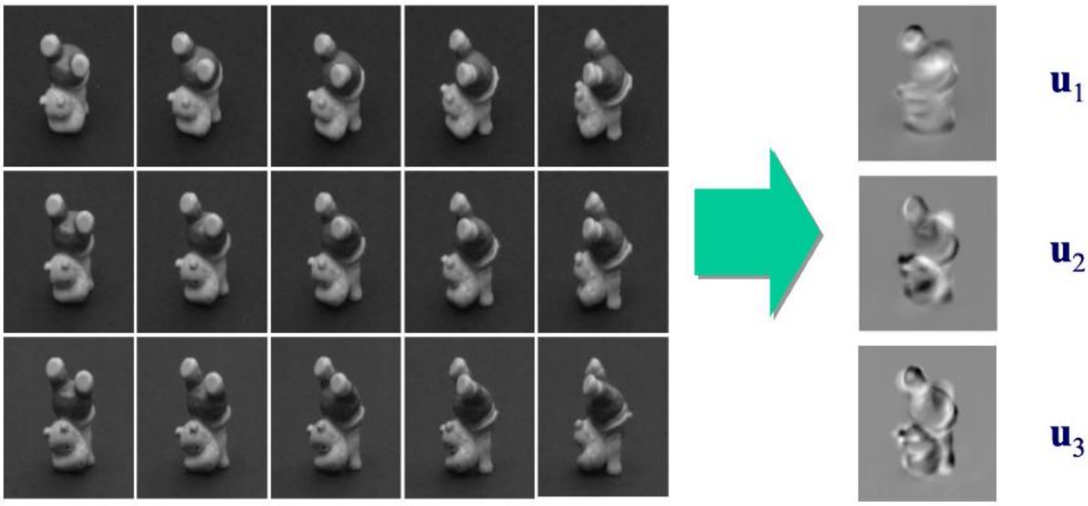
Principal Component Analysis



Principal Component Analysis



PCA Image Representation



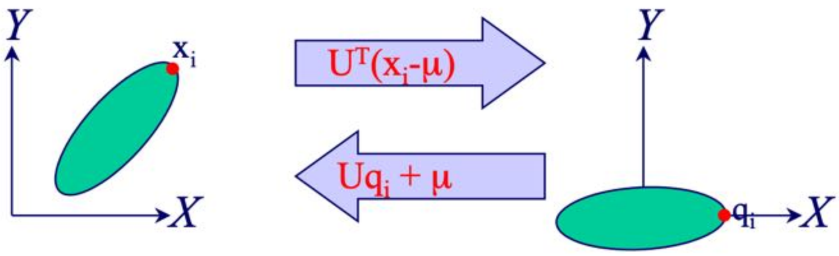
Properties of PCA

- Any point x_i can be projected to an appropriate point q_i by

$$q_i = U^T(x_i - \mu)$$

- and conversely

$$Uq_i + \mu = x_i$$



Properties of PCA

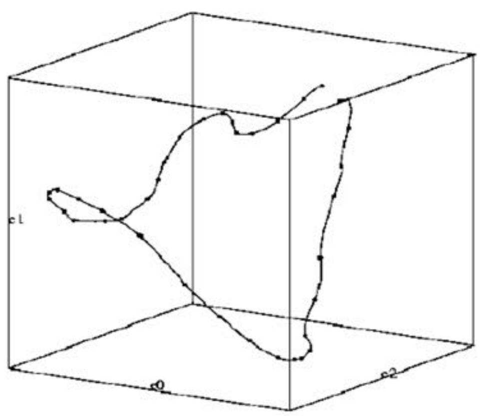
- It can be shown the MSE between x_i and its reconstruction using m eigenvectors is given by

$$\sum_{j=1}^{N} \lambda_j - \sum_{j=1}^{m} \lambda_j = \sum_{j=m+1}^{N} \lambda_j$$

- PCA minimizes the reconstruction error
- PCA maximizes the variance of projection
- Find a "natural" coordinate system for the sample data

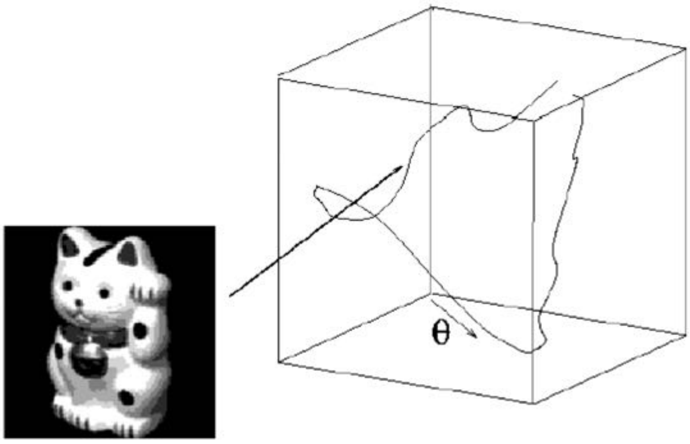
PCA for visual recognition and pose estimation

- Objects/images are represented as coordinates in an m-dimension space
- An example
  - 3D space with points representing objects on a manifold of parametric eigenspace such as orientation, pose, illumination, ...



PCA for visual recognition and pose estimation

- Calculate coefficients
- Search for nearest point on manifold
- Point determines / interpolates object and/or pose



Coefficient calculation

- To recover a_i the image is projected into the eigenspace

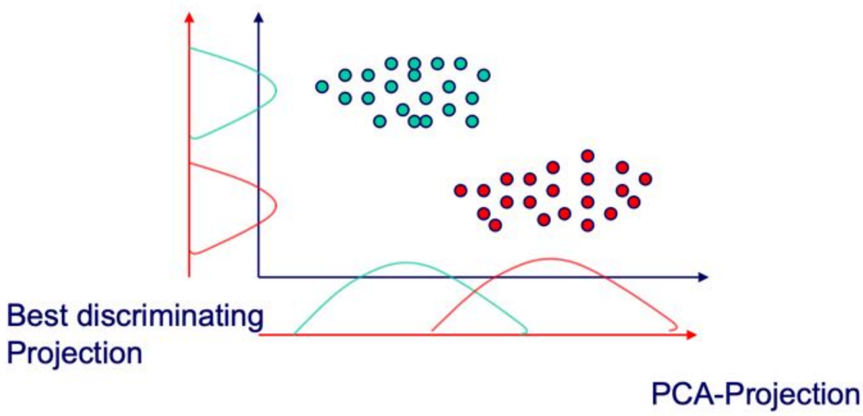
$$a_i(\mathbf{x}) = <\mathbf{x}, \mathbf{e}_i> = \sum_{j=1}^{m} x_j e_{i_j} \qquad 1 \le i \le p$$

- Complete image x_i is required to calculate a_i
- Corresponds to a least square solution

Outline

Linear Discriminate Analysis

- PCA minimizes the projection error



- PCA is unsupervised – no class information is used
- Discriminating information may be used

Linear Discriminate Analysis

- For LDA would would like to
  - Maximize distance between classes
  - Minimize distance within classes
- Fisher linear discriminant

$$\rho(W) = \frac{W^T S_B W}{W^T S_W W}$$

LDA: Problem Formulation

- n sample images: $\{x_1, \ldots, x_n\}$
- c classes: $\{\chi_1, \ldots, \chi_c\}$
- Average of each class: $\mu_i = \frac{1}{n_i} \sum_{x_k \in \chi_i} x_k$
- Total average: $\mu = \frac{1}{n} \sum_{k=1}^{N} x_k$

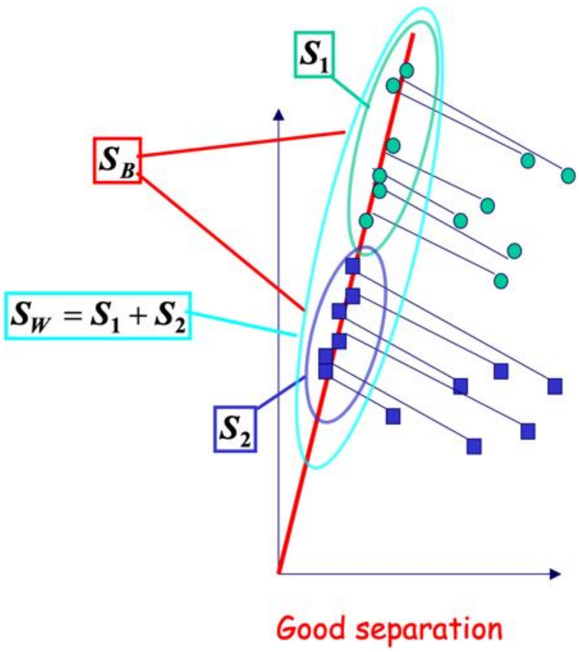
LDA: Practice

- Scatter of class i: $S_i = \sum_{x_k \in \chi_i} (x_k - \mu_i)(x_k - mu_i)^T$
- Within class scatter: $S_W = \sum_{i=1}^{c} S_i$
- Between class scatter: $S_B = \sum_{i=1}^{c} |\chi_i|(\mu_i - \mu)(\mu_i - \mu)^T$
- Total scatter: $S_T = S_W + S_B$

LDA: Practice

- After projection: $y_k = W^T x_k$
  - Between class scatter of y: $\tilde{S}_B = W^T S_B W$
  - Within class scatter of y: $\tilde{S}_W = W^T S_W W$

LDA Projection



LDA characteristics

- Maximization of

$$\rho(W) = \frac{W^T S_B W}{W^T S_W W}$$

- given by solution of generalized eigenvalue problem

$$S_B W = \lambda S_W W$$

- The the c-class case we obtain c-1 projections as the largest eigenvalue of

$$S_B W_i = \lambda S_W W_i$$

LDA in the wild

- How does one calculate LDA for high-dimensional images?
- Problem: S_W is always singular
  - Number of pixels in an image is larger than number of images in training set
- Fisherfaces example: reduce dimensionality by doing a PCA first and then LDA
- Simultaneous diagonalization of S_W and S_B

Ficherfaces

- First published by Belhumeur et al 1997
- Reduce dimensionality to n-c with PCA

$$U_{PCA} = \arg\max_U |U^T Q U| = [\ u_1\ u_2 \ldots\ u_{n-c}\ ]$$

- Further reduce to c-1 with LDA

$$W_{LDA} = \arg\max_w \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|} = [\ w_1\ w_2\ \ldots\ w_{c-1}\ ]$$

- The optimal projection is then

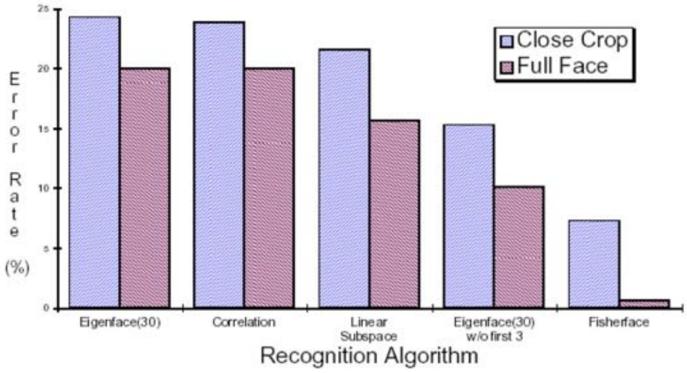
$$W_{opt} = W_{LDA}^T U^T$$

Example Fisherface

- Example Fisherface of recognition face w/wo glasses (Belhumeur et al, 1997)

Fisher example performance

- Small comparison of face recognition (old data)



- Significantly better performance than PCA for face recognition
- Noise sensitive
- Standard large scale Kaggle competitions today score 97%

Outline

Canonical Correlation Analysis (CCA)

- Also supervised method by motivated by regression / interpolation tasks such as pose estimation
- CCA related two sets of observations by determining pairs of directions that yield maximum correlation between the data sets
- Find a pair of directions (canonical factors): $w_x \in R^P$ *and* $w_y \in R^q$ so that the correlation of the projections $c = w_x^T x$ and $d = w_y^T y$ become maximal

CCA - the details

$$\rho = \frac{E[cd]}{\sqrt{E[c^2]\ E[d^2]}} =$$

$$\frac{E[w_x^T x\ y^t w_y]}{\sqrt{E[w_x^T x\ x^T w_x] E[w_y^T y\ y^t w_y]}} =$$

$$\frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}}$$

CCA - computations

- Finding solutions

$$w = \begin{bmatrix} w_x \\ w_y \end{bmatrix} \quad A = \begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \quad B = \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}$$

- Compute the Rayleigh Quotient

$$r = \frac{w^T A w}{w^T B w}$$

- Think of it as a generalized eigenvalue problem

$$Aw = \mu B w$$

CCA for images

- Same challenge as for LDA
- Computational analysis based on SVD

$$A = C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}}$$
$$A = UDV^T$$
$$w_{xi} = C_{xx}^{-\frac{1}{2}} u_i$$
$$w_{yi} = C_{yy}^{-\frac{1}{2}} v_i$$

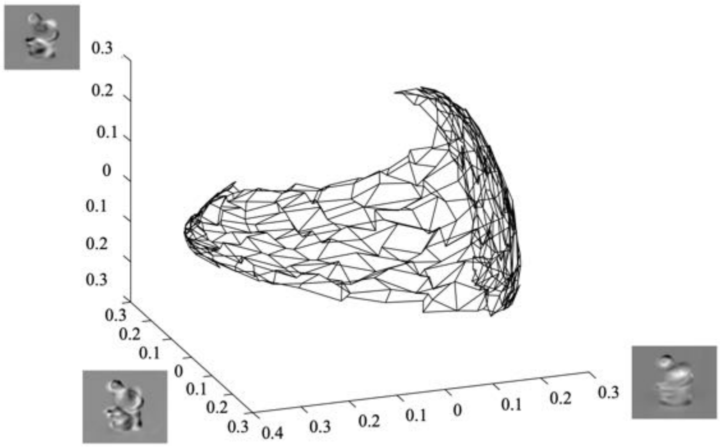
Properties of CCA

- At most min(p, q, n) CCA factors
- Invariant wrt to affine transformations
- Orthogonality of the canonical factors

$$w_{xi}^T C_{xx} w_{xj} = 0$$
$$w_{yi}^T C_{yy} w_{yj} = 0$$
$$w_{xi}^T C_{xy} w_{yj} = 0$$

CCA Example

- Parametric eigenspace obtained by PCA for 2 DOF pose space



Outline

Independent Component Analysis (ICA)

- ICA is a powerful technique from signal processing (blind source separation)
- Can we seen as an extension of PCA
- PCA takes statistics up to 2nd order into account
- ICA estimate components that are statistically independent
- Generates sparse/local descriptors - sparse coding

Independent Component Analysis (ICA)

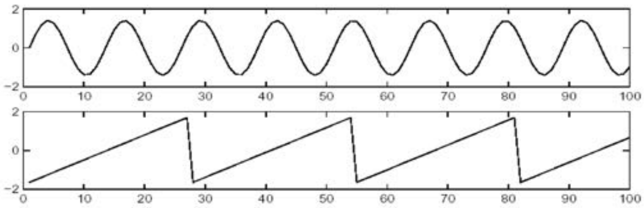
- m scalar variables - $X = (x_1, \ \ldots \ x_m)^T$
- Assumed to be a linear mixture of n sources - $S = (s_1, \ \ldots \ s_n)^T$

$$X = AS$$

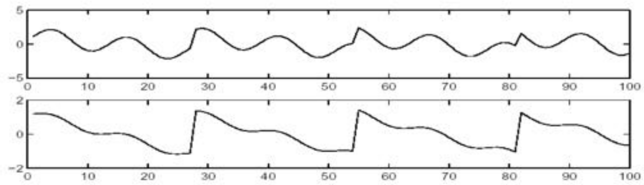
- Objective: Given X find estimates for A and S under the assumption S are independent

ICA Example

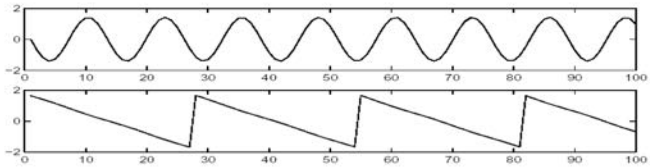
Original Sources



Mixtures

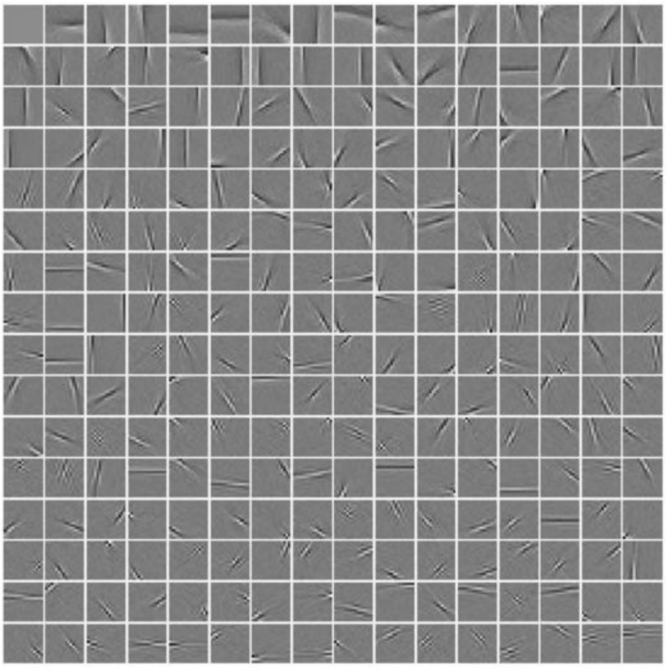


Recovered Sources



ICA Example

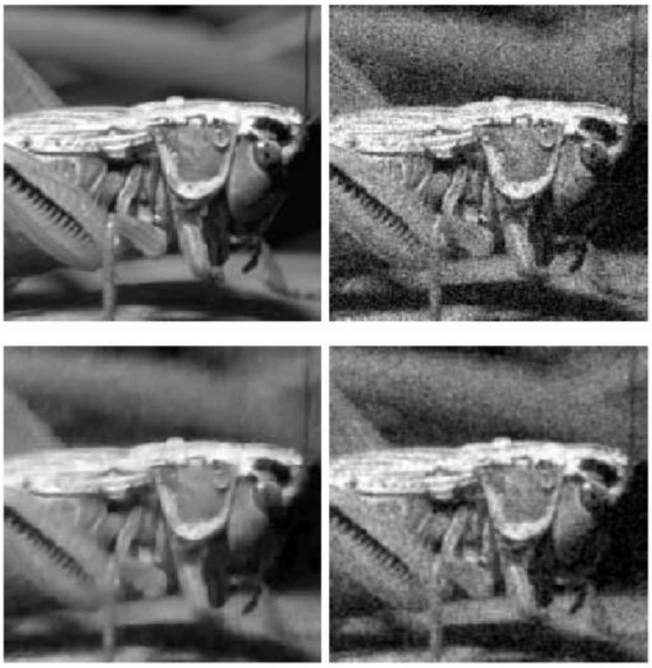
ICA basis obtained from 16x16 patches of natural images (Bell&Sejnowski 96)



ICA Algorithms

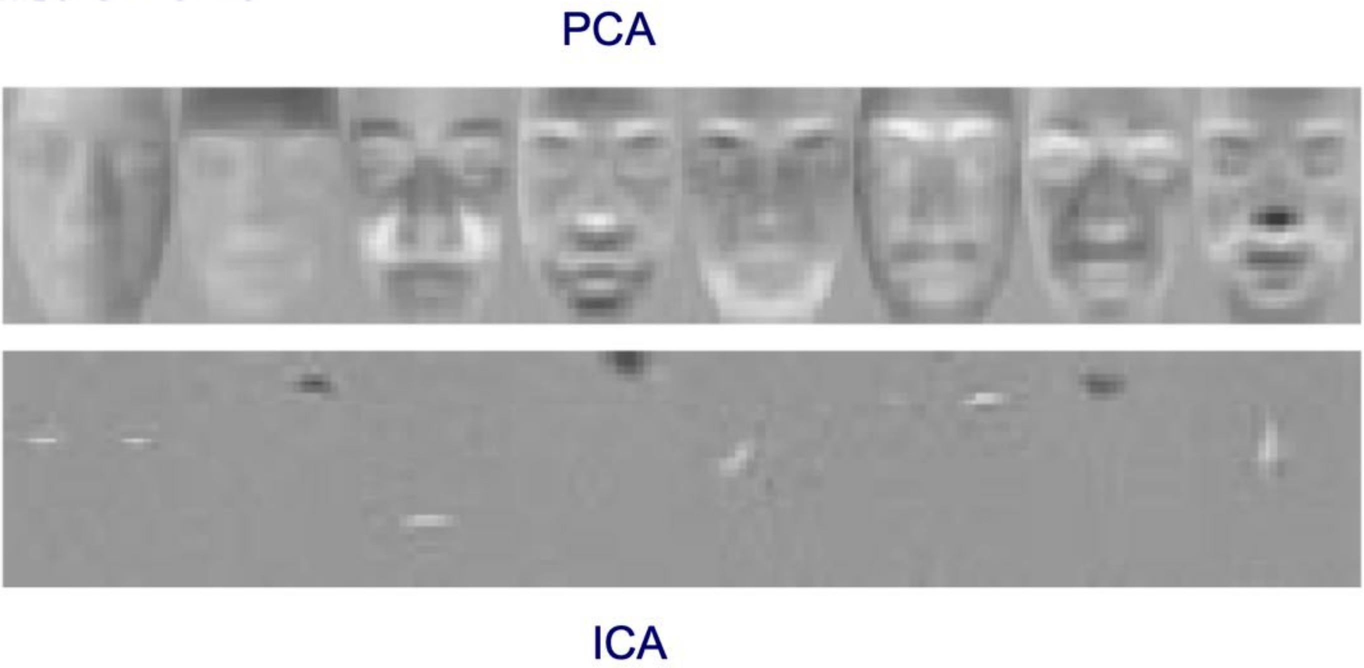
- Minimize a complex tensor function
- Adaptive algorithms based on stochastic gradient
  - Measure independence
  - Computer A recursively to maximize independence
- ICA only works for non-Gaussian sources
- Often whitening of data is performance
- ICA does not provide ordering
- ICA components are not orthogonal

ICA noise suppression example



Example from Hyvärinen, 1999

PCA vs ICA for face recognition



From Baek et al, 2002

Outline

Summary

- Brief overview of use of sub-space methods for data processing
- The exact task should dictate the choice of methods
- Other cascaded processing simplifies complexity
- Good standard tools available in most signal processing toolboxes

Questions

Questions