Advanced Component Analysis Methods for Signal Processing

Carnegie Mellon
THE ROBOTICS INSTITUTE

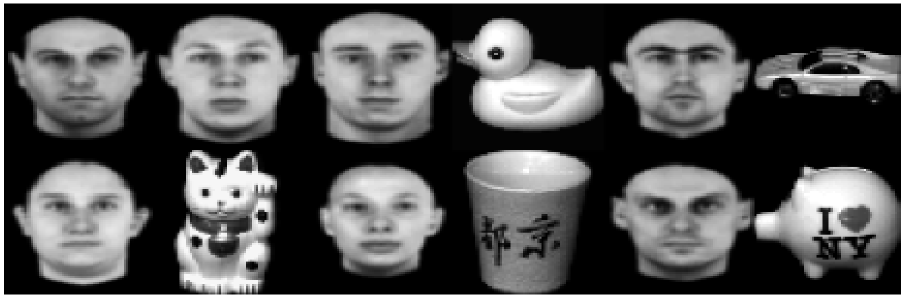
Fernando De la Torre (ftorre@cs.cmu.edu)

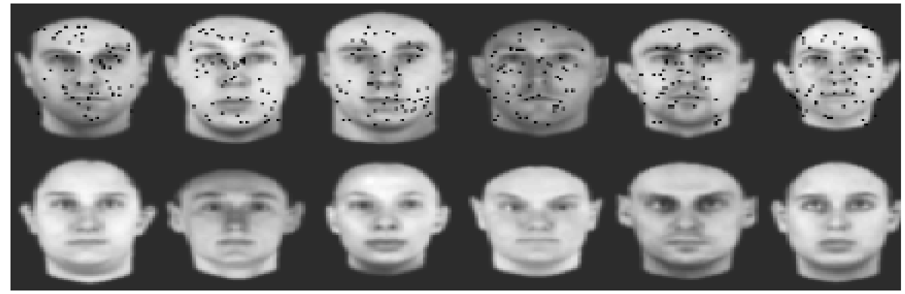
Outline

- Introduction
- Principal component analysis (PCA)
  - Robust Principal Component Analysis (RPCA)
  - PCA with missing data.
  - Parameterized Principal Component Analysis (PaPCA)
- Linear discriminant analysis (LDA)
  - Discriminative cluster analysis (DCA).
- K-means
  - Hierarchical aligned cluster analysis (HACA).
- Canonical correlation analysis (CCA)
  - Canonical time warping (CTW)
  - Dynamic coupled component analysis (DCCA)

1-Robust PCA

- Two types of outliers:

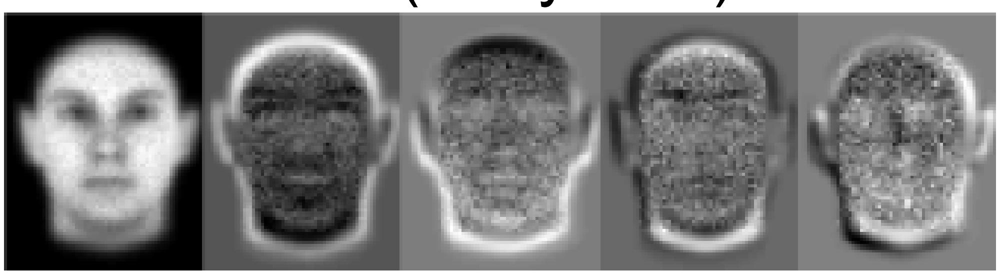


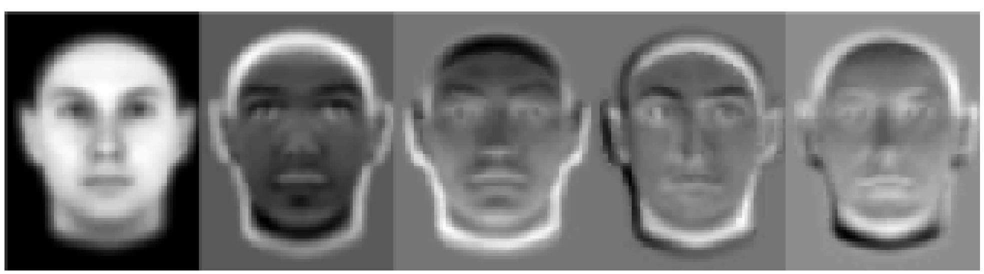


Sample outliers
(Xu & Yuille., 1995)

Intra-sample outliers
(de la Torre & Black, 2001b; Skocaj & Leonardis, 2003)

- Standard PCA solution (noisy data):





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)

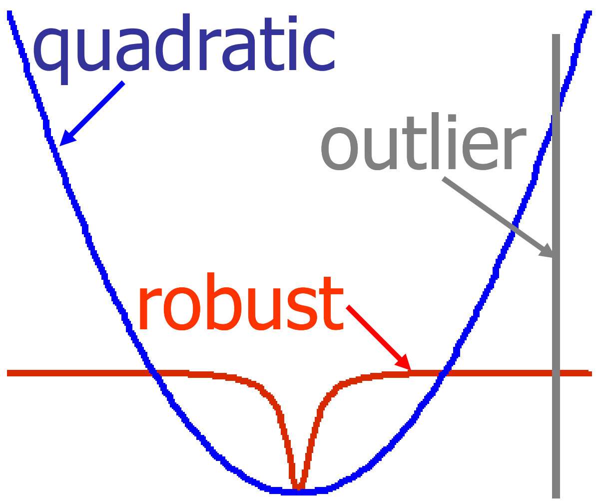
$$\left(\mathbf{C}\mathbf{C}^T\right)^{-1}\mathbf{R} = \mathbf{B}^T\mathbf{B}\mathbf{R}\Lambda^{-1}$$

Robust PCA

- Using robust statistics:
  (de la Torre & Black, 2001b; de la Torre & Black, 2003a)

Pixel residual

$$E_{rpca}(\mathbf{B},\mathbf{C},\boldsymbol{\mu}) = \sum_{i=1}^{n}\sum_{p=1}^{d}\rho\Big(d_{pi} - \mu_p - \sum_{j=1}^{k} b_{pj}c_{ji}, \sigma_p\Big)$$



mean

Basis (**B**) &
Coefficients(**c**)

Numerical Problems

- No closed form solution in terms of an eigen-equation.
- Deflation approaches do not hold.

$$\mathbf{A}' = \mathbf{A} - \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\mathrm{T}}$$

First eigenvector with highest eigenvalue.

$$\mathbf{A}'' = \mathbf{A}' - \lambda_2 \mathbf{u}_2 \mathbf{u}_2^{\mathrm{T}}$$

Second eigenvector with highest eigenvalue.

$$\cdots$$

- In the robust case all the basis have to be computed simultaneously (including the mean).

How to Optimize it?

$$E_{rpca}(\mathbf{B},\mathbf{C},\boldsymbol{\mu}) = \sum_{i=1}^{n}\sum_{p=1}^{d}\rho(d_{pi} - \mu_p - \sum_{j=1}^{k} b_{pj}c_{ji}, \sigma_p)$$

- Normalized Gradient descent

$$\mathbf{B}^{n+1} = \mathbf{B}^{n} - [\mathbf{H}_b]^{-1} \circ \frac{\partial E_{rpca}}{\partial \mathbf{B}} \qquad \mathbf{H}_{\mathbf{b}} = \max \; diag\left(\frac{\partial^2 E_{rpca}}{\partial \mathbf{b}_i \partial \mathbf{b}_i^T}\right)$$

$$\mathbf{C}^{n+1} = \mathbf{C}^{n} - [\mathbf{H}_c]^{-1} \circ \frac{\partial E_{rpca}}{\partial \mathbf{C}} \qquad \mathbf{H}_{\mathbf{c}} = \max \; diag\left(\frac{\partial^2 E_{rpca}}{\partial \mathbf{c}_i \partial \mathbf{c}_i^T}\right)$$

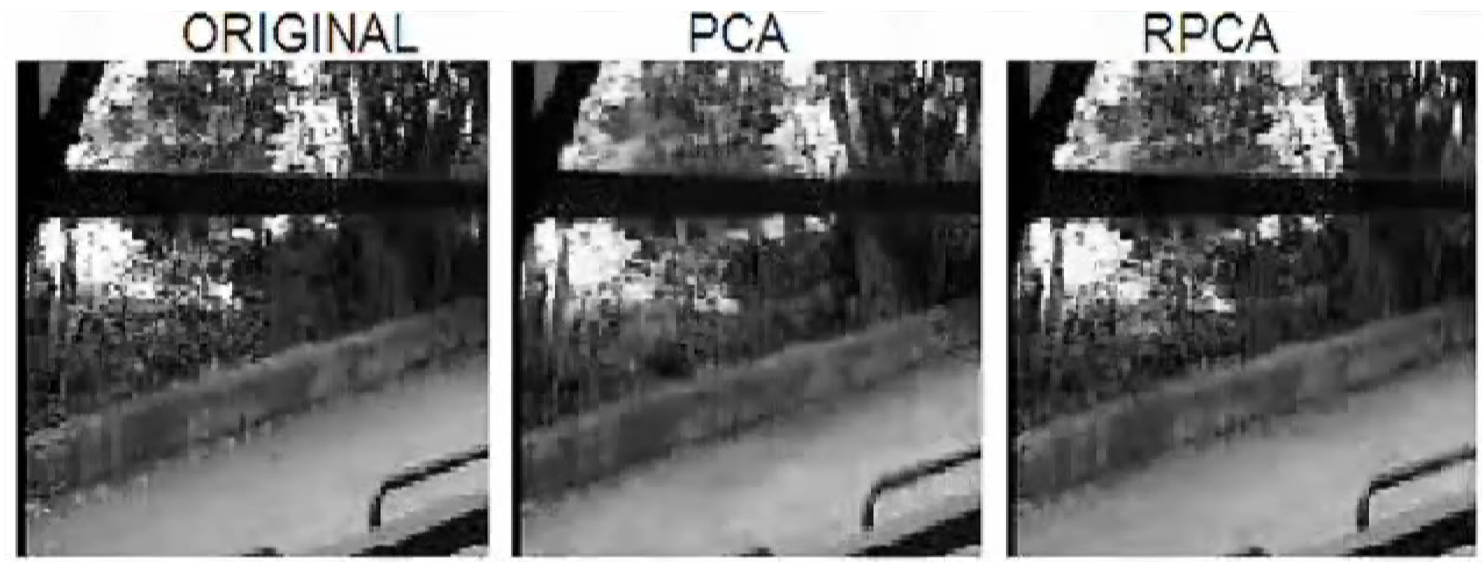
- Deterministic annealing methods to avoid local minima.

(Blake & Zisserman, 1987)

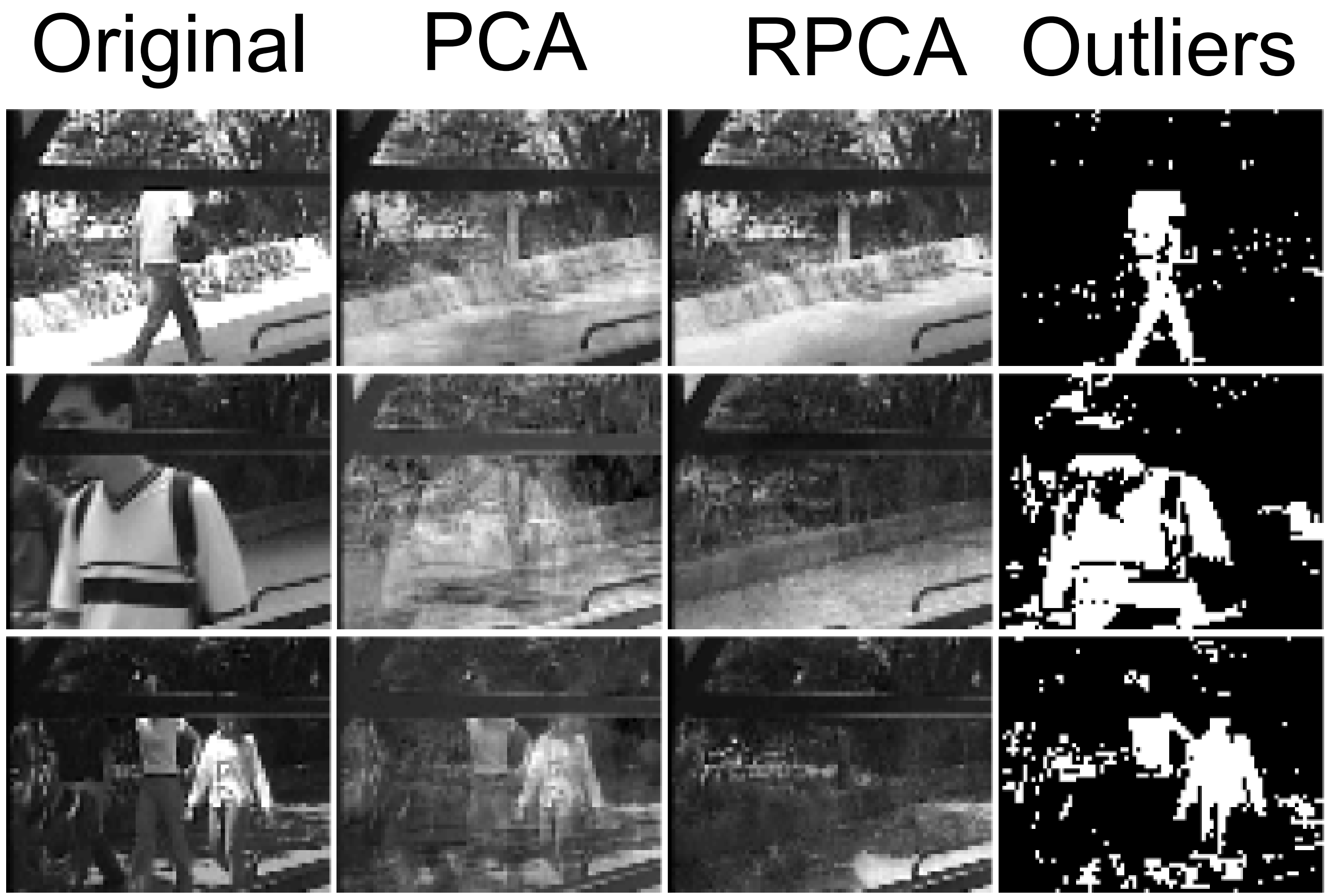
Example

Statistical outlier

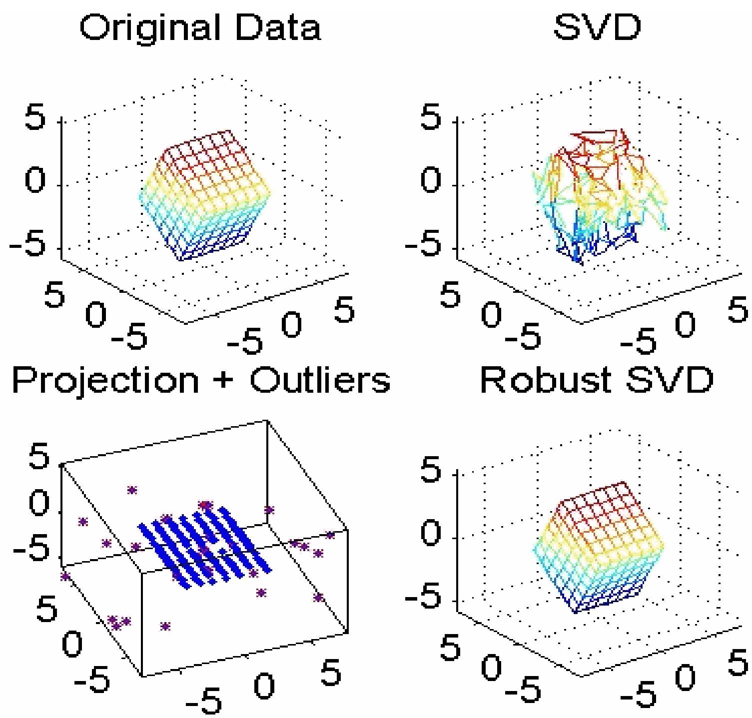
- Small region
- Short amount of time



Robust PCA



Structure from Motion



More work on Robust PCA

- Robust estimation of coefficients (Black & Jepson, 1998; Leonardis & Bischof, 2000; Ke & Kanade, 2004)
- Robust estimation of basis and coefficients (Gabriel & Odoro, 1984; Croux & Filzmoser., 1981; Skocaj et al., 2002;Skocaj & Leonardis, 2003; de la Torre & Black, 2001b; de la Torre & Black, 2003a)
- Other Robust PCA techniques (sample outliers) (Campbell, 1980; Ruymagaart, 1981; Xu & Yuille., 1995)

Outline

- Introduction
- Principal component analysis (PCA)
  - Robust Principal Component Analysis (RPCA)
  - PCA with uncertainty and missing data.
  - Parameterized Principal Component Analysis (PaPCA)
- Linear discriminant analysis (LDA)
  - Discriminative cluster analysis (DCA).
- K-means
  - Hierarchical aligned cluster analysis (HACA).
- Canonical correlation analysis (CCA)
  - Canonical time warping (CTW)
  - Dynamic coupled component analysis (DCCA)

2- PCA with Uncertainty and Missing Data

- Adding uncertainty $E_2(\mathbf{B},\mathbf{C}) = \|\mathbf{W} \circ (\mathbf{D} - \mathbf{BC})\|_F = \sum_{i=1}^{d}\sum_{j=1}^{n} w_{ij}(d_{ij} - \sum_{s=1}^{k} b_{is}c_{sj})^2$

- If weights are separable $\mathbf{W} = \mathbf{w}_r \mathbf{w}_c^T$ closed-form solution.

$$\mathbf{w}^c = \begin{pmatrix} w_1^c & w_2^c & \cdots & w_n^c \end{pmatrix}$$

$$\mathbf{w}^r = \begin{pmatrix} w_1^r \\ w_2^r \\ \cdots \\ w_d^r \end{pmatrix} \qquad \begin{pmatrix} d_{11} & \cdots & \cdots & d_{1n} \\ d_{21} & \ddots & & d_{2n} \\ \vdots & & \ddots & \vdots \\ d_{d1} & \cdots & \cdots & d_{dn} \end{pmatrix} = \mathbf{D}$$

$\mathbf{W} \in \Re^{d\times n}$

$w_{ij} \geq 0$

$\circ$ *Hadamard product*

- Generalized SVD

(Greenacre, 1984; Irani & Anandan, 2000;)

General Case

- For arbitrary weights no closed-form solution.

$$E_2(\mathbf{B},\mathbf{C}) = \left\| \mathbf{W} \circ (\mathbf{D} - \mathbf{BC}) \right\|_F = \sum_{i=1}^{n} (\mathbf{d}_i - \mathbf{Bc}_i)^T diag(\mathbf{w}_i)(\mathbf{d}_i - \mathbf{Bc}_i) =$$

$$\sum_{p=1}^{d} (\mathbf{d}^p - \mathbf{C}^T\mathbf{b}^p)^T diag(\mathbf{w}^p)(\mathbf{d}^p - \mathbf{C}^T\mathbf{b}^p)$$ (Torre & Black, 2003a)

- Alternated least squares algorithms
  – Slow convergence, easy implementation.
- Damped Newton Algorithm
  – Fast convergence. (Buchanan & Fitzgibbon., 2005)

$$E_2(\mathbf{B},\mathbf{C}) = \left\| \mathbf{W} \circ (\mathbf{D} - \mathbf{BC}) \right\|_F + \lambda_1 \|\mathbf{B}\|_F + \lambda_2 \|\mathbf{C}\|_F$$

$$\mathbf{v} = \begin{bmatrix} vec(\mathbf{B}) \\ vec(\mathbf{C}) \end{bmatrix} \quad \mathbf{v}^{(n+1)} = \mathbf{v}^n - \left[\frac{[\partial^2 E_2]}{\partial^2 \mathbf{v}} \right]^{-1} \frac{\partial E_2}{\partial \mathbf{v}}$$

– **H** definite positive: $\mathbf{H} = \frac{\partial^2 E_2}{\partial^2 \mathbf{v}} + \lambda \mathbf{I}$

$repeat$

$$\mathbf{H} = \frac{\partial^2 E_2}{\partial^2 \mathbf{v}} \quad \mathbf{g} = \frac{\partial E_2}{\partial \mathbf{v}}$$

$repeat$

$\lambda = 10\lambda$

$\mathbf{y} = \mathbf{x} - (\mathbf{H} + \lambda I)^{-1}\mathbf{g}$

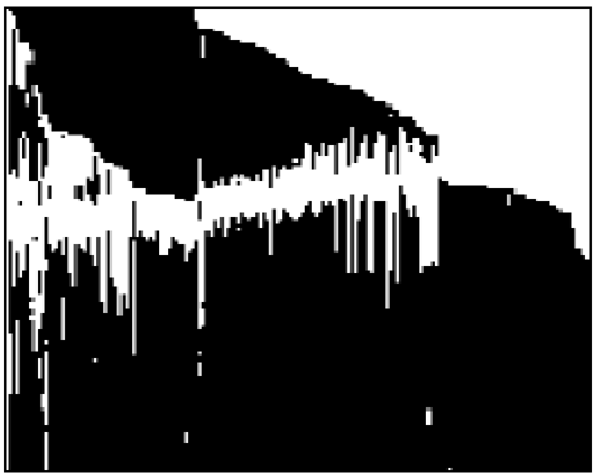
$until \;\; F(\mathbf{y}) < F(\mathbf{x})$

$\mathbf{x} = \mathbf{y}; \lambda = \frac{\lambda}{10}$

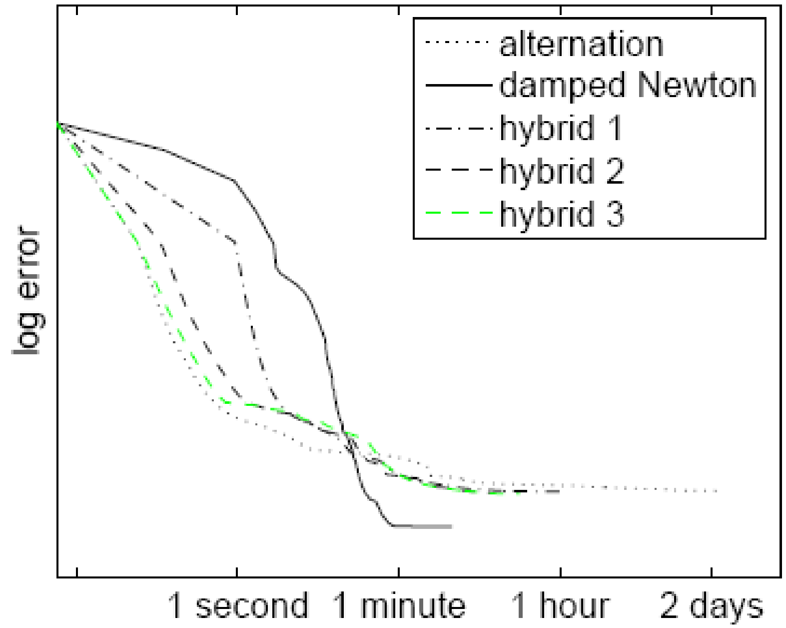
$until \;\; convergence$

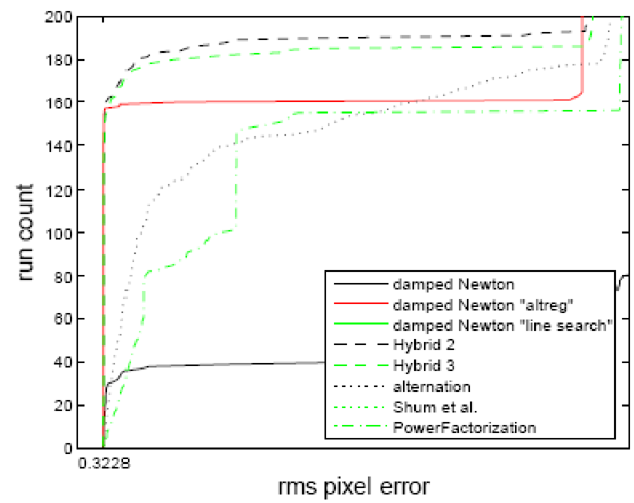
Experiments

240×167: 70% known



Total: 500 runs





Related work

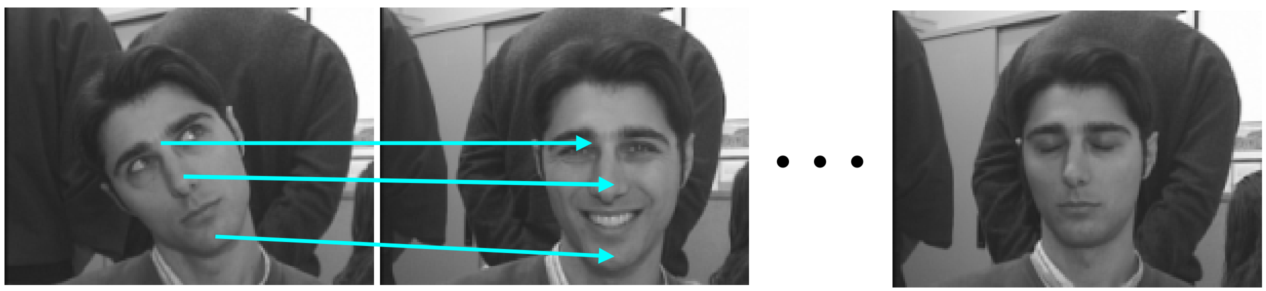
- Iterative (Wiberg, 1976; Shum et al., 1995; Morris & Kanade, 1998; Aans et al.,2002; Guerreiro & Aguiar, 2002)
- Closed-form (Aguiar & Moura, 1999; Irani & Anandan, 2000)
- Power factorization (Hartley & Schaalitzky, 2003)
- Bayesian estimation (L.Torresani & Bregler, 2004)

Outline

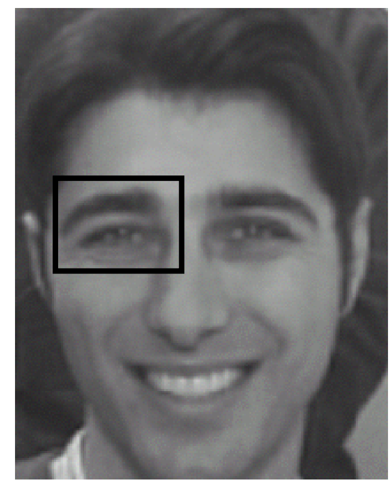
- Introduction
- Principal component analysis (PCA)
  - Robust Principal Component Analysis (RPCA)
  - PCA with uncertainty and missing data.
  - Parameterized Principal Component Analysis (PaPCA)
- Linear discriminant analysis (LDA)
  - Discriminative cluster analysis (DCA).
- K-means
  - Hierarchical aligned cluster analysis (HACA).
- Canonical correlation analysis (CCA)
  - Canonical time warping (CTW)
  - Dynamic coupled component analysis (DCCA)

3- Parameterized Component Analysis (PaCA) (de la Torre & Black, 2003b)

- Learn a subspace invariant to geometric transformations?



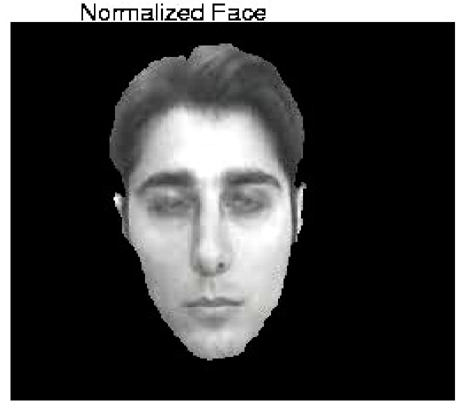
- Data has to be **geometrically** normalized
  - Tedious manual cropping.
  - Inaccuracies due to matching ambiguities.
  - Hard to achieve sub-pixel accuracy.



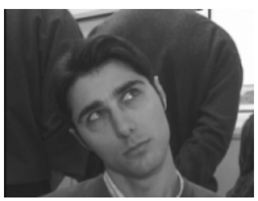
Error function for PaCA

Original Data

Normalized Face



$$E(\mathbf{B},\mathbf{C},\mathbf{a}) = \sum_{t=1}^{T} \left\| \mathbf{d}_t(\mathbf{f}(\mathbf{x},\mathbf{a}_t)) - \mathbf{B}\mathbf{c}_t \right\|_{\mathbf{W}_1}^2 + p_1(\mathbf{a}) + p_2(\mathbf{c})$$



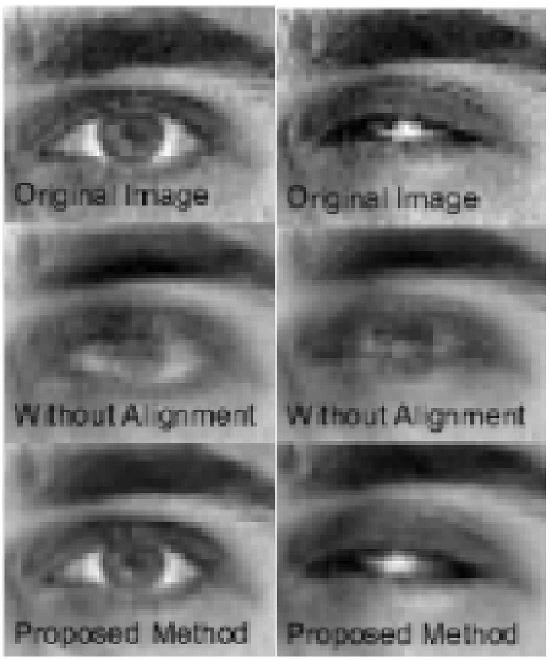
Motion (warping)

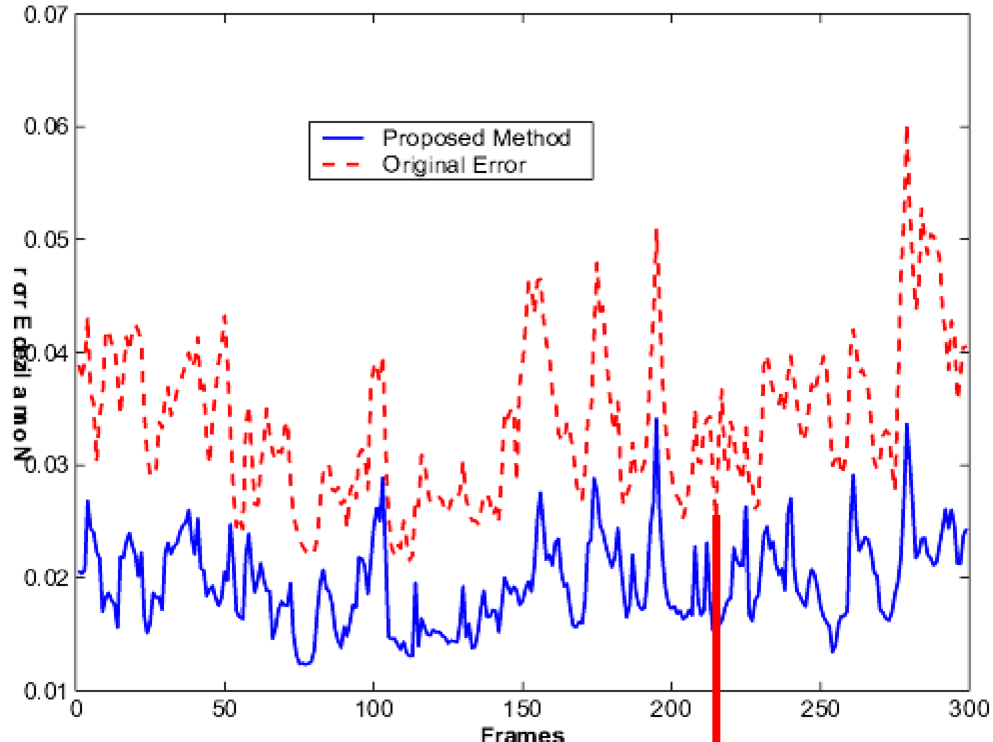
Basis (**B**) & coefficients (**c**)

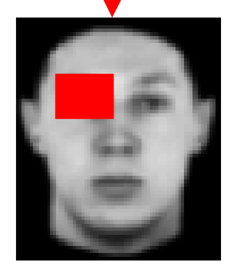
Regularization

$$\sum_{t=1}^{T}\sum_{l=1}^{L} \lambda_1 \left\| \mathbf{c}_t - \mathbf{\Gamma}_c \mathbf{c}_{t-1} \right\|_{\mathbf{W}2}^2 + \lambda_2 \left\| \mathbf{a}_t - \mathbf{\Gamma}_a \mathbf{a}_{t-1} \right\|_{\mathbf{W}3}^2$$

EigenEye Learning







More on Parameterized CA

- Probabilistic model
  - Search scales exponentially with the number of motion parameters (Frey & Jojic, 1999a; Frey & Jojic, 1999b; Williams & Titsias, 2004)
- Other continuous approaches.

  (Schewitzer, 1999; Rao, 1999; Shashua et al., 2002)
- Invariant clustering

  (Fitzgibbon & Zisserman, 2003)
- Non-rigid motion

  (Baker et al., 2004)
- Invariant recognition

  (Black & Jepson, 1998)
- Invariant support vector machines (Avidan, 2001)
- Parameterized Kernel Component Analysis (De la Torre, 2008)

Outline

- Introduction
- Principal component analysis (PCA)
  - Robust Principal Component Analysis (RPCA)
  - PCA with uncertainty and missing data.
  - Parameterized Principal Component Analysis (PaPCA)
- Linear discriminant analysis (LDA)
  - Discriminative cluster analysis (DCA).
- K-means
  - Hierarchical aligned cluster analysis (HACA).
- Canonical correlation analysis (CCA)
  - Canonical time warping (CTW)
  - Dynamic coupled component analysis (DCCA)

Problems of k-means

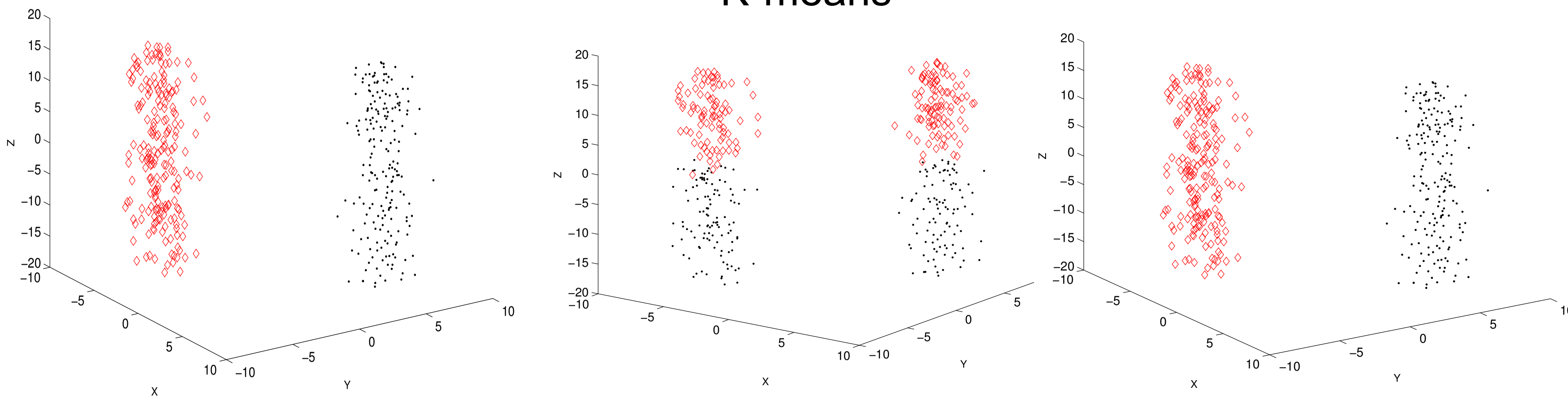
- K-means optimizes:

$$E(\mathbf{A},\mathbf{B}) = \| \mathbf{D} - \mathbf{B}\mathbf{A}^T \|_F = \sum_{i=1}^{c} \sum_{j \in C_i} \| \mathbf{d}_j - \mathbf{a}_i \|$$

  - Not efficient for high dimensional data.
  - Multiple local minima.
  - No mechanism to remove irrelevant features for clustering

K-means

DCA



Discriminative cluster analysis (DCA)

- LDA optimizes:

$$E(\mathbf{A},\mathbf{B}) = \| (\mathbf{G}^T\mathbf{G})^{-\frac{1}{2}}(\mathbf{G}^T - \mathbf{B}\mathbf{A}^{\mathbf{T}}\mathbf{D}) \|_F$$

Supervised

$$\mathbf{G}^T = \begin{bmatrix} 1 & \dots & 0 \\ 0 & \dots & 1 \\ 0 & \dots & 0 \end{bmatrix}$$

Unsupervised

- DCA optimizes:

$$E(\mathbf{A},\mathbf{B},\mathbf{G}) = \| (\mathbf{G}^T\mathbf{G})^{-\frac{1}{2}}(\mathbf{G}^T - \mathbf{B}\mathbf{A}^{\mathbf{T}}\mathbf{D}) \|_F$$

$$g_{ij} \in \{0,1\} \qquad \mathbf{G}\mathbf{1}_c = \mathbf{1}_n$$

- In LDA **G** is known because it is a supervised technique. In DCA **G** is estimated.
- Simultaneous dimensionality reduction and clustering.

Optimization

- Eliminate **A**

S_t S_b

$$E(\mathbf{B},\mathbf{G}) \propto tr((\mathbf{B}^T\mathbf{D}\mathbf{D}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{D}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{D}^T\mathbf{B}))$$

- Optimize for **B**

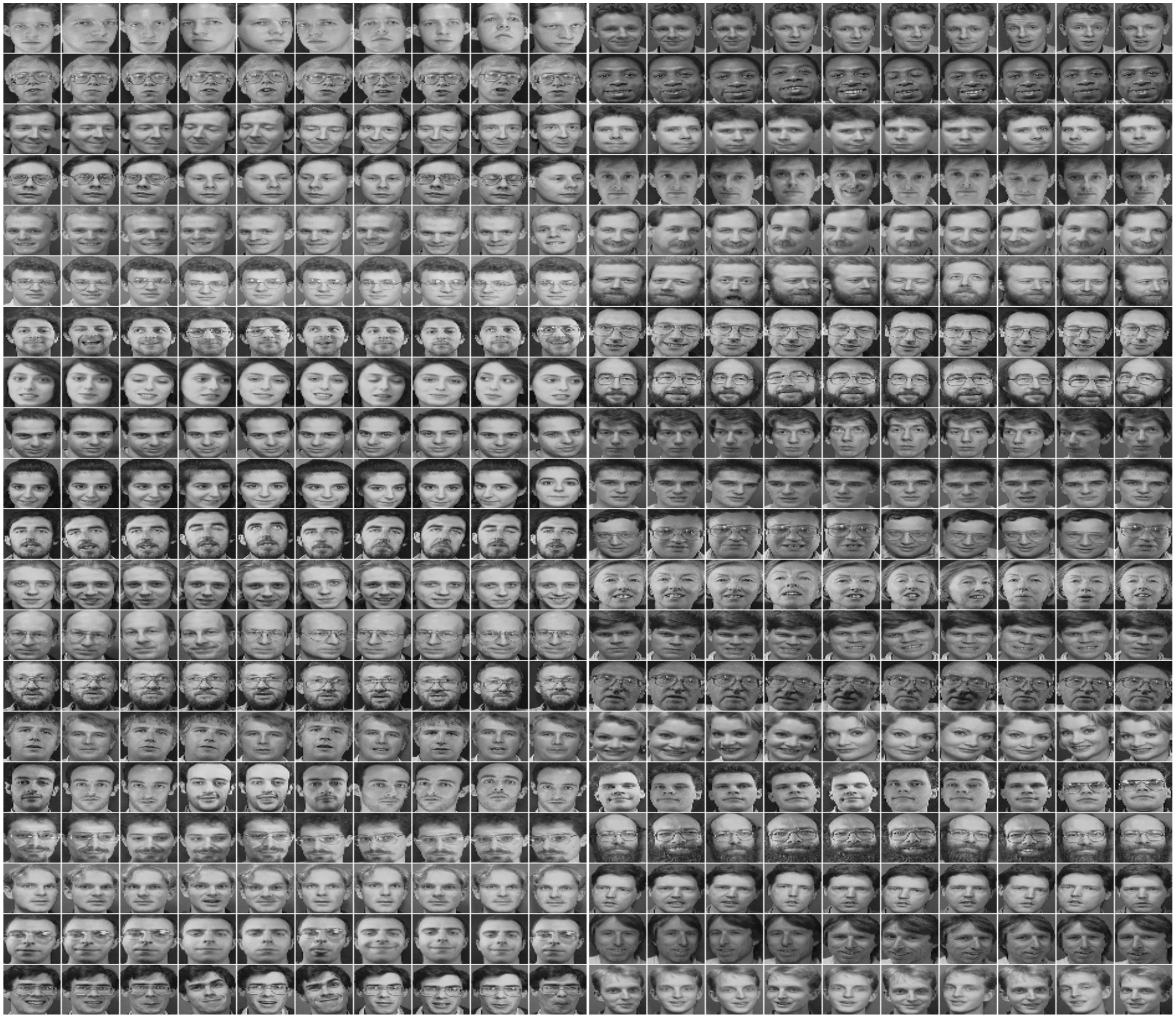
$$\mathbf{D}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{D}^T\mathbf{B} = \mathbf{D}\mathbf{D}^T\mathbf{B}\Lambda$$

- Optimize for **G** $\quad \mathbf{A} = \mathbf{C}^T(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C} \quad \mathbf{C} = \mathbf{B}^T\mathbf{D}$

$$\mathbf{G} = \mathbf{V} \circ \mathbf{V} \qquad \mathbf{V}^{(n+1)} = \mathbf{V}^{(n)} - \eta\frac{\partial E}{\partial \mathbf{V}}$$

$$\frac{\partial E}{\partial \mathbf{V}} = (\mathbf{I}_C - \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T)\mathbf{A}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}$$

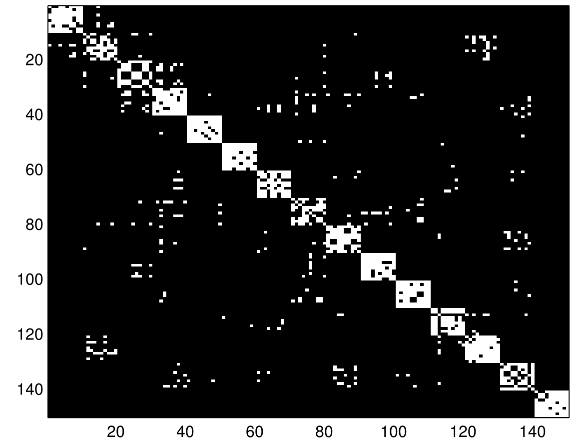
Clustering faces



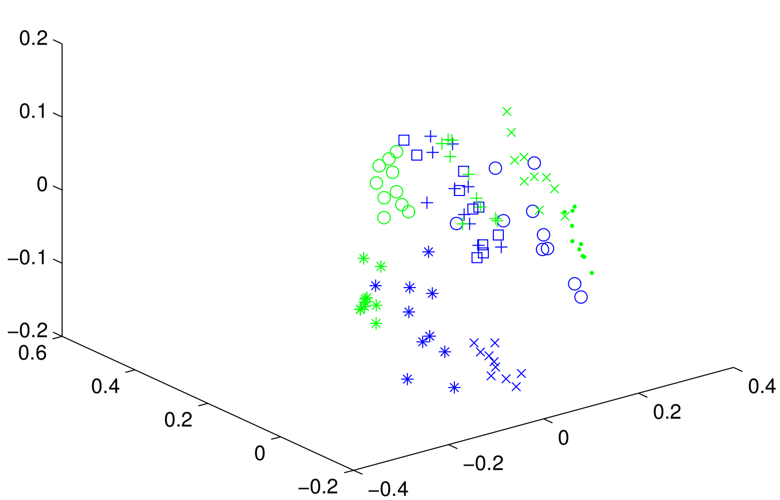
92x112 pixels.
40 people
10 samples/person

Experiments

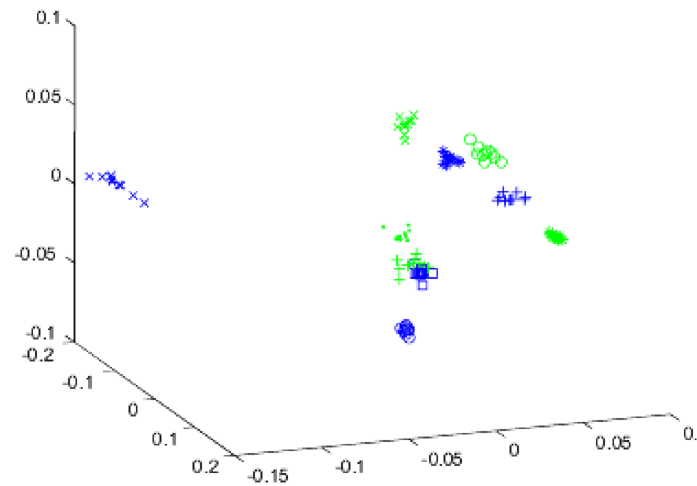
$$\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T \approx$$



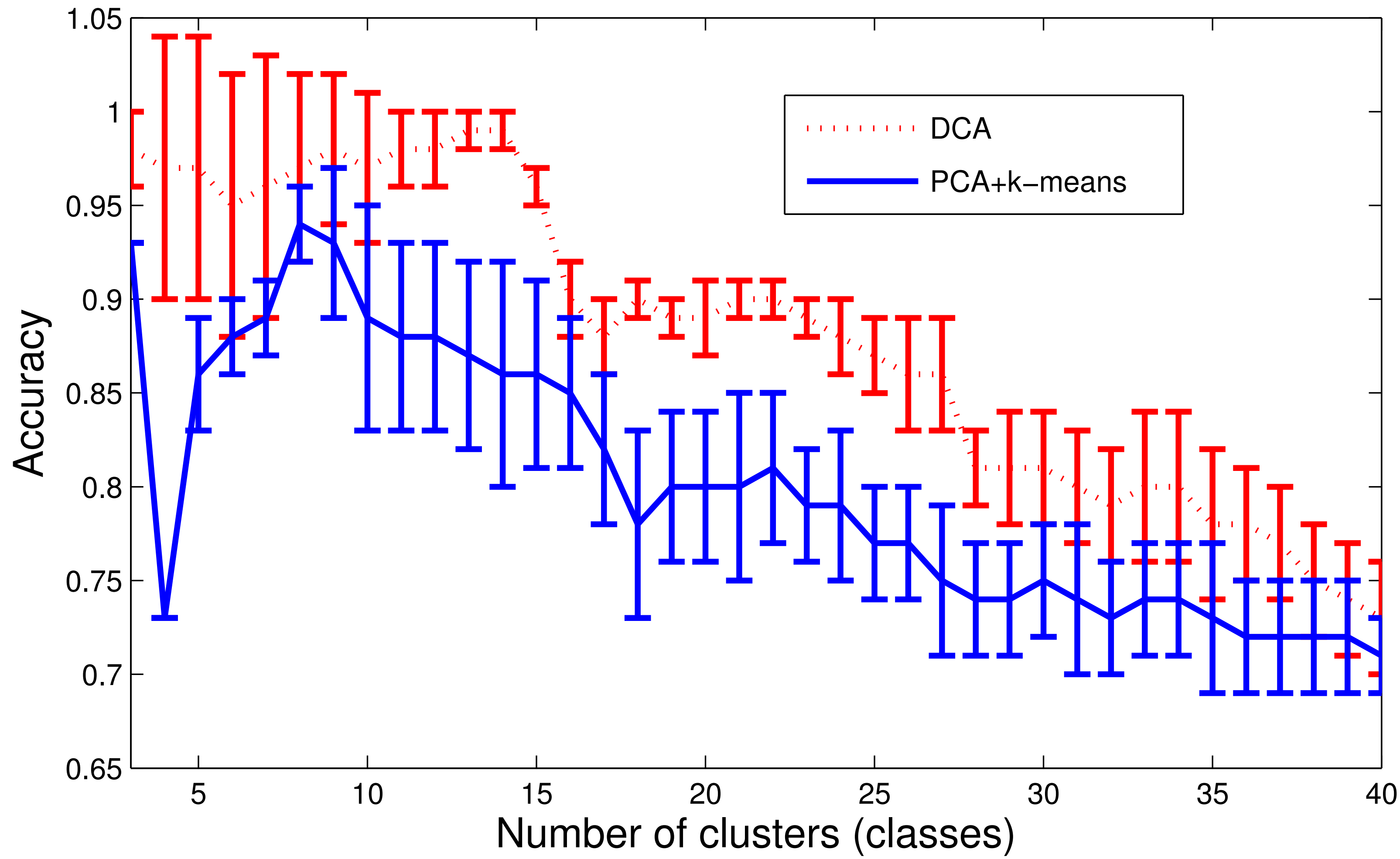
PCA

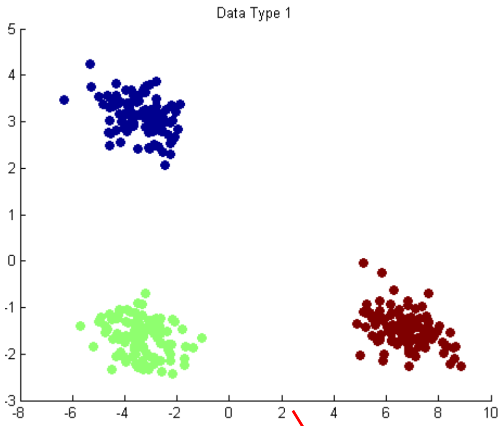


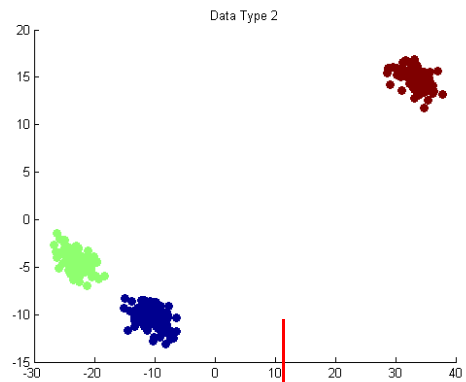
DCA

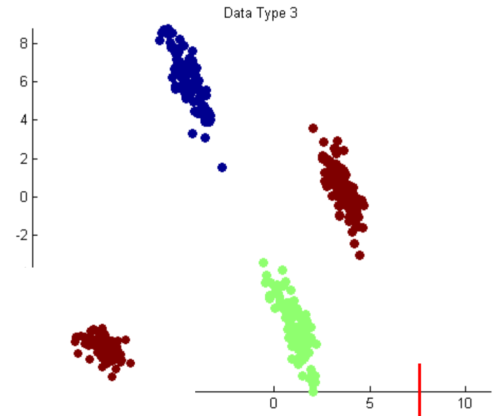


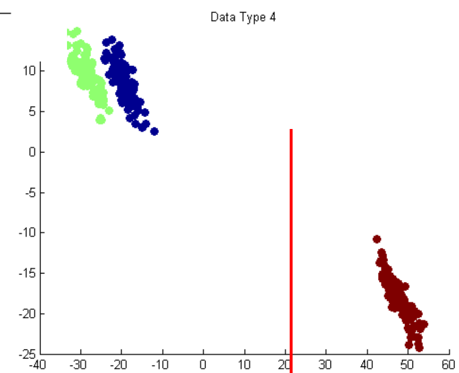
DCA vs. PCA+k-means











1000 realizations

	Datatype			
	1	2	3	4
k-Means	53.57%	38.22%	41.50%	11.88%
DCA	84.71%	82.11%	86.50%	85.55%

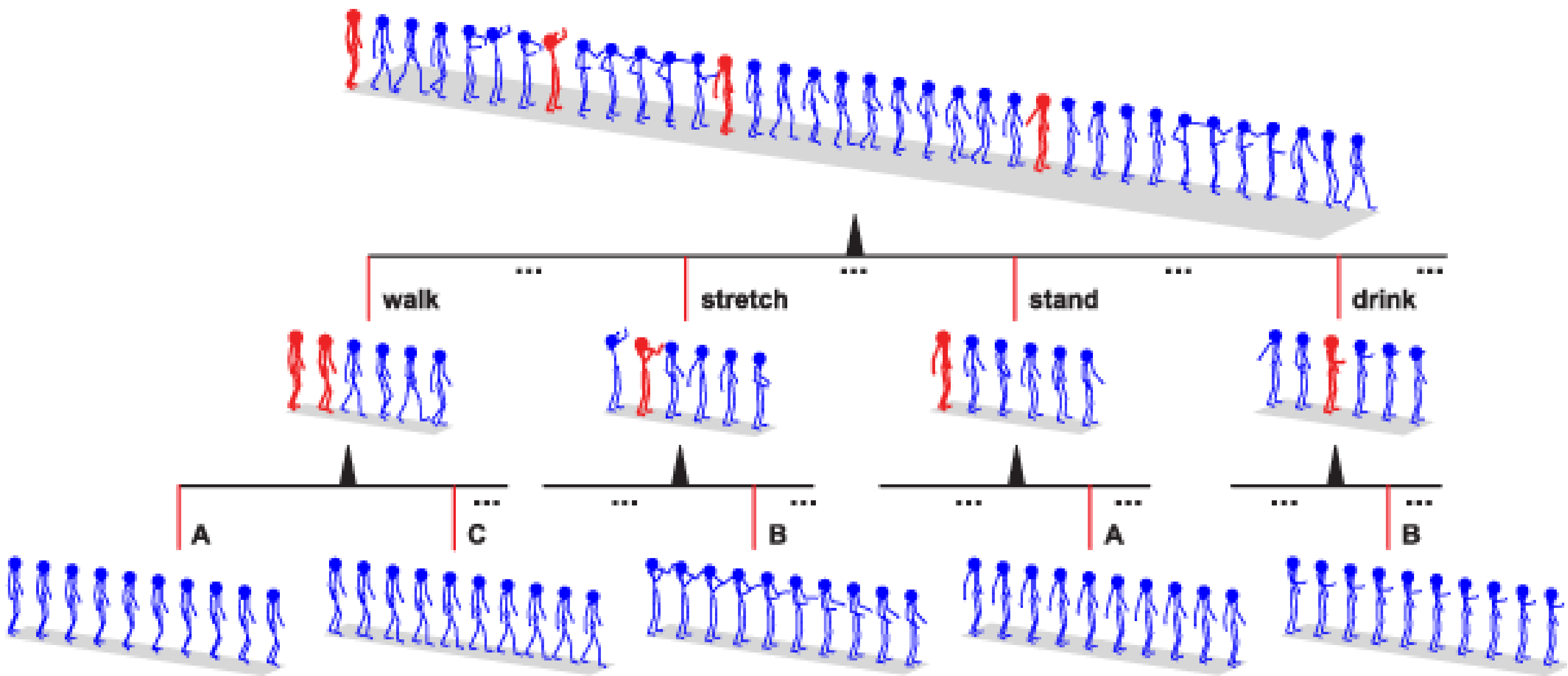
Outline

- Introduction
- Principal component analysis (PCA)
  - Robust Principal Component Analysis (RPCA)
  - PCA with uncertainty and missing data.
  - Parameterized Principal Component Analysis (PaPCA)
- Linear discriminant analysis (LDA)
  - Discriminative cluster analysis (DCA).
- K-means
  - Hierarchical aligned cluster analysis (HACA).
- Canonical correlation analysis (CCA)
  - Canonical time warping (CTW)
  - Dynamic coupled component analysis (DCCA)

Problem I

(Learning motion primitives)

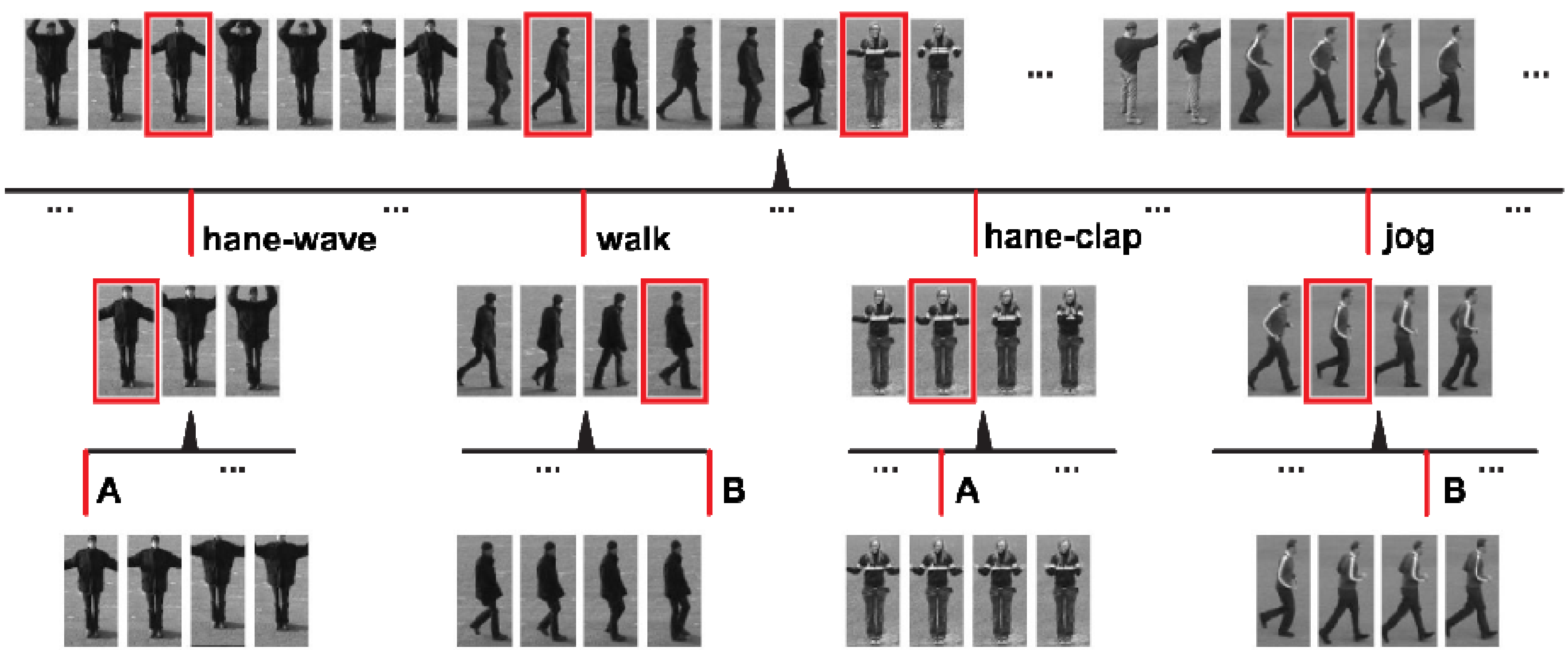
- How to factorize a stream of motion capture data into motion primitives?



Problem II

(Unsupervised segmentation of actions in video)

- How to factorize a stream of video of different people into actions?



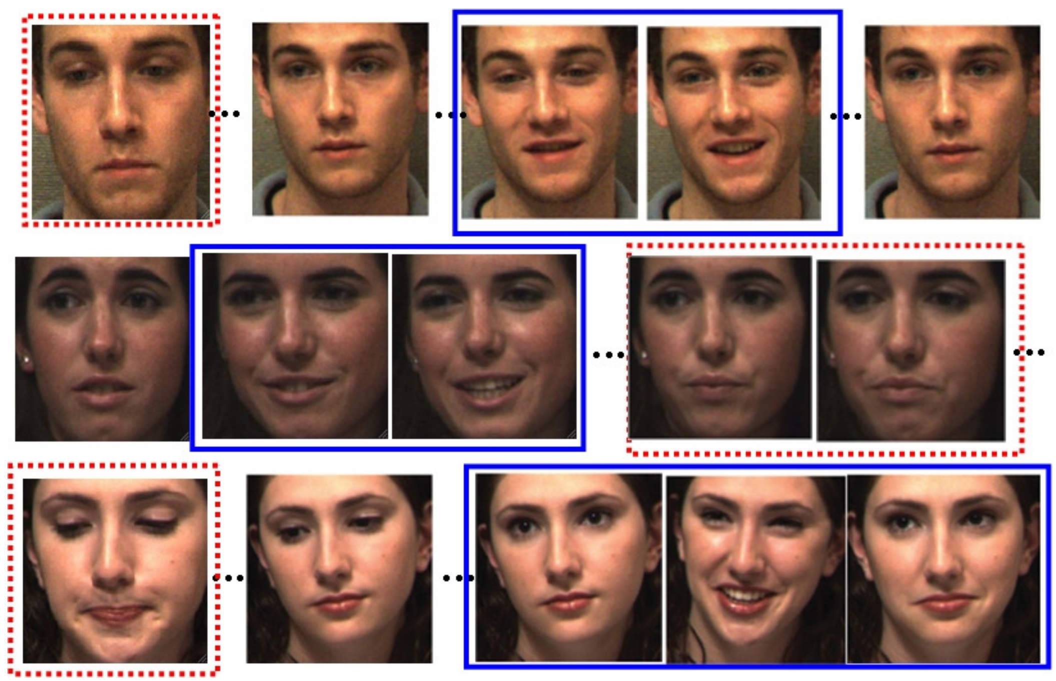
Problem III

(Learning Facial Action Coding System)

- How to learn a vocabulary for facial expressions?

Outer Brow Raiser **(AU2)** Brow lowerer **(AU4)** Upper lid raiser **(AU5)** Nose Wrinkler **(AU9)**

Lip corner puller **(AU12)** Chin raiser **(AU17)** Lip stretcher **(AU20)** Dimpler **(AU14)** Lip corner depressor **(AU15)**

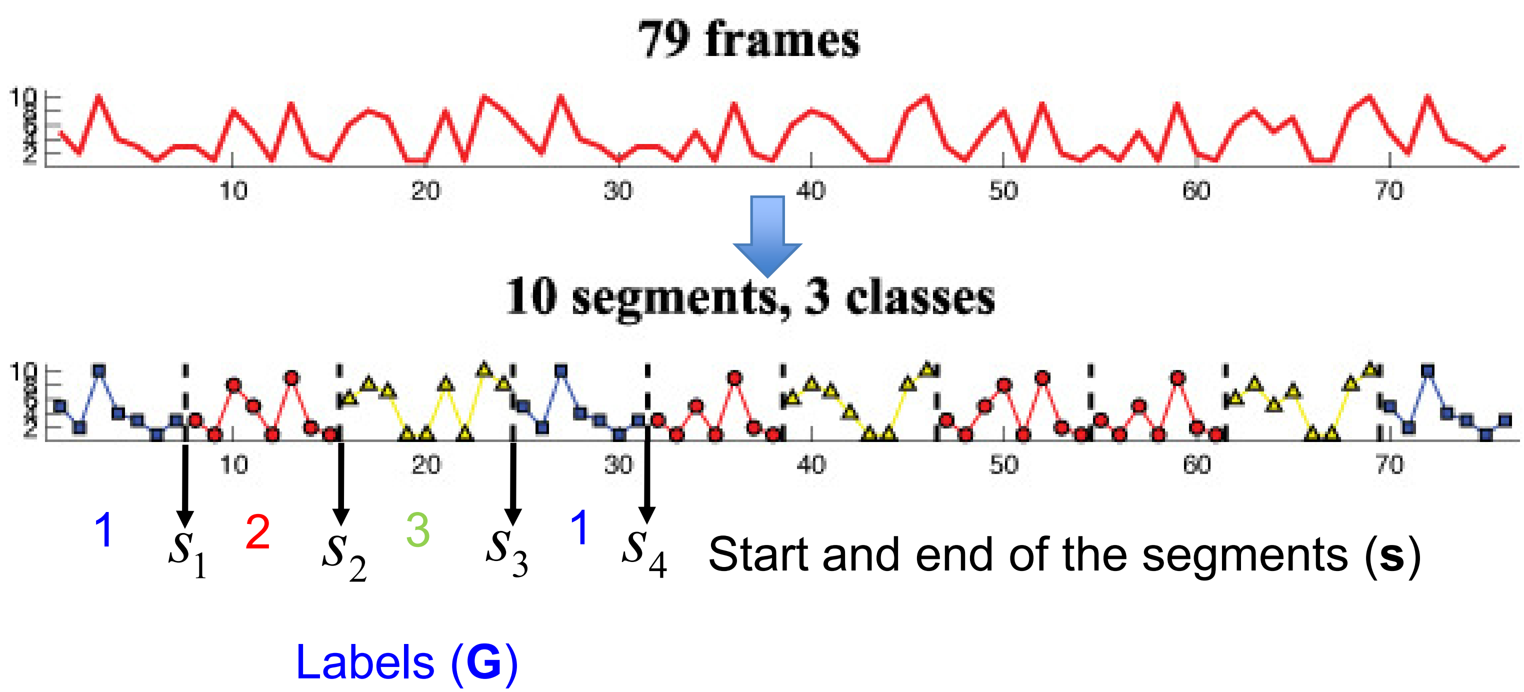


Previous Work

- **Change point detection** (Fearnhead 06, *Murphy et al. 07, Harchaoui et al. 09,* Boysen *et al. 09,)*
- **Bayesian networks:**
  - **HMMs** *(Kohlmorgen et al. 01)* , **switching LDS** *(Oh et al. 08, Fox et al. 08), **MM Markov Networks*** (Xu 06).
- **PCA, GMM, spectral Clustering** *(Barbic et al.04, Irani et al. 01, De la Torre et al. 07)*
- **Other methods: Zero-velocity detection** (Jenkins et al. 02)**, Frequency analysis** (Davis et al. 00)
- ***ACA***
  - ***Extension on kernel k-means to cluster time series.***
  - ***Efficient solution with dynamic programming.***

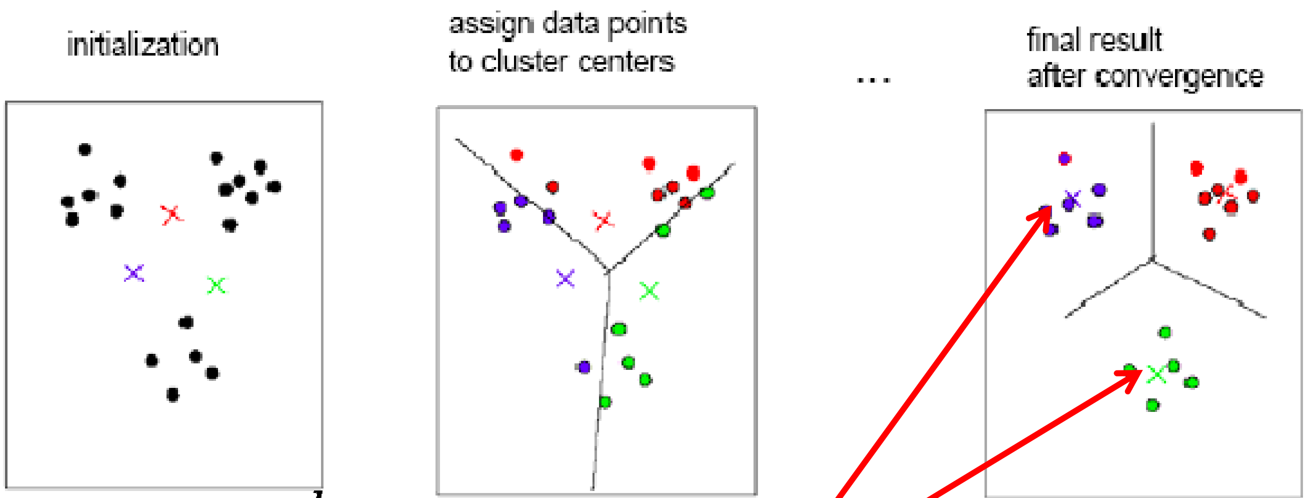
Problem Formulation for ACA

- Given a sequence $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \Re^{d \times n}$, Aligned Cluster Analysis (ACA) decompose **X** into ***m*** disjoint segments belonging to one of ***k*** classes.



k-means Clustering

- Partition the data set in c-disjoint “clusters” of data points.



$$J_{km}(\mathbf{G},\mathbf{M}) = \sum_{c=1}^{k}\sum_{i=1}^{n} g_{ci}\|\mathbf{x}_i - \mathbf{m}_c\|_2^2 = \|\mathbf{X} - \mathbf{MG}\|_F^2$$

n=samples

d=features

$$\mathbf{X} = [x_1, \cdots, x_n] \in \mathbb{R}^{d\times n} \quad \mathbf{M} \in \mathbb{R}^{d\times k} \quad \mathbf{G}^T = \begin{array}{cc} \text{n} & \\ \begin{bmatrix} 1 & \dots & 0 \\ 0 & \dots & 1 \\ 0 & \dots & 0 \end{bmatrix} & \text{c} \end{array}$$

$$g_{ij} \in \{0,1\}$$

$$\mathbf{G1}_c = \mathbf{1}_n$$

k-means and kernel *k*-means

- K-means (Zha et al., 01;Ding & He,04; DelaTorre,06)

$$E_{km}(\mathbf{M},\mathbf{G}) = \sum_{c=1}^{k}\sum_{i=1}^{n} g_{ci}\left\|\mathbf{x}_i - \mathbf{m}_c\right\|_2^2 = \left\|\mathbf{X} - \mathbf{MG}\right\|_F^2$$

$$\mathbf{G} \in \{0,1\}^{k\times n} \quad \& \quad \mathbf{G}^T 1_k = 1_n \qquad \mathbf{G} = \begin{bmatrix} 1 & 1 \dots & 0 \\ 0 & 0 \dots & 1 \\ 0 & 0 \dots & 0 \end{bmatrix}$$

- Kernel K-means (Dhillon 04; DelaTorre,06)

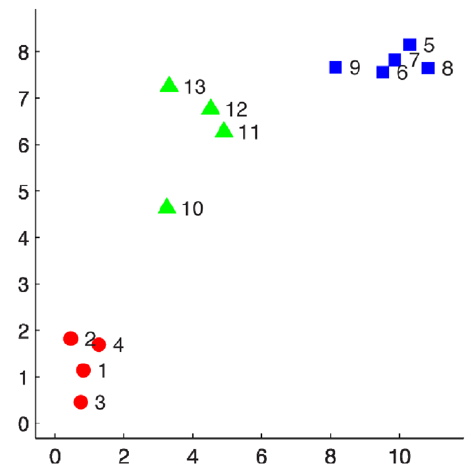
$$E_{kkm}(\mathbf{M},\mathbf{G}) = \sum_{c=1}^{k}\sum_{i=1}^{n} g_{ci}\left\|\varphi(\mathbf{x}_i) - \mathbf{m}_c\right\|_2^2 = \left\|\varphi(\mathbf{X}) - \mathbf{MG}\right\|_F^2$$

$$E_{kkm}(\mathbf{G}) = tr(\mathbf{K}(\mathbf{I}_n - \mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G}))$$

$\mathbf{K} = \varphi(\mathbf{X})^T \varphi(\mathbf{X}) \in \Re^{n\times n}$ is the kernel matrix

kernel *k*-means (Dhillon et al. 04, De la Torre 06)

$$J_{kkm}(\mathbf{G},\mathbf{M}) = \sum_{c=1}^{k}\sum_{i=1}^{n} g_{ci}\|\phi(\mathbf{x}_i) - \mathbf{m}_c\|_2^2 = \|\phi(\mathbf{X}) - \mathbf{MG}\|_F^2$$

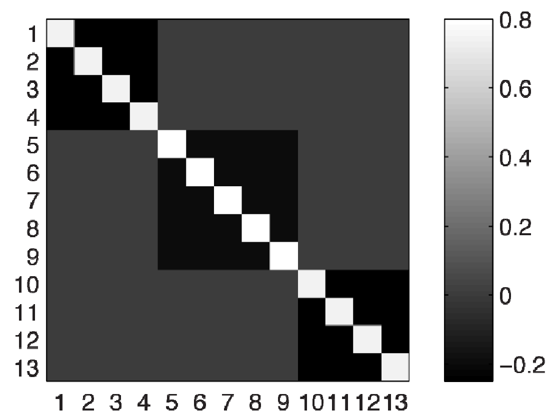


Matrix form

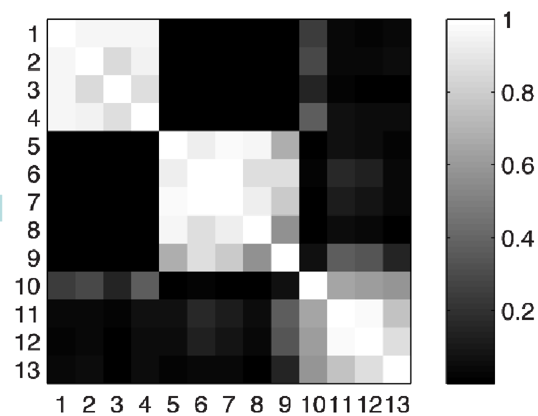
$$J_{kkm} = \mathrm{Tr}\left(\mathbf{LK}\right) \quad \text{with} \quad \mathbf{L} = \mathbf{I}_n - \mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G}$$

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	1	1	1

$\mathbf{G} \in \{0,1\}^{k\times n}$



$\mathbf{L} \in \mathbb{R}^{n\times n}$



$\mathbf{K} = \phi(\mathbf{X})^T\phi(\mathbf{X}) \in \mathbb{R}^{n\times n}$

Dynamic Time Alignment Kernel (DTAK)

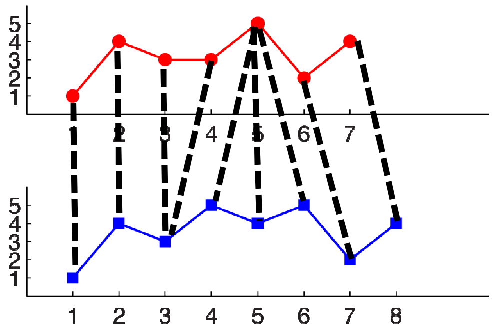
- How to measure a distance between two segment **X** and **Y**?
  - Use the Dynamic Time Alignment Kernel

$$\mathbf{X} \doteq [\mathbf{x}_1, \cdots, \mathbf{x}_{n_x}] \in \mathbb{R}^{d \times n_x} \qquad \mathbf{Y} \doteq [\mathbf{y}_1, \cdots, \mathbf{y}_{n_y}] \in \mathbb{R}^{d \times n_y}$$

$$\tau(\mathbf{X}, \mathbf{Y}) = \frac{p_{n_x, n_y}}{n_x + n_y} \quad \text{frame kernel: } \kappa_{ij} = \phi(\mathbf{x}_i)^T \phi(\mathbf{y}_j)$$

$$\text{with} \qquad p_{i,j} = \max \begin{cases} p_{i-1,j} + \kappa_{ij} \\ p_{i-1,j-1} + 2\kappa_{ij} \\ p_{i,j-1} + \kappa_{ij} \end{cases}$$

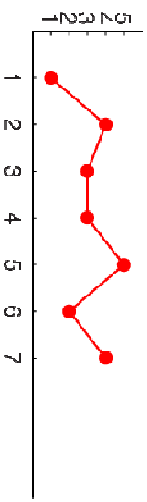
An example



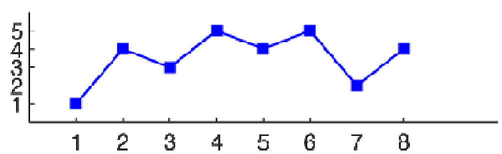
with

$$\tau(\mathbf{X},\mathbf{Y}) = \frac{p_{n_x,n_y}}{n_x+n_y}$$

$$p_{i,j} = \max \begin{cases} p_{i-1,j} + \kappa_{ij} \\ p_{i-1,j-1} + 2\kappa_{ij} \\ p_{i,j-1} + \kappa_{ij} \end{cases}$$

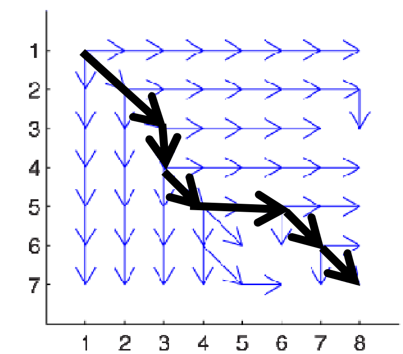


	1	2	3	4	5	6	7	8
1	1	0	0	0	0	0	0	0
2	0	1	0	0	1	0	0	1
3	0	0	1	0	0	0	0	0
4	0	0	1	0	0	0	0	0
5	0	0	0	1	0	1	0	0
6	0	0	0	0	0	0	1	0
7	0	1	0	0	1	0	0	1



Similarity matrix

		2		4		6		8
1	2	2	2	2	2	2	2	2
2	2	4	4	4	5	5	5	6
3	2	4	6	6	6	6	6	6
4	2	4	7	7	7	7	7	7
5	2	4	7	9	9	10	10	10
6	2	4	7	9	9	10	12	12
7	2	5	7	9	11	11	12	14



Error function for ACA

- Introduce a new variable **s** that defines the starting and ending of the segment $\mathbf{X}_{[s_i, s_{i+1})} = [\mathbf{x}_{s_i}, \mathbf{x}_{s_i+1}, \cdots, \mathbf{x}_{s_{i+1}-1}]$
- ACA minimizes:

$$J_{aca}(\mathbf{M}, \mathbf{G}, \mathbf{s}) = \sum_{c=1}^{k} \sum_{i=1}^{m} g_{ci} \|\psi(\mathbf{X}_{[s_i, s_{i+1})}) - \mathbf{m}_c\|^2 = \|\psi(\mathbf{X}) - \mathbf{MG}\|_F^2$$

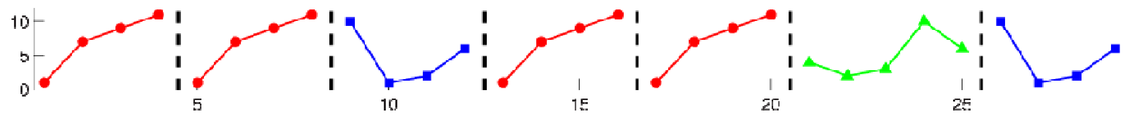
DTAK

	ACA	Kernel K-means
Features	segments with consecutive frames	single samples
Variables	s,G,M	G,M

Matrix form

$$J_{aca} = \mathrm{Tr}\Big((\mathbf{L} \circ \mathbf{W})\mathbf{K}\Big) \quad \text{with} \quad \mathbf{L} = \mathbf{I}_n - \mathbf{H}^T\mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G}\mathbf{H}$$

n = 28 (frames) m = 7 (segments) k = 3 (clusters)



clusters / segments

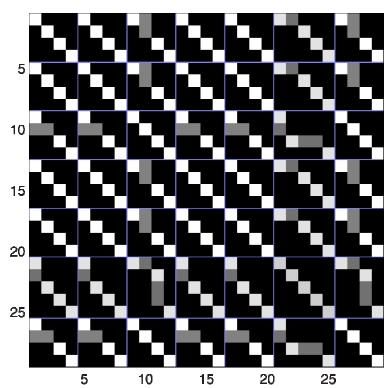
	1	2	3	4	5	6	7
1	1	1	0	1	1	0	0
2	0	0	1	0	0	0	1
3	0	0	0	0	0	1	0

$\mathbf{G} \in \{0,1\}^{k \times m}$

segments / samples

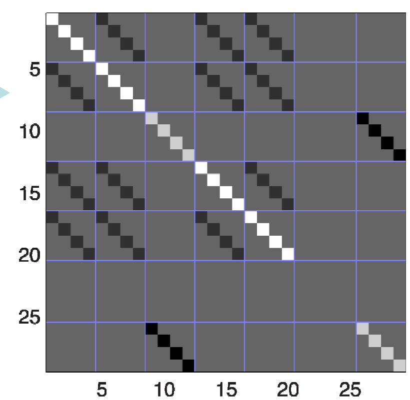
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

$\mathbf{H} \in \{0,1\}^{m \times n}$

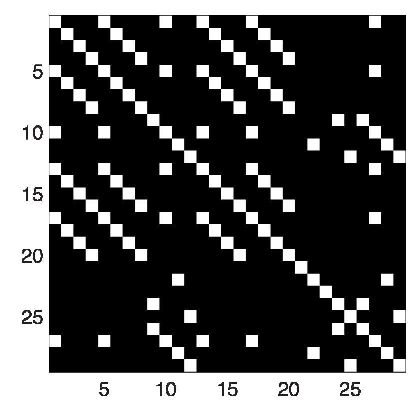


$\mathbf{W} \in \mathbb{R}^{n \times n}$

$\mathbf{L} \circ \mathbf{W} \in \mathbb{R}^{n \times n}$



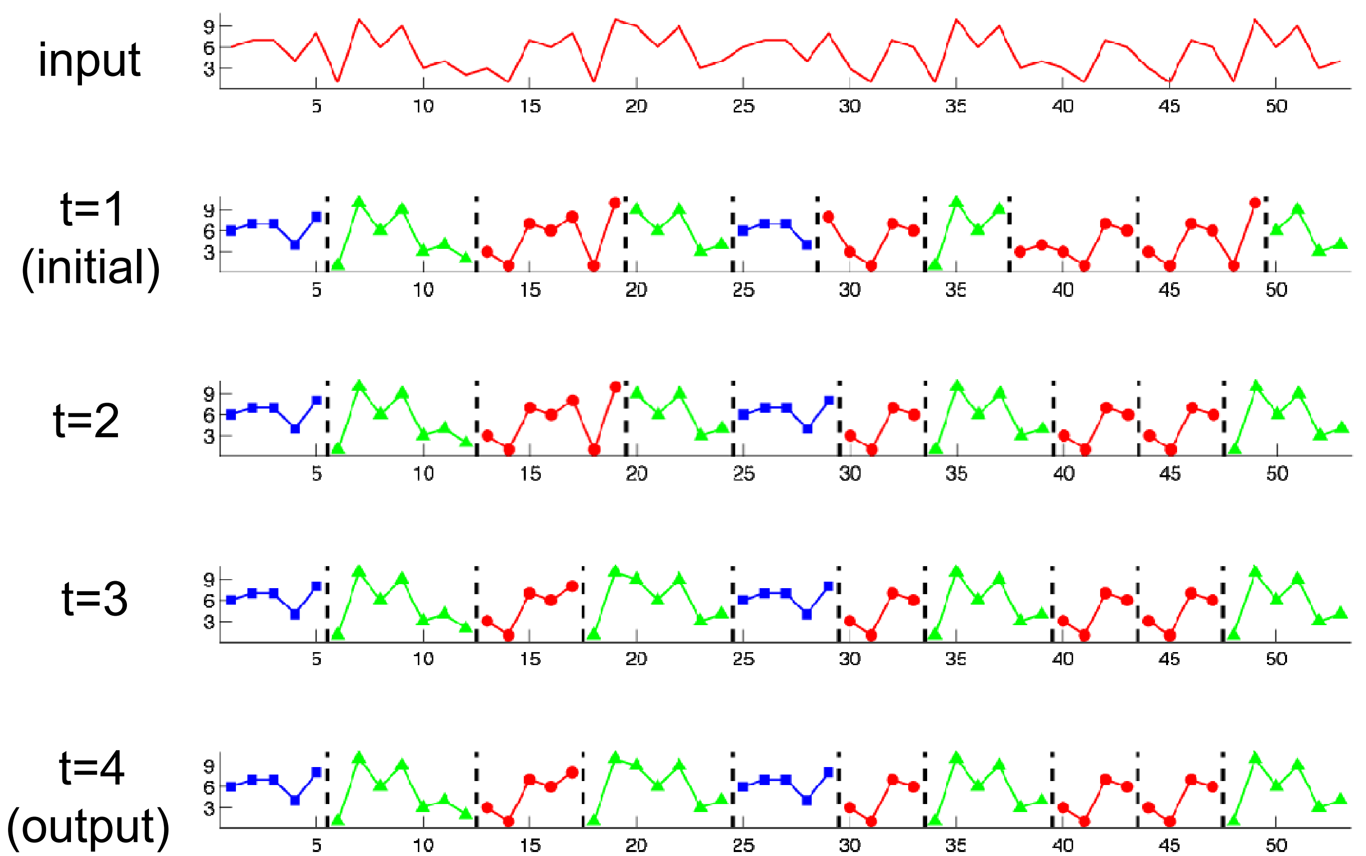
$\mathbf{K} \in \mathbb{R}^{n \times n}$

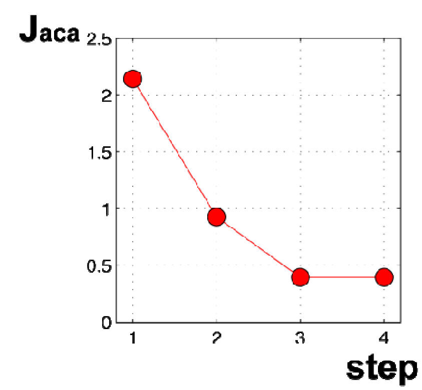


Optimizing ACA

- ACA is optimized by coordinate-descent

$$\mathbf{G}^t, \mathbf{s}^t = \arg\min_{\mathbf{G},\mathbf{s}} \sum_{c=1}^{k} \sum_{i=1}^{m} g_{ci} \|\psi(\mathbf{Y}_i) - \mathbf{m}_c^{t-1}\|^2$$

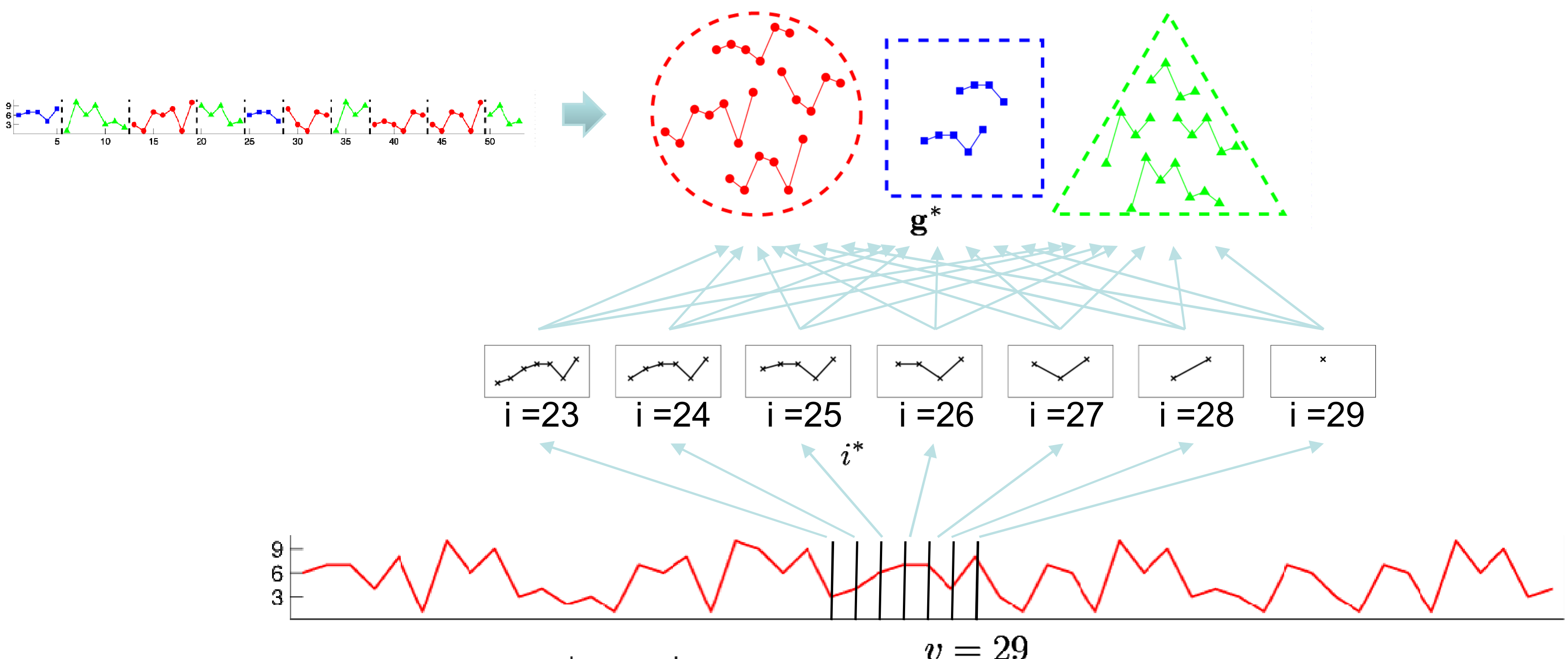




Optimizing ACA (Forward step)

- Each step is solved with Dynamic Programming.

$$\mathbf{L}(1,v) = \min_{v-w_{\max}<i\leq v}\left(\mathbf{L}(1,i-1) + \min_{\mathbf{g}}\sum_{c=1}^{k} g_c\|\psi(\mathbf{X}_{[i,v]}) - \mathbf{m}_c^0\|^2\right)$$

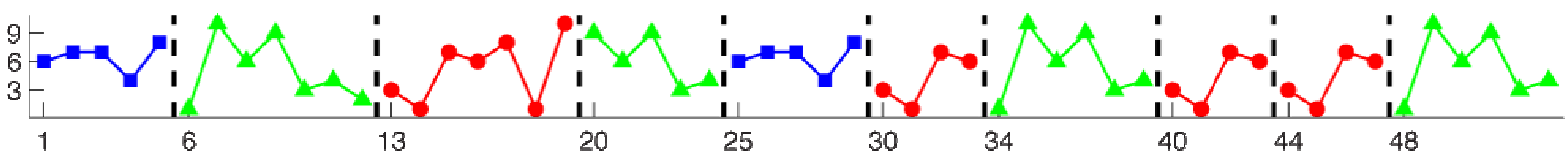


v	i_v^*	$\mathbf{g}_v^*$
$\cdots$	$\cdots$	$\cdots$
29	25	■

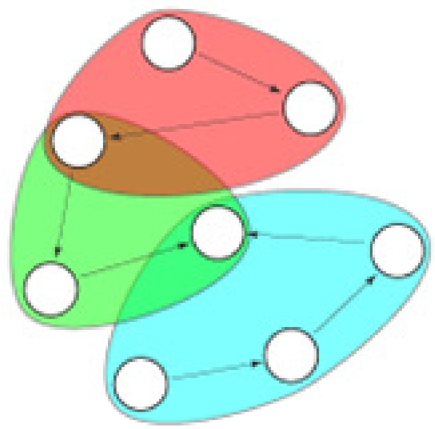
Move forward for v > 29

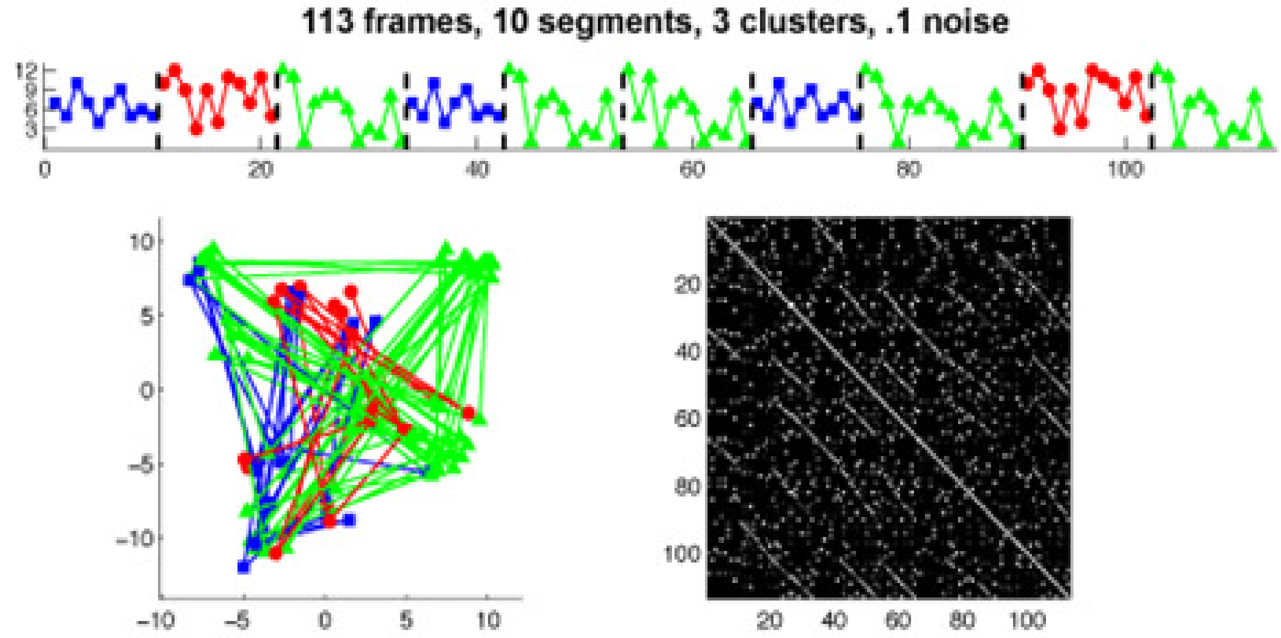
Optimizing ACA (Backward step)

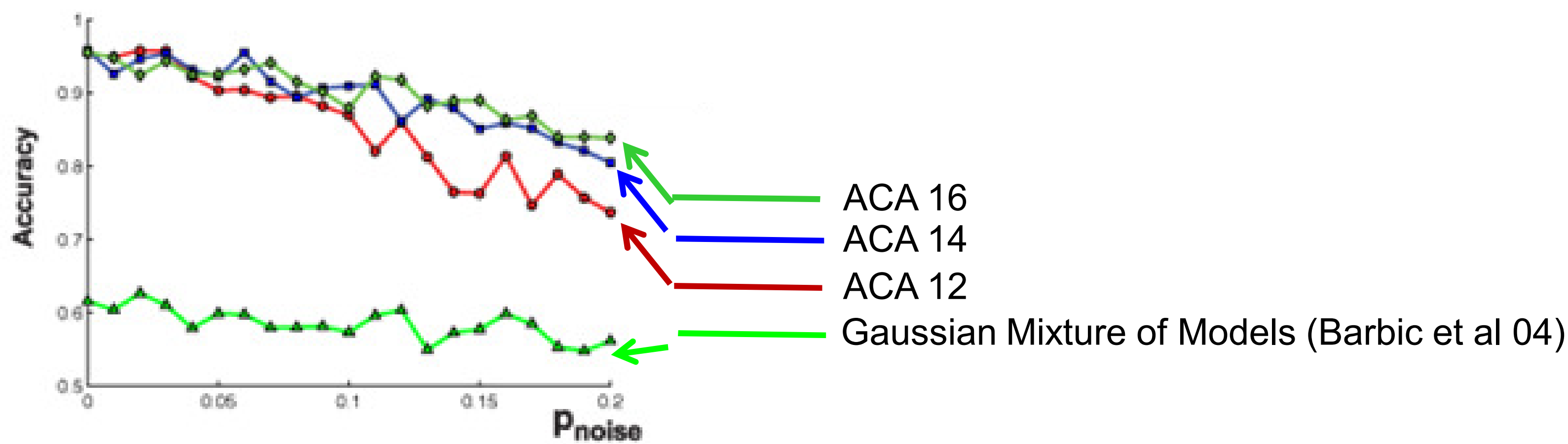
v	i_v^*	$\mathbf{g}_v^*$
...	...	...
29	25	■
...	...	...
33	30	●
...	...	...
39	34	▲
...	...	...
43	40	●
...	...	...
47	44	●
...	...	...
53	48	▲



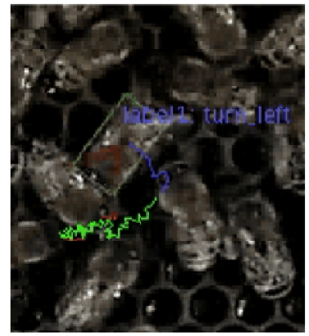
Synthetic data

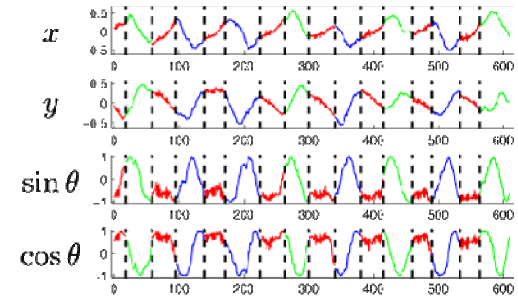




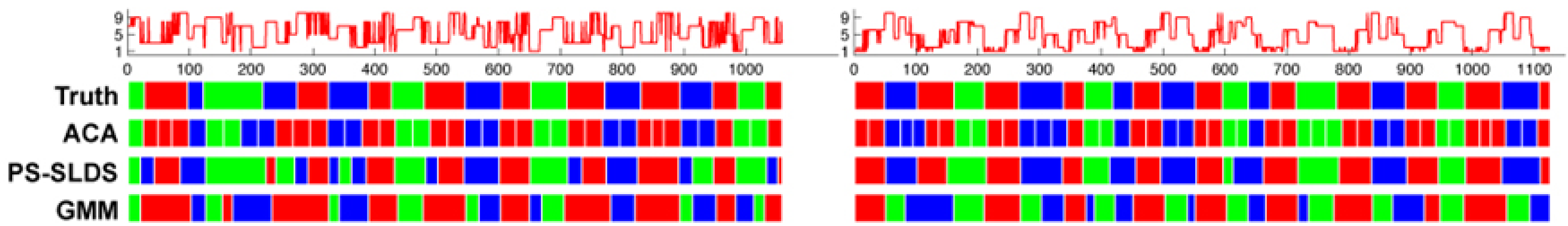


Honey bee dance data





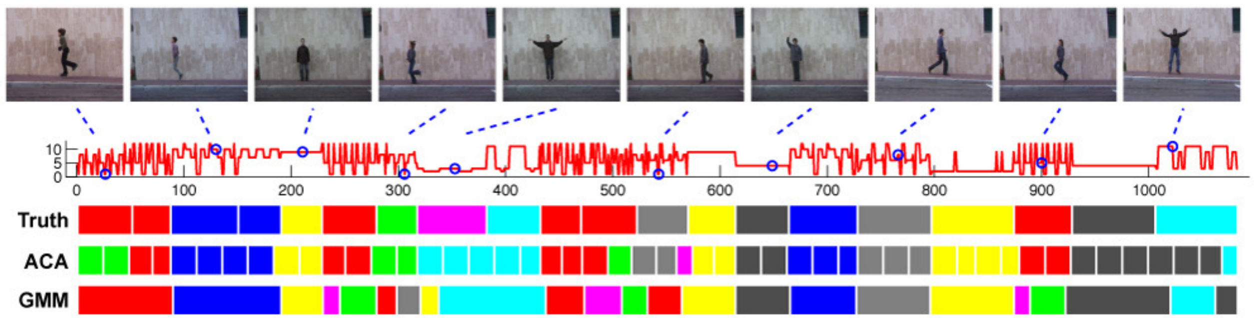
3 clusters:
waggle, left turn and right turn.



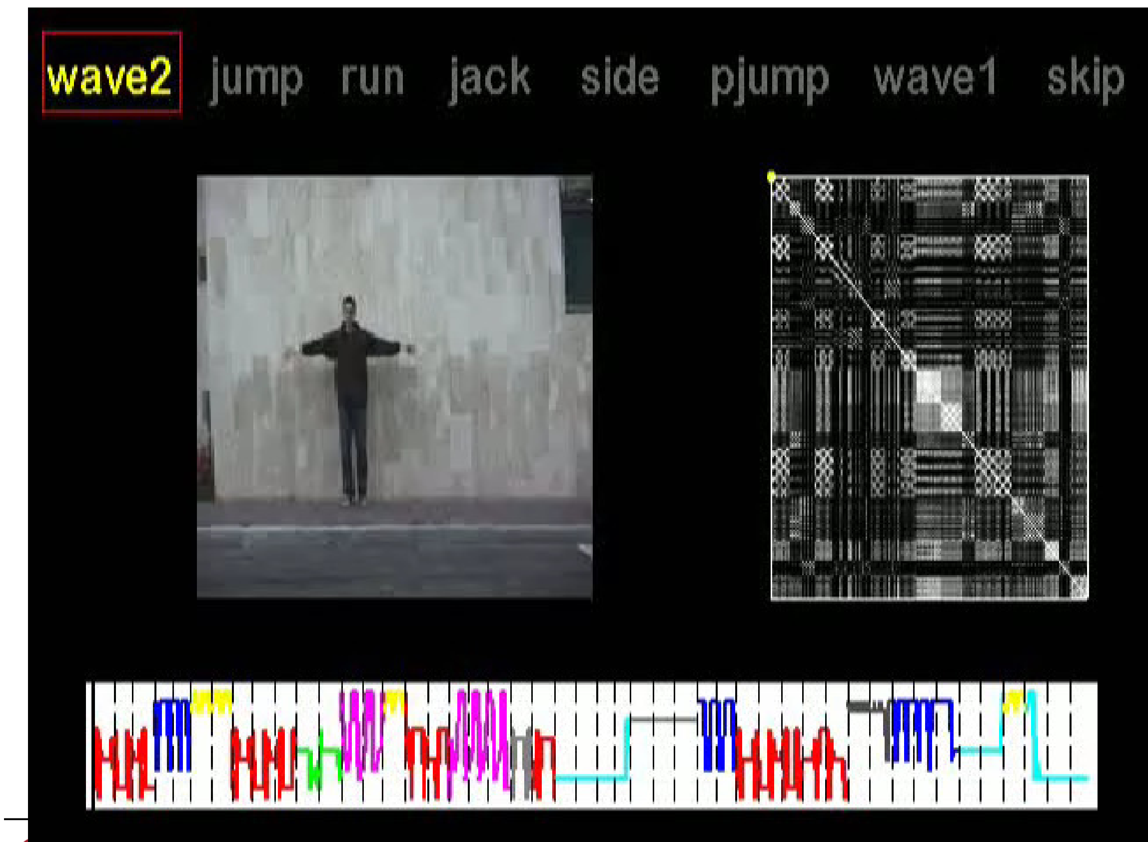
Parametric Segmental Switching Linear Dynamic System (PS-SLDS) (Oh et al. 08)

	Seq 1	Seq 2	Seq 3	Seq 4	Seq 5	Seq 6
ACA	**0.86**	**0.94**	0.67	0.92	**0.90**	**0.91**
PS-SLDS	0.76	0.92	**0.83**	**0.93**	**0.90**	0.90
GMM	0.74	0.67	0.44	0.71	0.58	0.71

Video data



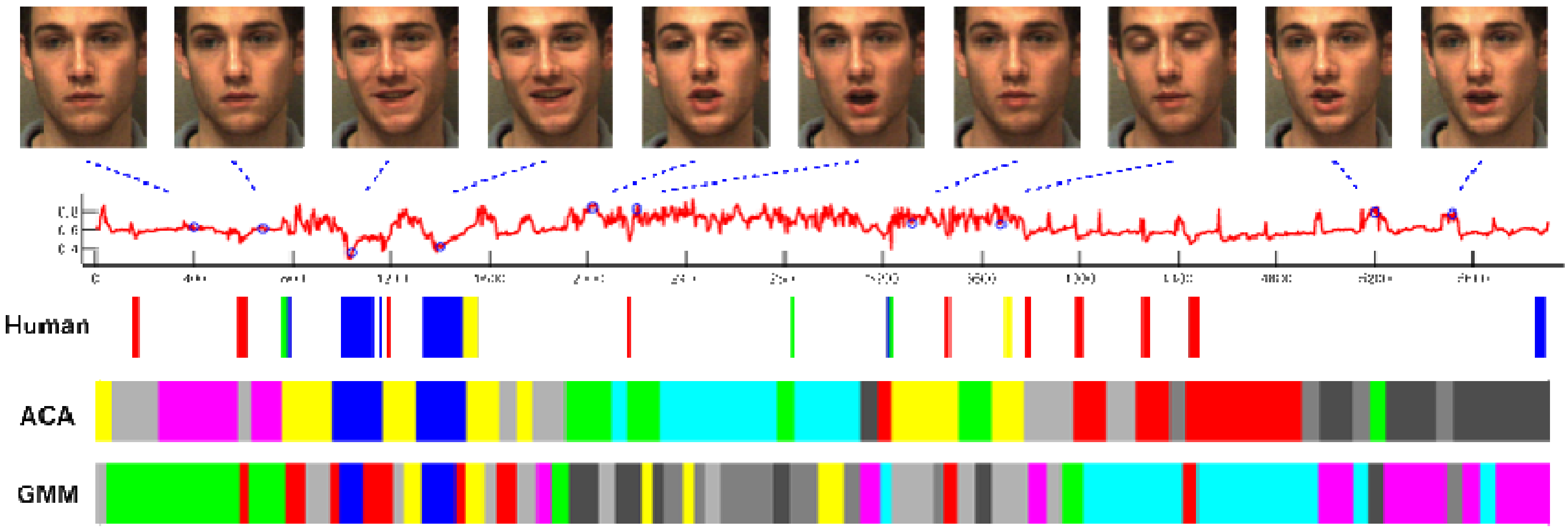
Weizmann dataset
(90 videos, 9 actions,
10 subjects):

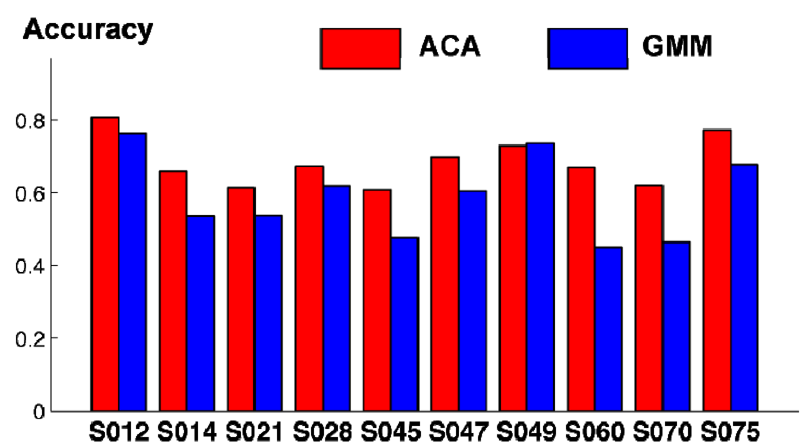


	Weizmann
ACA	0.76
GMM	0.63

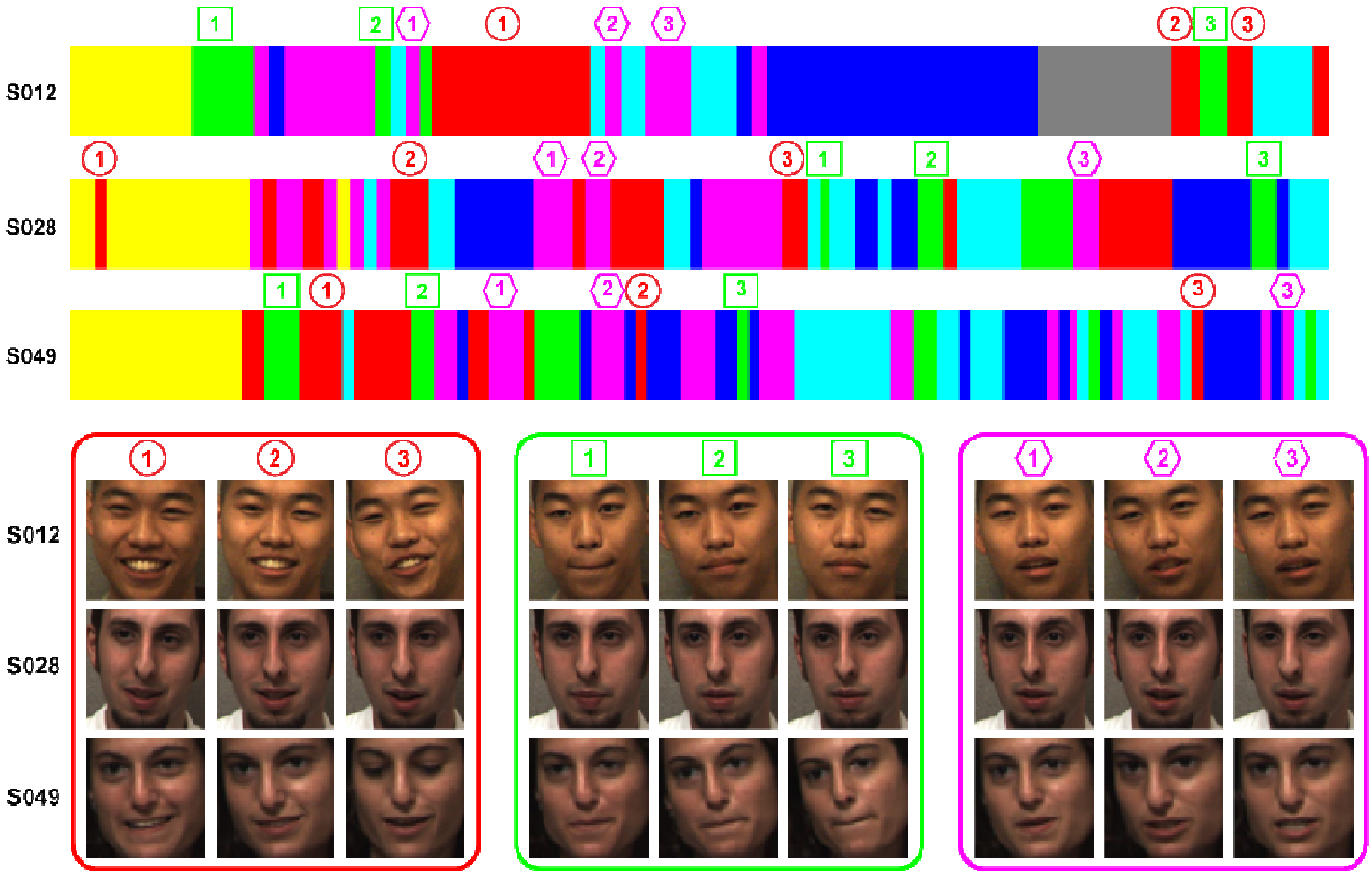
150 videos, 6
actions, 25
subjects

Unsupervised facial action discovery





Learning a vocabulary for facial expression

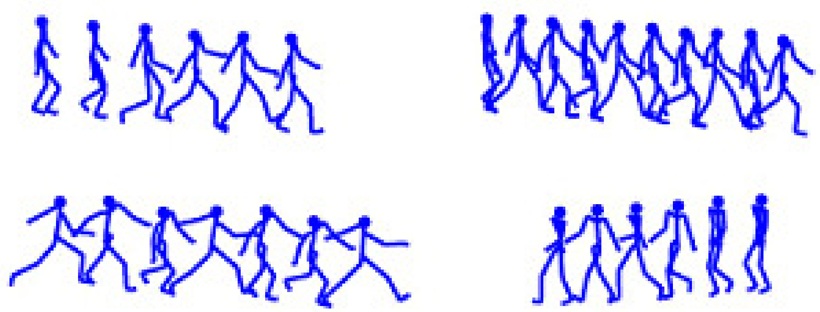


Outline

- Introduction
- Principal component analysis (PCA)
  – Robust Principal Component Analysis (RPCA)
  – PCA with uncertainty and missing data.
  – Parameterized Principal Component Analysis (PaPCA)
- Linear discriminant analysis (LDA)
  – Discriminative cluster analysis (DCA).
- K-means
  – Hierarchical aligned cluster analysis (HACA).
- Canonical correlation analysis (CCA)
  – Canonical time warping (CTW)
  – Dynamic coupled component analysis (DCCA)

Alignment of human behavior

- Temporal alignment of human behavior is useful in many human sensing tasks:

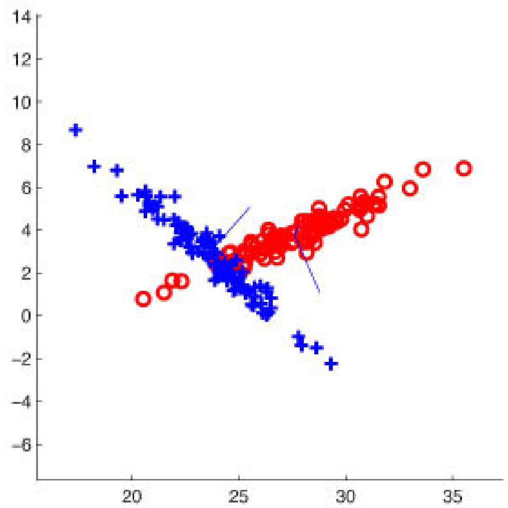


Previous work

- Data mining: *Derivative Dynamic Time Warping* *(Keogh et al. 01)*
  - DTW + Derivatives of features
- Computer Graphics: *Style Translation* *(Hsu et al. 05)*
  - DTW + Least-square
- Computer Vision: *View-invariant Action Recognition* *(Rao et al. 03 09)*
  - DTW + Homography
- Other Area: *Multiple Alignment of Continuous Time Series* *(Listgarten et al. 04)*
  - HMM + Account for changes in the amplitude of the signals
- Our work: *Canonical Time Warping* *(appear in NIPS 2009)*
  - DTW + Canonical Correlation Analysis (CCA)

Canonical Correlation Analysis (CCA)

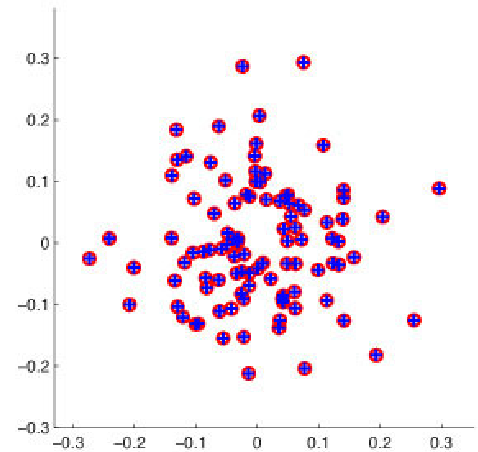
- Canonical Correlation Analysis *(Hotelling 1936)*



$$\mathbf{X} \in \mathbb{R}^{d_x \times n}, \mathbf{Y} \in \mathbb{R}^{d_y \times n}$$

Same #columns (samples)
Different #rows (dimensionality)

CCA projection



$$J_{cca}(\mathbf{V}_x, \mathbf{V}_y) = \|\mathbf{V}_x^T \mathbf{X} - \mathbf{V}_y^T \mathbf{Y}\|_F^2$$

$$\mathbf{V}_x \in \mathbb{R}^{d_x \times b}, \mathbf{V}_y \in \mathbb{R}^{d_y \times b}$$

$$\text{s.t. } \mathbf{V}_x^T \mathbf{X}\mathbf{X}^T \mathbf{V}_x = \mathbf{V}_y^T \mathbf{Y}\mathbf{Y}^T \mathbf{V}_y = \mathbf{I}_b$$

A least-square formulation for DTW

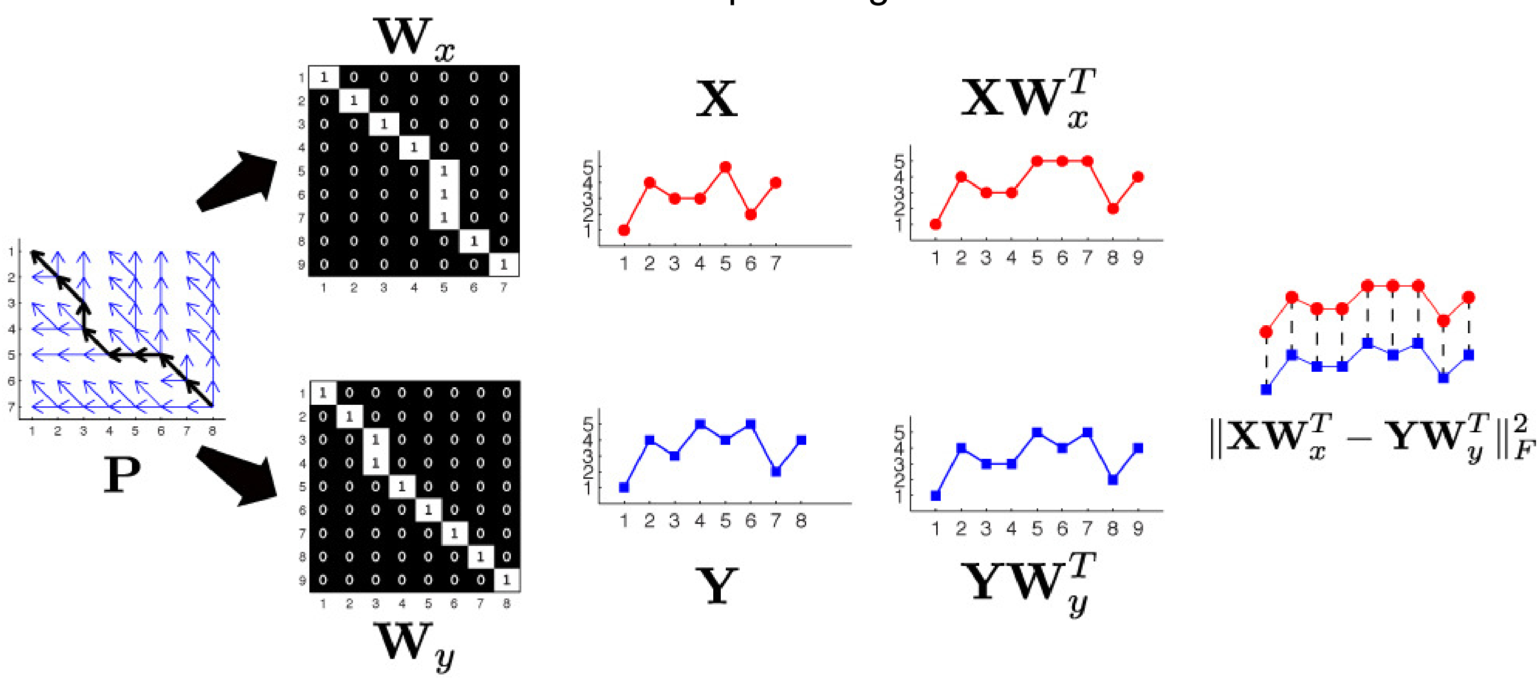
same #rows, different #columns

$$\mathbf{X} \in \mathbb{R}^{d \times n_x}, \mathbf{Y} \in \mathbb{R}^{d \times n_y}$$

$$J_{dtw}(\mathbf{W}_x, \mathbf{W}_y) = \|\mathbf{X}\mathbf{W}_x^T - \mathbf{Y}\mathbf{W}_y^T\|_F^2$$

$$\mathbf{W}_x \in \{0,1\}^{m \times n_x}, \mathbf{W}_y \in \{0,1\}^{m \times n_y}$$

temporal alignment



Canonical Time Warping (CTW)

Reminder

$$J_{dtw}(\mathbf{W}_x, \mathbf{W}_y) = \|\mathbf{X}\mathbf{W}_x^T - \mathbf{Y}\mathbf{W}_y^T\|_F^2$$

$$J_{cca}(\mathbf{V}_x, \mathbf{V}_y) = \|\mathbf{V}_x^T\mathbf{X} - \mathbf{V}_y^T\mathbf{Y}\|_F^2$$

different #rows, different #columns

$$\mathbf{X} \in \mathbb{R}^{d_x \times n_x}, \mathbf{Y} \in \mathbb{R}^{d_y \times n_y}$$

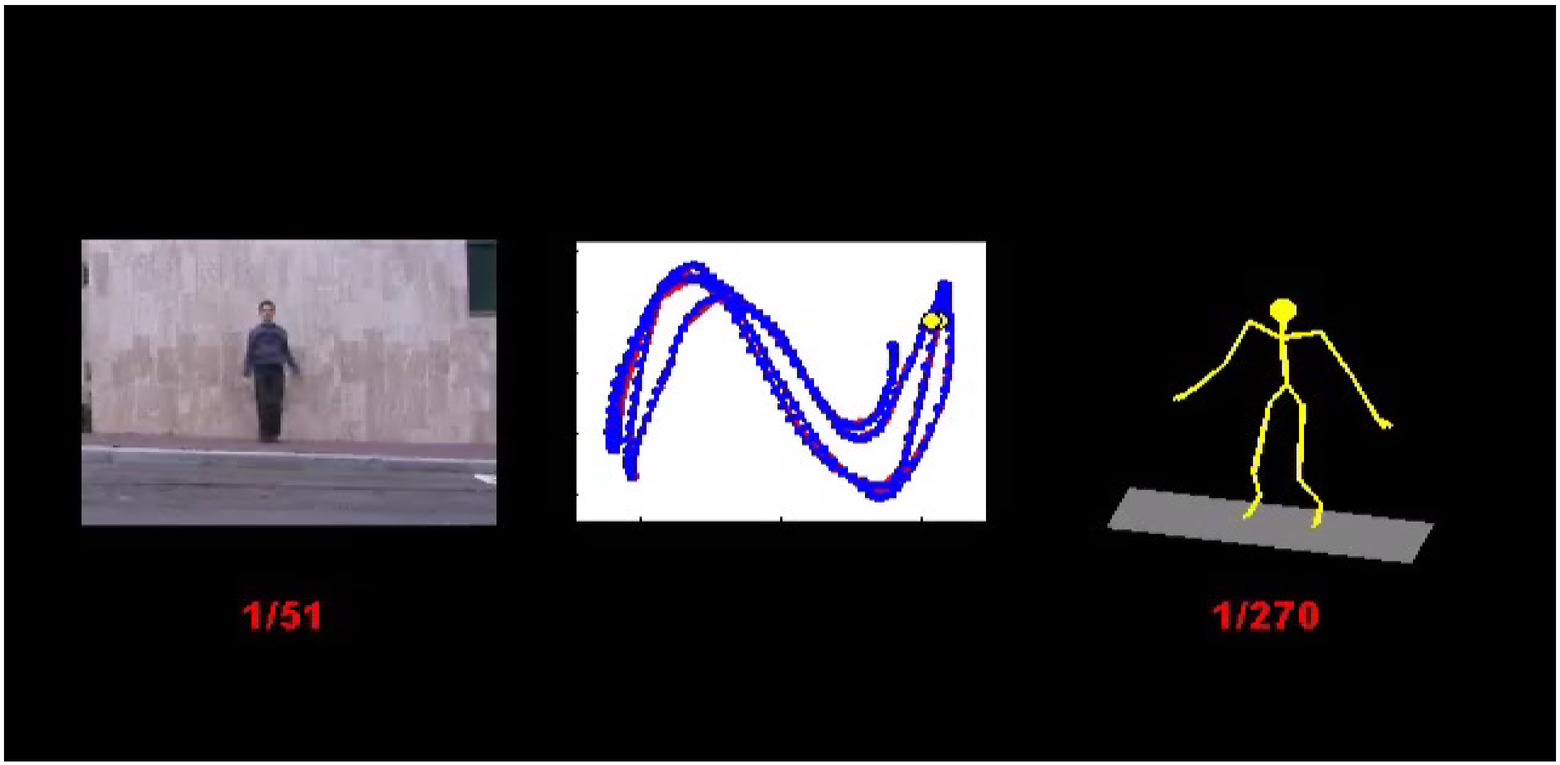
spatial transformation

$$J_{ctw}(\mathbf{W}_x, \mathbf{W}_y, \mathbf{V}_x, \mathbf{V}_y) = \|\mathbf{V}_x^T\mathbf{X}\mathbf{W}_x^T - \mathbf{V}_y^T\mathbf{Y}\mathbf{W}_y^T\|_F^2$$

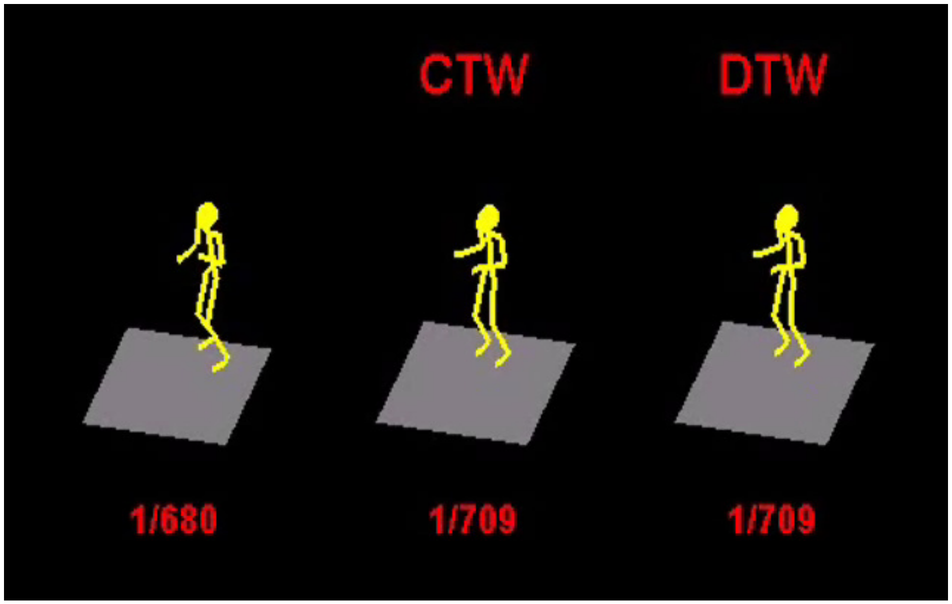
temporal alignment

$$\mathbf{V}_x^T\mathbf{X}\underbrace{\mathbf{W}_x^T\mathbf{W}_x}_{\mathbf{D}_x}\mathbf{X}^T\mathbf{V}_x - \mathbf{V}_y^T\mathbf{Y}\underbrace{\mathbf{W}_y^T\mathbf{W}_y}_{\mathbf{D}_y}\mathbf{Y}^T\mathbf{V}_y - \mathbf{I}_b$$

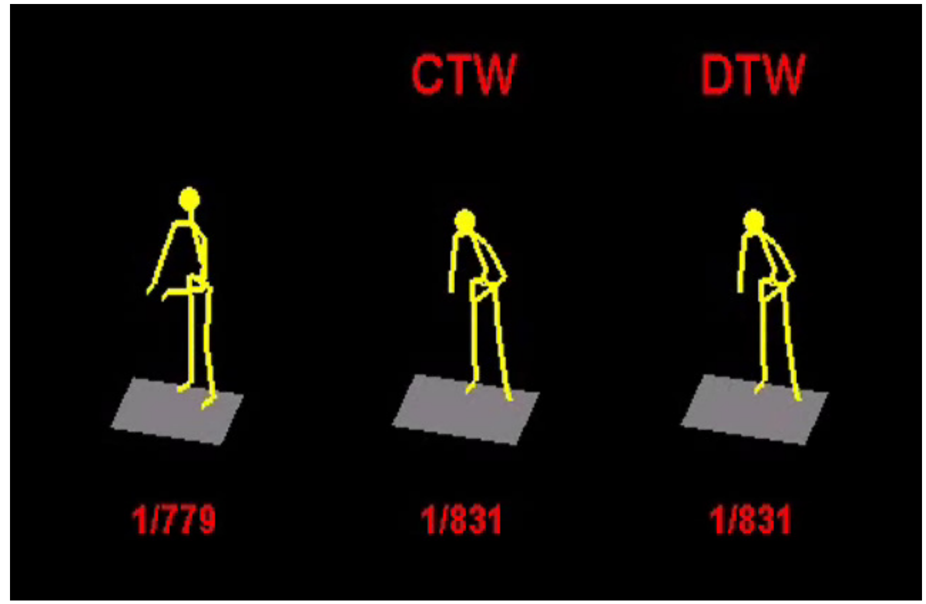
Aligning motion capture and video



Aligning motion capture data of same action (different people)

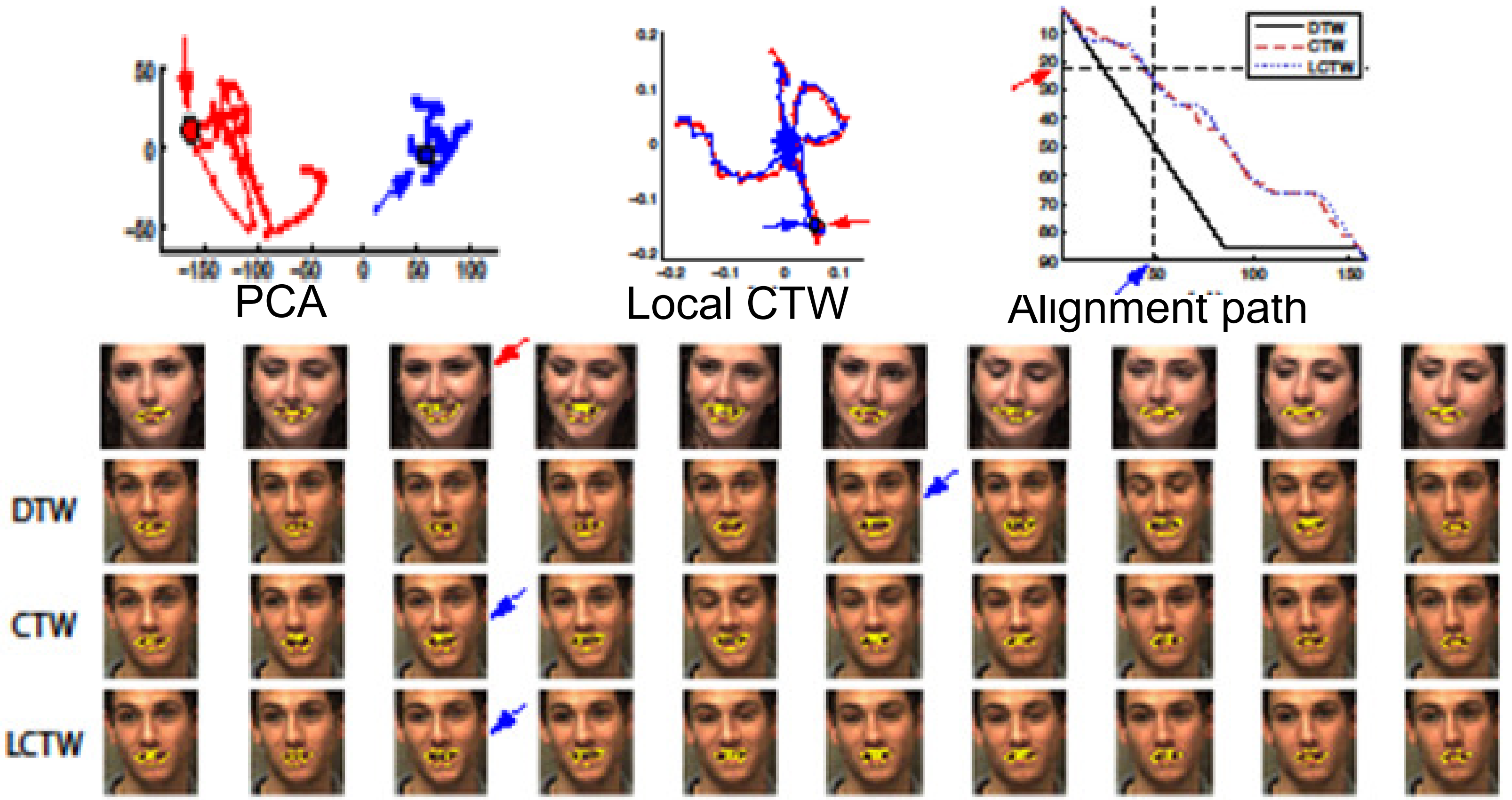


Boxing



Opening a cabinet

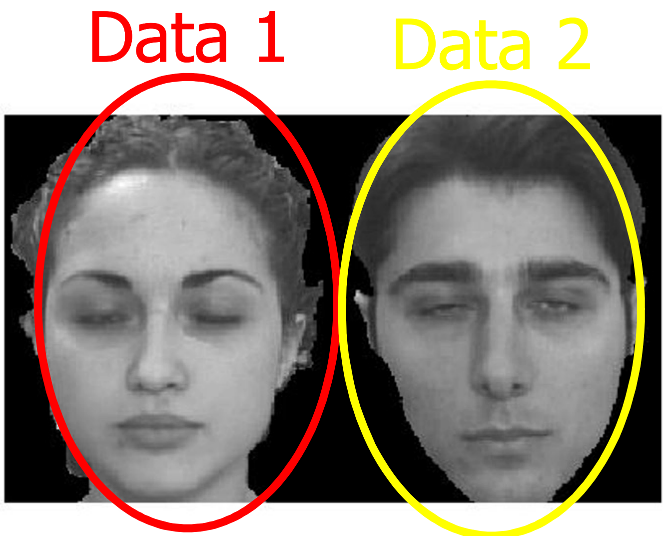
Facial expression alignment



Outline

- Introduction
- Principal component analysis (PCA)
  - Robust Principal Component Analysis (RPCA)
  - PCA with uncertainty and missing data.
  - Parameterized Principal Component Analysis (PaPCA)
- Linear discriminant analysis (LDA)
  - Discriminative cluster analysis (DCA).
- K-means
  - Hierarchical aligned cluster analysis (HACA).
- Canonical correlation analysis (CCA)
  - Canonical time warping (CTW)
  - Dynamic coupled component analysis (DCCA)

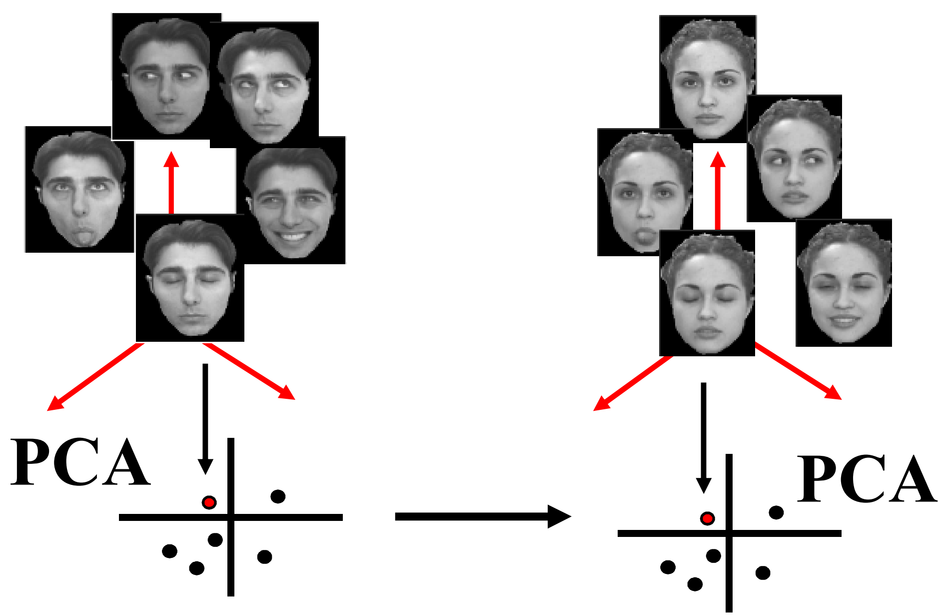
Dynamic Coupled Component Analysis (DCCA) (de la Torre & Black, 2001a)



- Learning the coupling.
- High dimensional data.
- Limited training data.

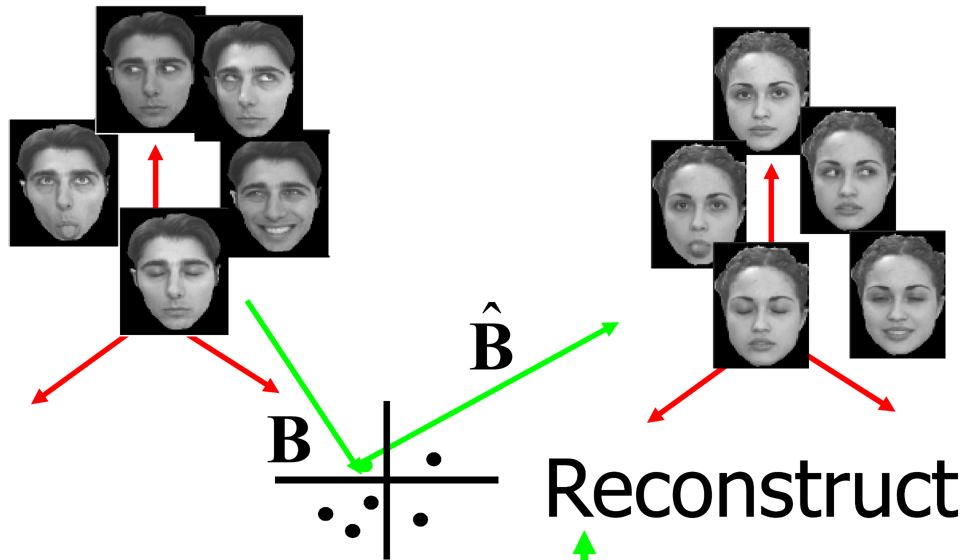
Solutions?

- PCA independently and general mapping



- Signals dependent signals with small energy can be lost.

DCCA



Reconstruction

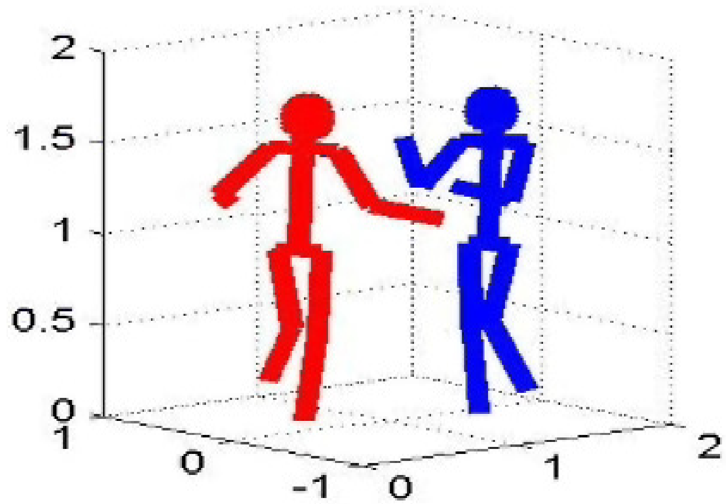
$$E_{cca}(\mathbf{B},\hat{\mathbf{B}},\mathbf{A},\mathbf{C},\boldsymbol{\mu},\hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} \left\| \hat{\mathbf{d}}_i - \hat{\boldsymbol{\mu}} - \hat{\mathbf{B}}\mathbf{c}_i \right\|_{\mathbf{W}1_i}^2$$

Projection

$$+ \lambda \sum_{i=1}^{n} \left\| \mathbf{c}_i - \mathbf{B}^T(\mathbf{d}_i - \boldsymbol{\mu}) \right\|_{\mathbf{W}2_i}^2 + \lambda_2 \sum_{i=1}^{n} \left\| \mathbf{c}_i - \mathbf{A}\mathbf{c}_{i-1} \right\|_{\mathbf{W}3_i}^2$$

Dynamics

Dynamic Coupled Component Analysis




CA Can Do It!
POST FEB. 15 TO FEB. 28
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.*