Canonical Correlation Analysis (CCA), its variants with applications

Chen Songcan (陈松灿)
s.chen@nuaa.edu.cn
http://parnec.nuaa.edu.cn

Dept. of Computer Science and Engineering
Nanjing University of Aeronautics and Astronautics (NUAA)

Contents

- Part 1: CCA, its variants, applications
- Part 2: Limitations & approaches
- Part 3: Our works

Part 1: CCA, its variants & applications

1. **An introduction of CCA**
2. **Variants of CCA**
3. **Applications**

1.1 Introduction of CCA

Original paper **[Hot1936]**

Relations Between Two Sets of Variates

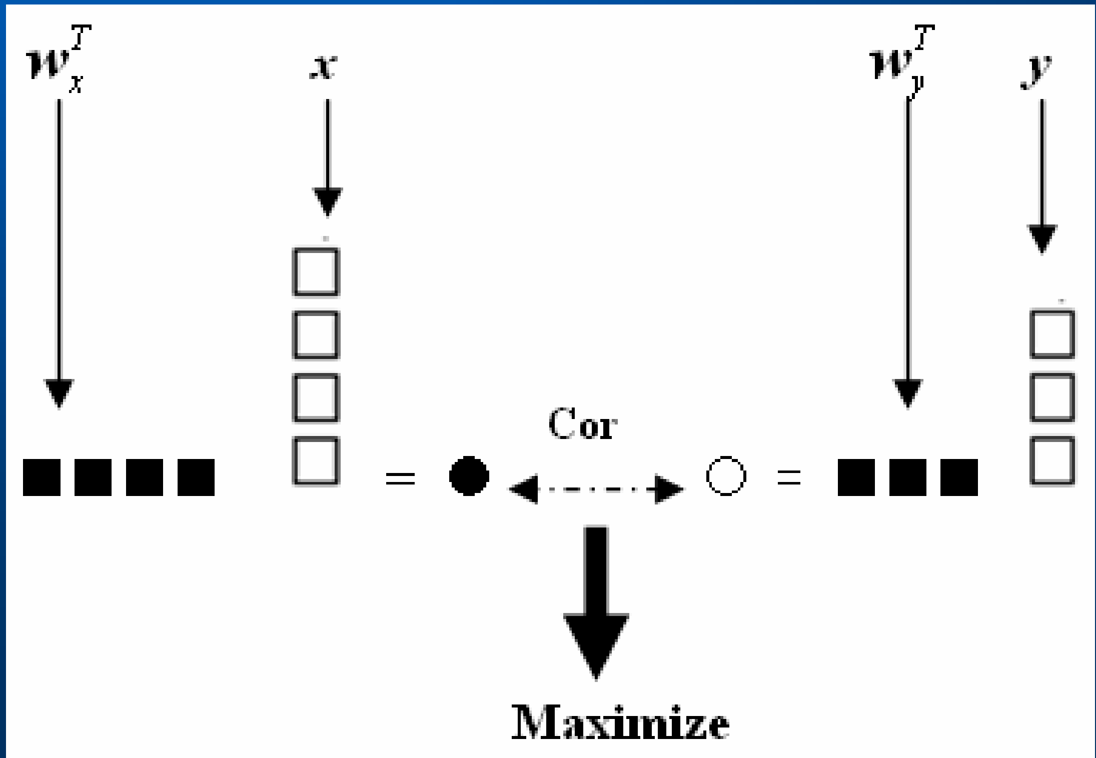
Harold Hotelling

Biometrika, Vol. 28, No. 3/4 (Dec., 1936), 321-377.

As early as Fisher LDA **[Fis1936].**

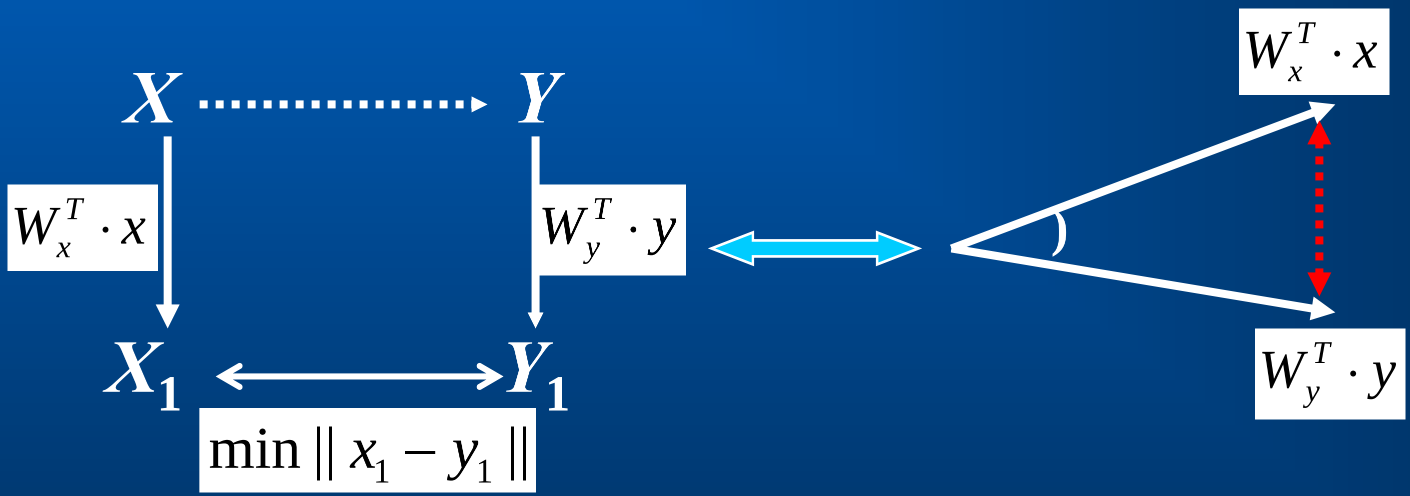
Description of CCA

Given X set $\{x_1,\ldots,x_n\} \in R^p$, and Y set $\{y_1,\ldots,y_n\} \in R^q$, CCA aims to simultaneously seek $w_x \in R^p$, $w_y \in R^q$, to ensure [Bor1999]



An equivalent description

We want to find a function relationship between X and Y, but we don't do so in the original space, instead we seek W_x and W_y such that



Minimize the distance between X_1 and Y_1 is equivalent to minimize the angle between X_1 and Y_1, i.e. maximize the correlation between them.

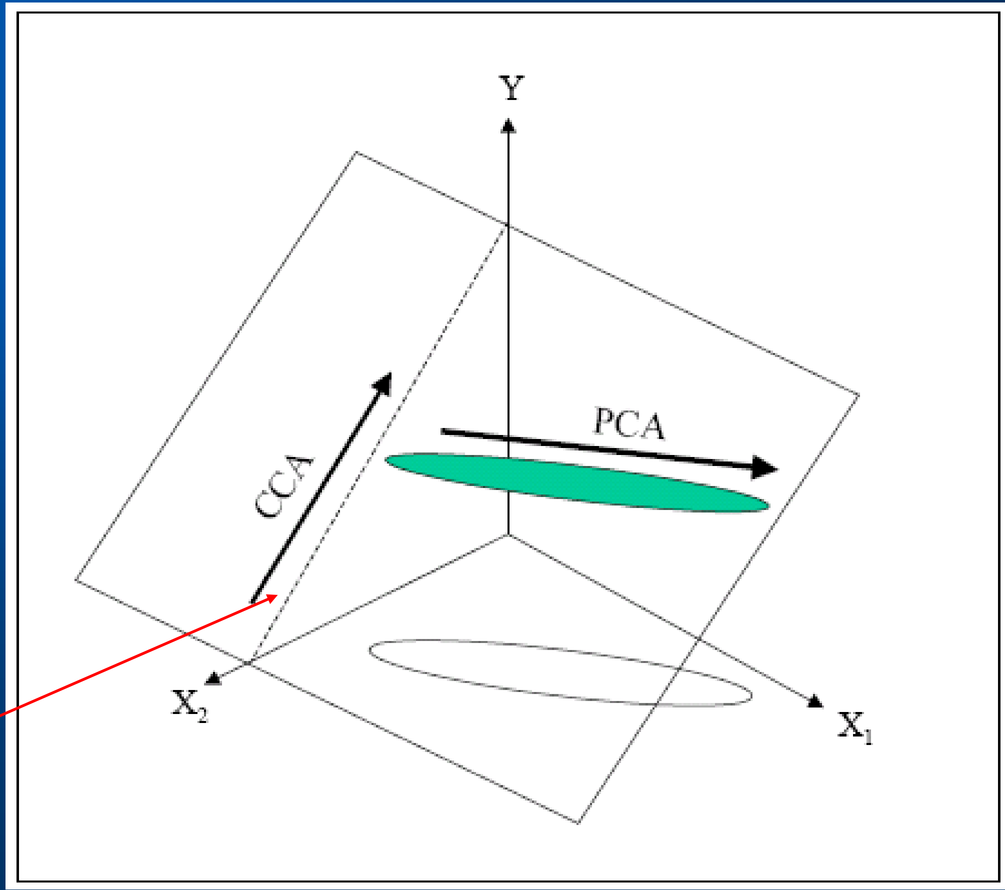
PCA vs. CCA

	PCA	CCA
Input	$X=(x^1,x^2...x^N)$	$X=(x^1,x^2...x^N)$ $Y=(y^1,y^2...y^N)$
Output	one projection matrix *W* $x_1=W^T\cdot x$	two projection matrixes Wx and Wy $\begin{cases} x_1=W_x^T\cdot x \\ y_1=W_y^T\cdot y \end{cases}$
Purpose	$\min\|\|x-W\cdot x_1\|\|$	$\min\|\|x_1-y_1\|\|$

PCA vs. CCA

Comparison between PCA and CCA.
It can be seen that the CCA provides useful information about the linear correlation between **x** and **y**, while the PCA fails to expose this information.

$y=1-x_2+noise$



$y=1-x_2$

[Yac2004]

A Unified framework

PCA, MLR & CCA can be rewritten as a unified generalized eigenvalue equation $Aw = \lambda Bw$ [BLK1992], and in some special case, CCA is equivalent to LDA [BR2003]

	A	B
PCA	C_{xx}	I
MLR	$\begin{pmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{pmatrix}$	$\begin{pmatrix} C_{xx} & 0 \\ 0 & I \end{pmatrix}$
CCA	$\begin{pmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{pmatrix}$	$\begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix}$

1.2 Variants of CCA

Up to now, CCA has been extended to

- **Kernel CCA (KCCA) [MRB2003, TC2004]**
- **Multiple sets CCA (mCCA) [Ket1971, Nie2002]**
- **Nonlinear CCA [GF2004, Lai2000, BL1983, BF1985, Buj1990, ST1992, Win1992]**
- **etc…**

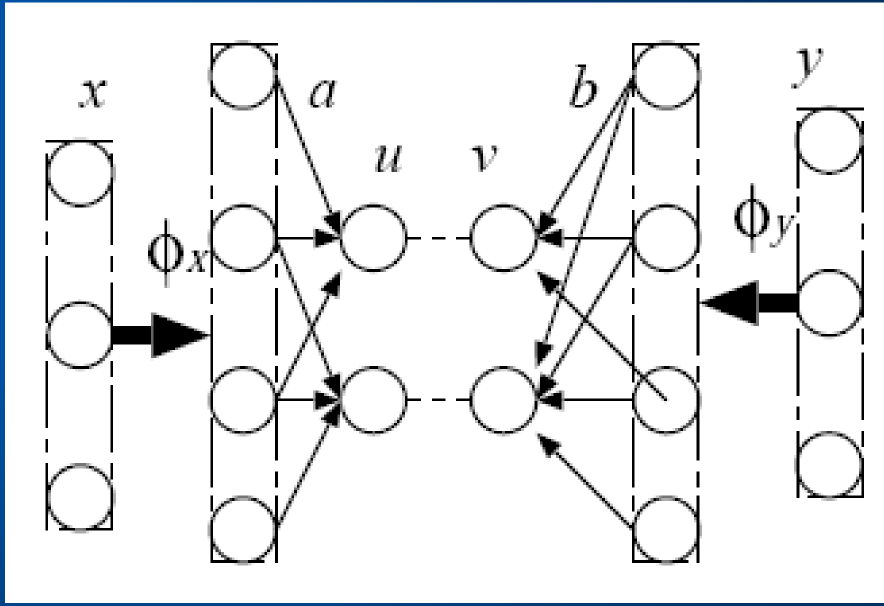
We focus on KCCA in this section.

KCCA: illustration

Input data: *X* and *Y*

Nonlinear mappings:

$$\phi_x : x \mapsto \phi_x(x)$$
$$\phi_y : y \mapsto \phi_y(y)$$



Performing CCA on $\phi_x(x), \phi_y(y)$ in *feature spaces* $\boldsymbol{F_x}$ and $\boldsymbol{F_y}$ using so-called "*kernel trick*", the linear correlation in $\boldsymbol{F_x}$ and $\boldsymbol{F_y}$ indicates the nonlinear correlation in original spaces. **[TC2004]**

1.3 Applications

CCA and its variants have been applied to

image processing and analysis [LGD2005], [Yac2004] , [Bor1998, FBL2001, Hil2001]

image retrieval [HSS2004].

blind source separation [BK2001].

remote sensing data processing [Nie2002, Nie1994,NCK1998, NCK1999].

pattern recognition [BR2003, Yo2004 , GSB2001].

text translation and analysis (Cross–language text mining, Text document retrieval, Text categorization, Image-Text Retrieval , Machine Translation) [For2004],

functional analysis using fMRI [IHA2004][Fri2001]

computer vision [ESE2005],

pose estimation [MRB2003] ,

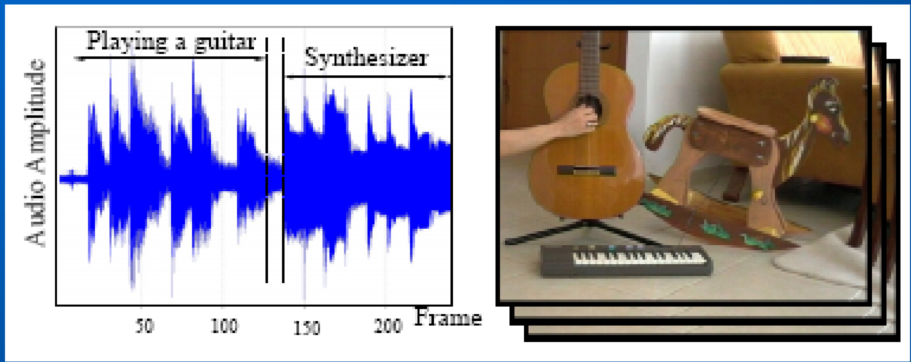
Biometrics [YVN2003][VK2003][NGK2001] and other fields.

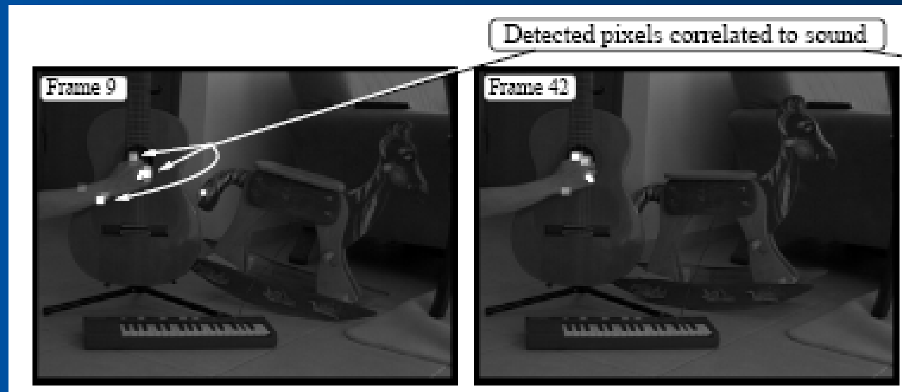
Applications: image retrieval

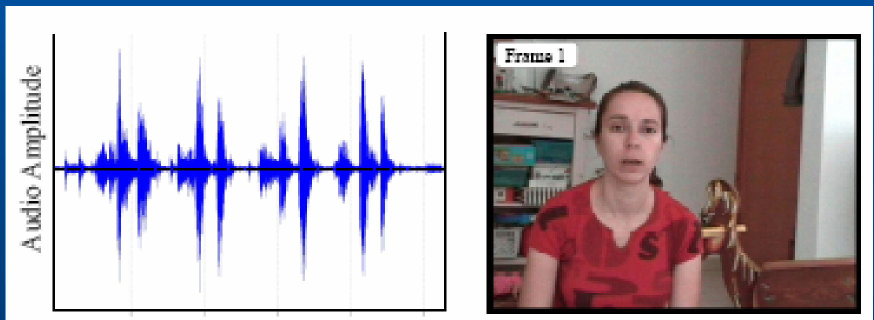
[HSS2004]
(NC)

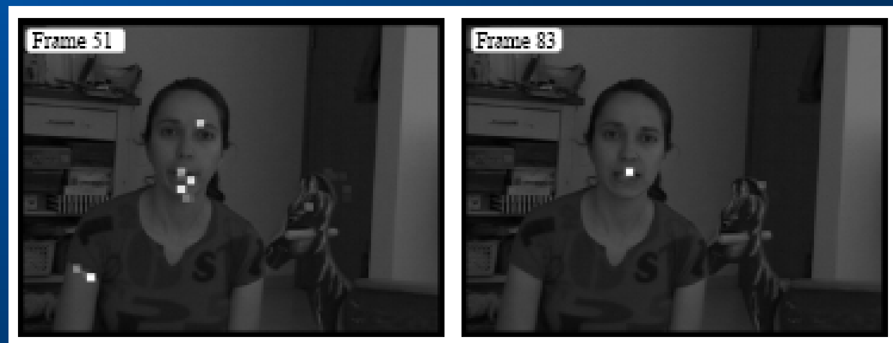
Images retrieved for the text query: “at phoenix sky harbor on July 6, 1997. 757-2s7, n907wa phoenix suns taxis past n902aw teamwork America west America west 757-2s7, n907wa phoenix suns taxis past n901aw Arizona at phoenix sky harbor on July 6, 1997.” The actual match is the middle picture in the first row.

Applications: computer vision



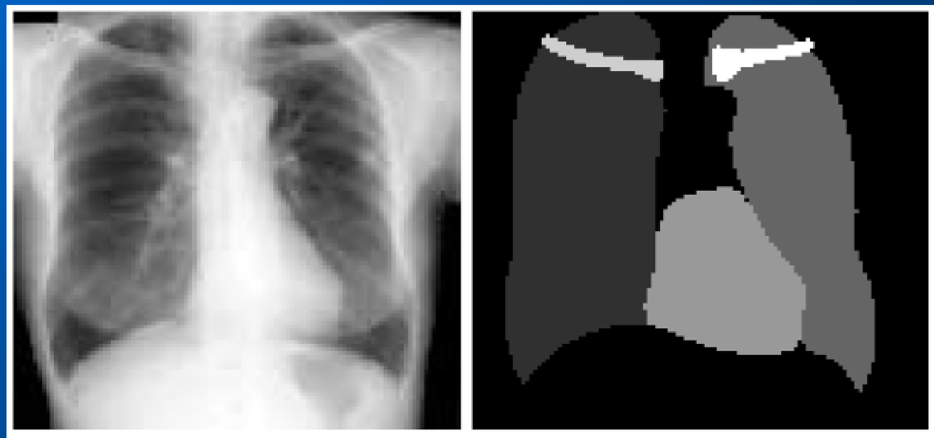






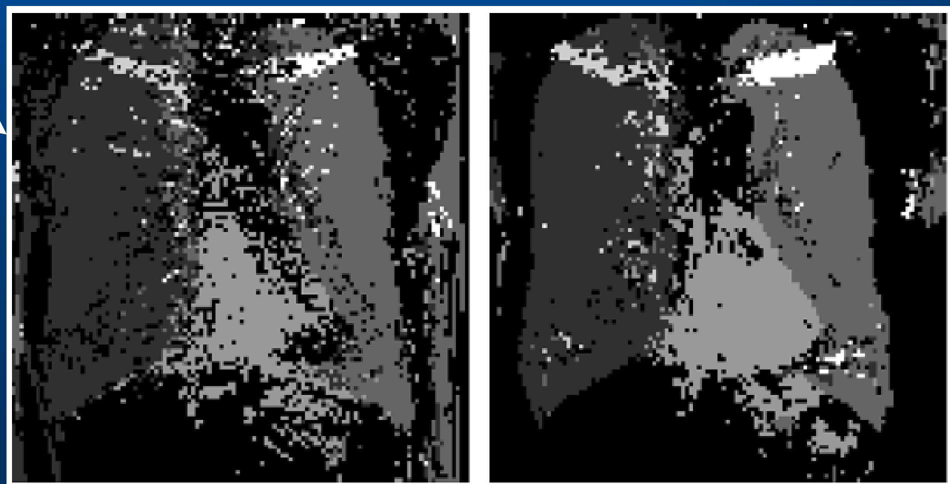
detecting pixels correlated to sound by CCA on audio & video [ESE2005] (CVPR05)

Applications: image segmentation



by expert

by LDA



by CCA

[LGD2005](PR)

Applications: image processing

Demosaicing**[Yac2004]** (HP report)

Object: color bands R, G, B

Processed result

Original

Mosaiced

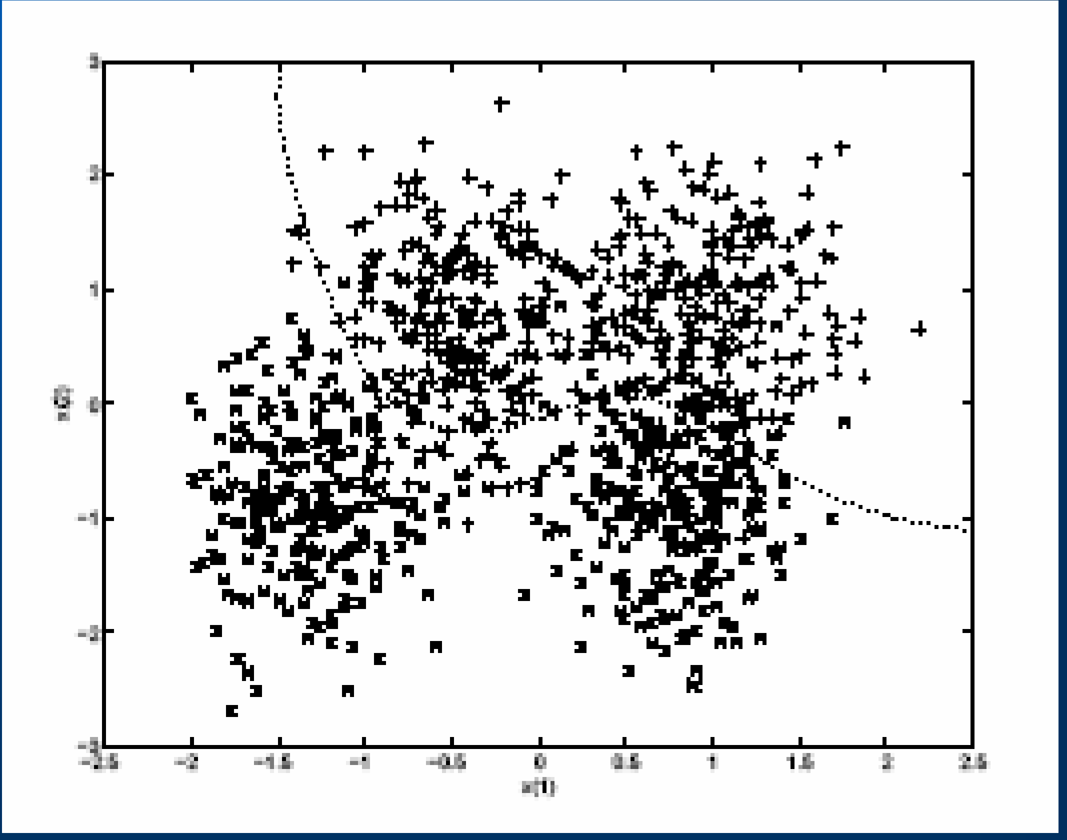
The demosaicing problem deals with the reconstruction of a full color image ***f*** from its partial sampling ***m***.

Applications: classification (1)

Binary classification [GSB2001] (ICANN)

$$\phi_x : x \mapsto \phi_x(x)$$

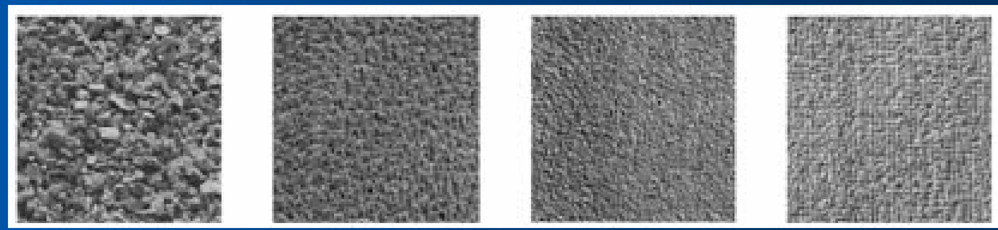
$$y = \begin{cases} 1, & \text{if } \boldsymbol{x} \in \omega_1 \\ 0, & \text{if } \boldsymbol{x} \in \omega_2 \end{cases}$$



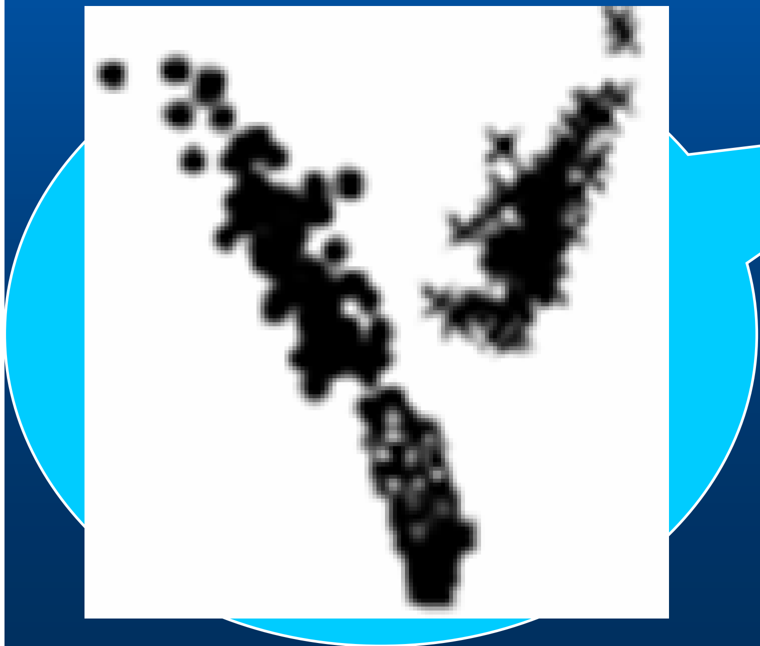
Applications: classification (2)

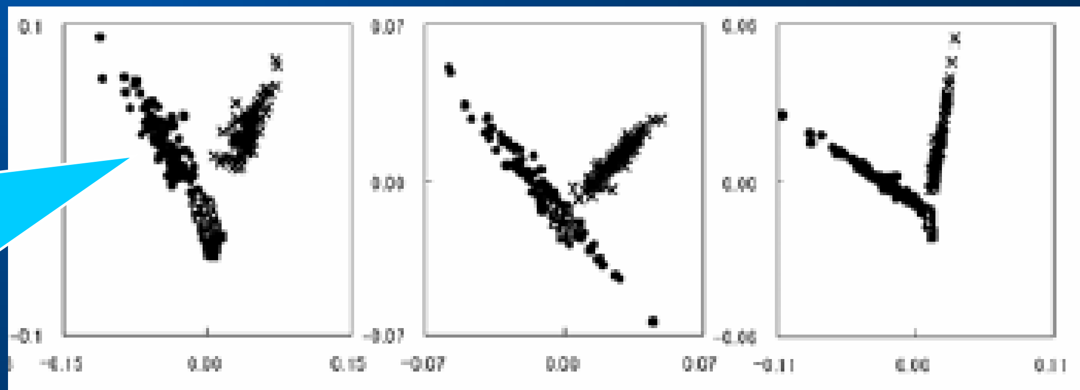
Multi-class recognition

[Yo2004]



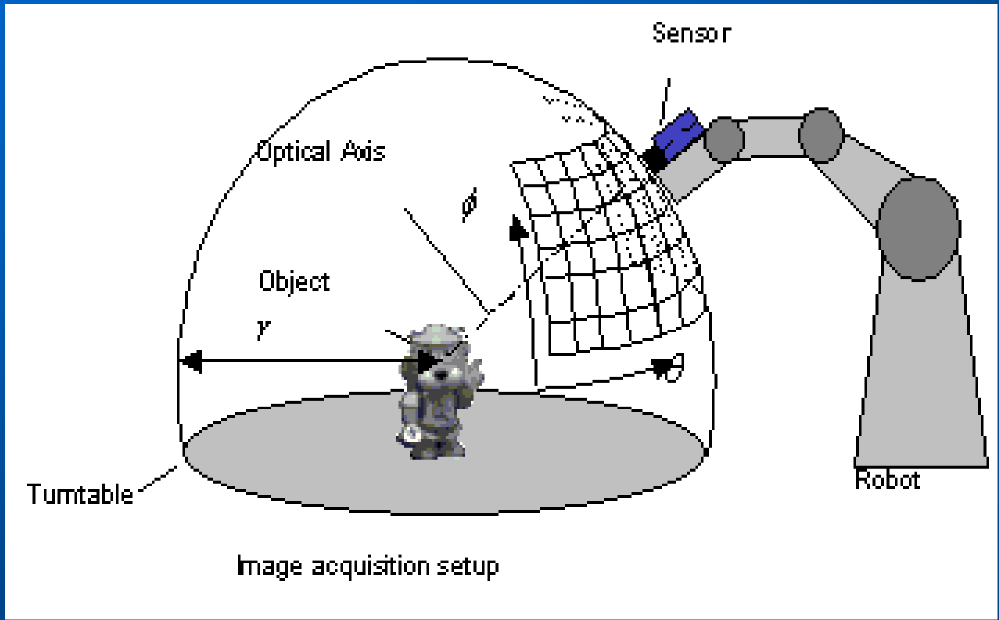
Samples from 4 classes

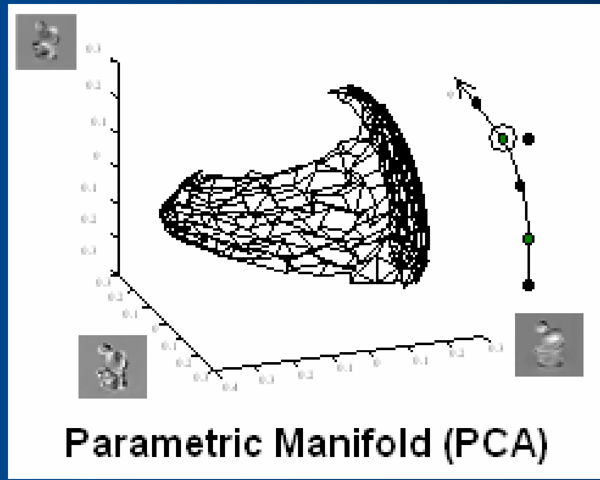


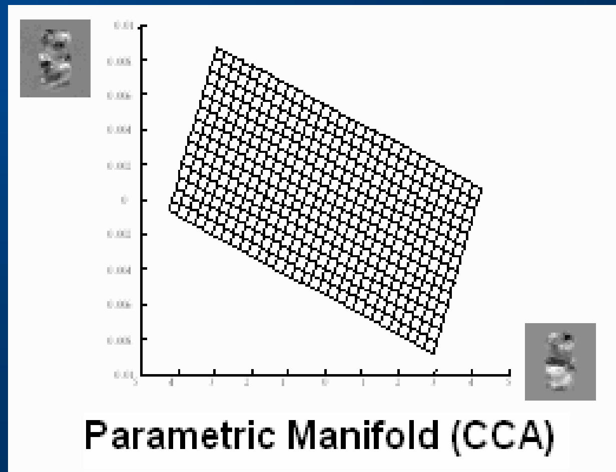


Separation between classes

Applications: pose estimation







In this case, CCA is more suitable for clear description of manifold than PCA, and result shows that CCA outperforms PCA in estimation accuracy. [MRB2003] (PR)

Part 2: Limitations and solutions

1. Limitations of CCA

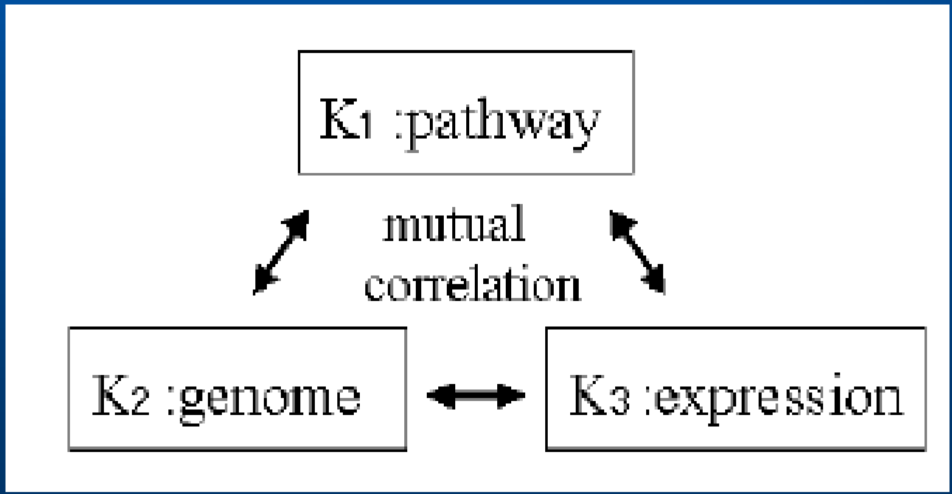
2. Solutions to break the limitations

Limitations of CCA

- 1. Only two data sets considered, e.g. X and Y
- 2. Small sample size (SSS) problem
- 3. Linearity
- 4. Globality (sometimes, merit)

Limitation 1: only *X* & *Y*

In real world, such cases are often encountered that multiple (>2) rather than just 2 factors correlate together, so common CCA (considering only 2 data sets) can not deal with such cases. **[YVN2003]**



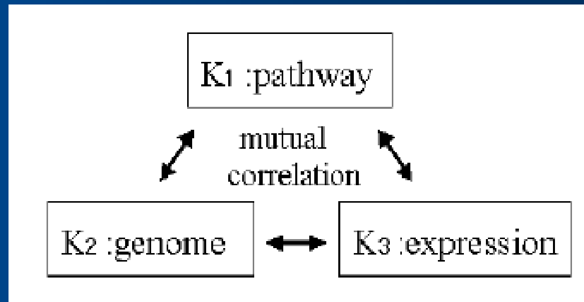
Pathway represents a higher level of biological functions than single genes.

Solutions to Limitation 1

Multisets CCA (mCCA) [Ket1971,Nie2002,YVN2003]

Goal: maximize the sum of correlation between any two data sets

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{n1} & \Sigma_{n2} & \cdots & \Sigma_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \lambda \begin{bmatrix} \Sigma_{11} & 0 & \cdots & 0 \\ 0 & \Sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$



where $\sum_{ij}$ denotes the covariance between data set X_i and X_j, when n=2, mCCA reduces into CCA.

Limitation 2 : $(XX^T)^{-1}$ 奇异性

When CCA is performed, $(XX^T)^{-1} \in R^{p\times p}$ is needed, where $X=[x_1,\ldots,x_n] \in R^{p\times n}$, p is the sample dimensionality and n the sample set size.

In general, $(XX^T)^{-1}$ exists,

however,

when dealing with images, e.g., 200 images of size 112×92, $n<<p$, so

$$\mathrm{rank}(XX^T) = \mathrm{rank}(X)<=\min(n,p)=n$$

but $(XX^T)^{-1} \in R^{p\times p}$, thus small sample size (SSS) problem occurs, resulting in the non-existence of $(XX^T)^{-1}$. CCA is infeasible.

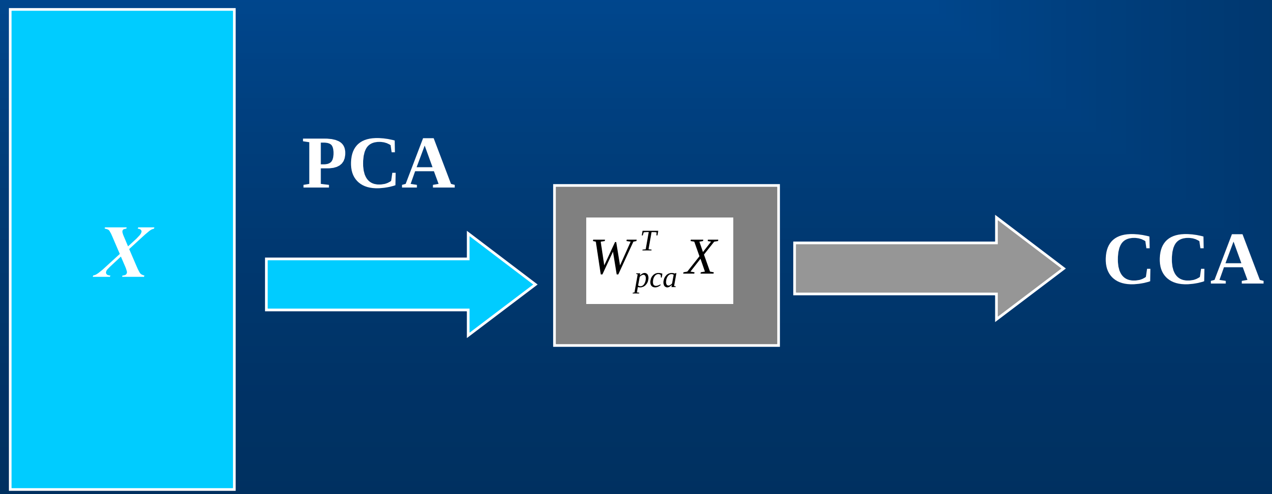
Solutions to Limitation 2

When SSS problem occurs, $(XX^T)^{-1}$ does not exists. However, we attack it by

1) performing PCA on the original samples;

2) performing CCA to reduce the dimensionality.

Such a preprocess of dimensionality by PCA is not only necessary in application, but also meaningful in algebra theory. [SZL2005]



Solutions to Limitation 2

regularization technique

i.e.

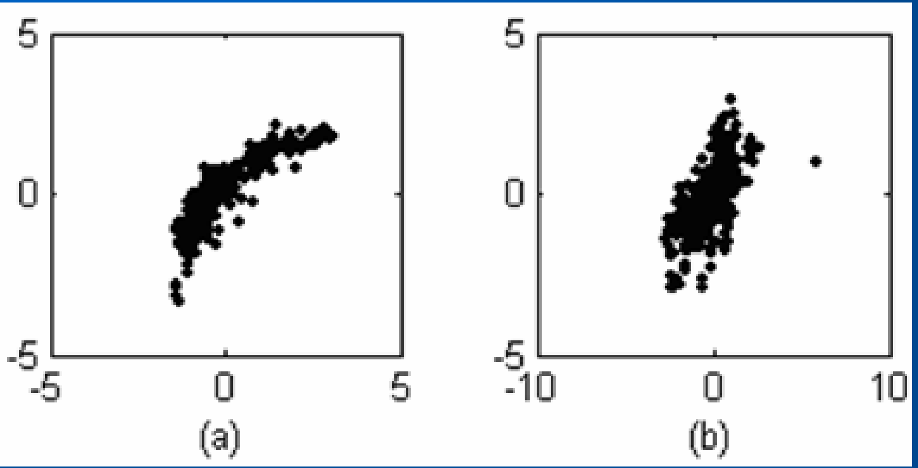
replace $(XX^T)^{-1}$ with $(XX^T + kI)^{-1}$

where I denotes identity matrix,

k the regularization parameter.

However, the optimal value of k is not easily specified. **[MRB2003, Fri1989]**

Limitation 3: Linearity



[Mat2005]

CCA is a global and linear methodology but frustrated when dealing with the data involving nonlinear correlation, maybe there is a nonlinear relationship, which can not be discovered by CCA, so CCA fails to discover the possible nonlinear relationship hidden between the data sets due to the ignorance of the details of local structures.

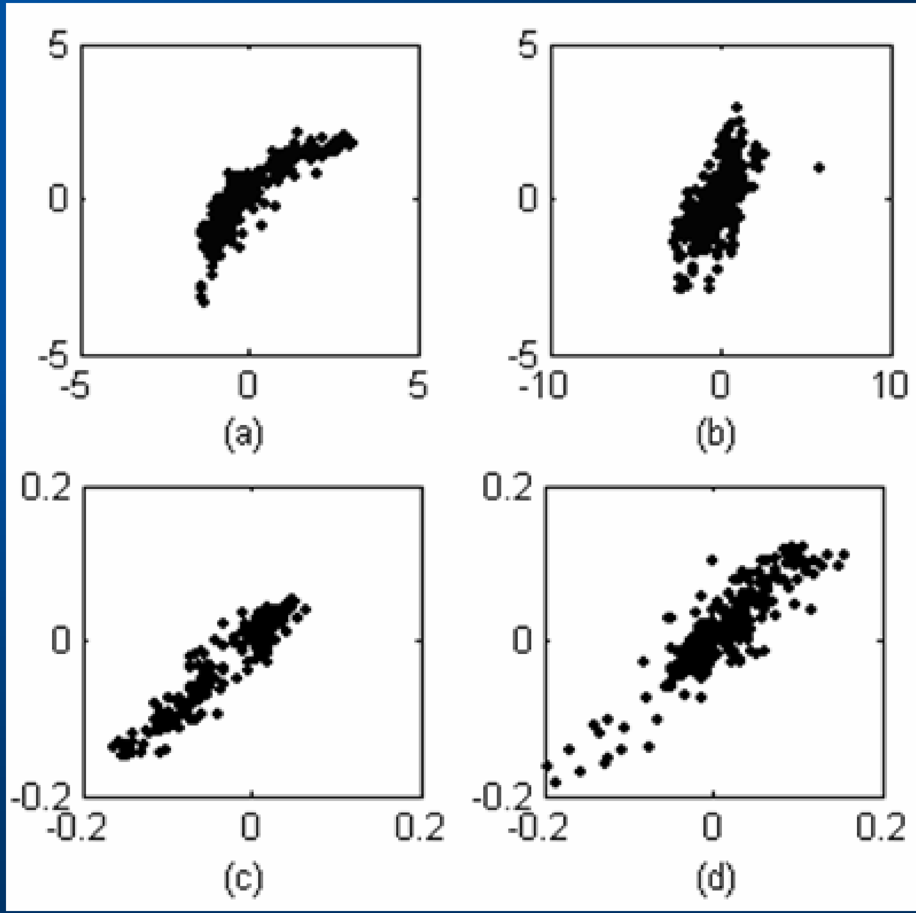
Solution to Limitation 3

CCA

CCA on

$$\phi_x : x \mapsto \phi_x(x)$$
$$\phi_y : y \mapsto \phi_y(y)$$

$\longrightarrow$ **KCCA**



Solutions to Limitation 3

There are other nonlinear versions of CCA,
e.g., CCA based on neural networks, with them,
--- nonlinear correlation can be discovered.

But the neural networks suffer from some intrinsic problems:
long-time training, slow convergence and local minima.
[Lai2000, GF2004]

Limitation 4: Globality

CCA is global and linear;

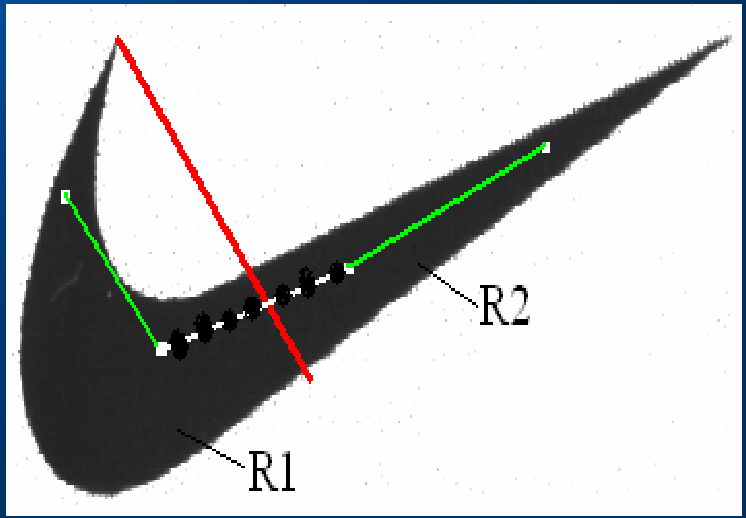
KCCA is nonlinear but global.

Both ignore the details of the local structure, the nonlinear mapping $\phi_x(\cdot)$ is uniform anywhere, i.e, the induced kernel function, e.g., $K_x(x_1, x_2) = \exp\left(-\|x_1 - x_2\|^2 / 2\sigma^2\right)$ is applied to all data pairs.

Suppose x_1, x_2 in R1, σ is set to $\sigma 1$

x_3, x_4 in R2, σ is set to $\sigma 2$.

What σ if x_1 (R1) & x_3 (R2) ?



Limitations of KCCA

1) The choice of both the problem-dependent kernel function and its parameters, e.g. σ in RBF kernel, is still a sticky problem and also a hot research topic. [ACS2005, PAM2003]

2) The overfitting problem, i.e., all eigenvalues equal to one for KCCA, can be attacked by the regularization technique, i.e.

replace the kernel matrix $\boldsymbol{K_x}^2$

with $\boldsymbol{K_x}^2 + kI$ or $(\boldsymbol{K_x} + kI)^2$ [MRB2003, TC2004]

Possible solutions to limitation 4

Both CCA and KCCA deal with correlation in global way, however, how do they do when some complex, nonlinear problems are encountered ?

A *possible* solution to this limitation:

Locality based method !

Part 3: Our works

3.1 Locality based methods

1) introduction

2) applications

3) A review

3.2 Our works: Locality preserving CCA (LP-CCA)

1) deviation of LP-CCA

2) experiments based on LP-CCA

3) further extensions: KLP-CCA

4) experiments based on KLP-CCA

3.1 Locality based methods

Recently developed locality based methods:

LLE [RS2000],
Isomap [TSL2000],
Locality Preserving Projection (LPP) [HN2004],
Locality Pursuit Embedding (LPE) [MLH2004],
Local PCA [KL1997],
Locally LDA [KK2005],
Eigenmap [BN2001],
etc.

We will demonstrate LPP & LPE.

Locality Preserving Projections (LPP)

LPP $$a = \arg \min_{a^T X D X^T a = 1} \sum_i \sum_j \boldsymbol{w}_{ij} a^T \left(x_i - x_j \right) \left(x_i - x_j \right)^T a$$

w_{ij} incurs a heavy penalty if neighboring points $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are mapped far apart. Then minimizing it is an attempt to ensure that if $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are "close" then $\boldsymbol{a^T x_i}$ and $\boldsymbol{a^T x_j}$ are close as well. [HN2004]

LPP is the generalization of PCA.

Applications of LPP

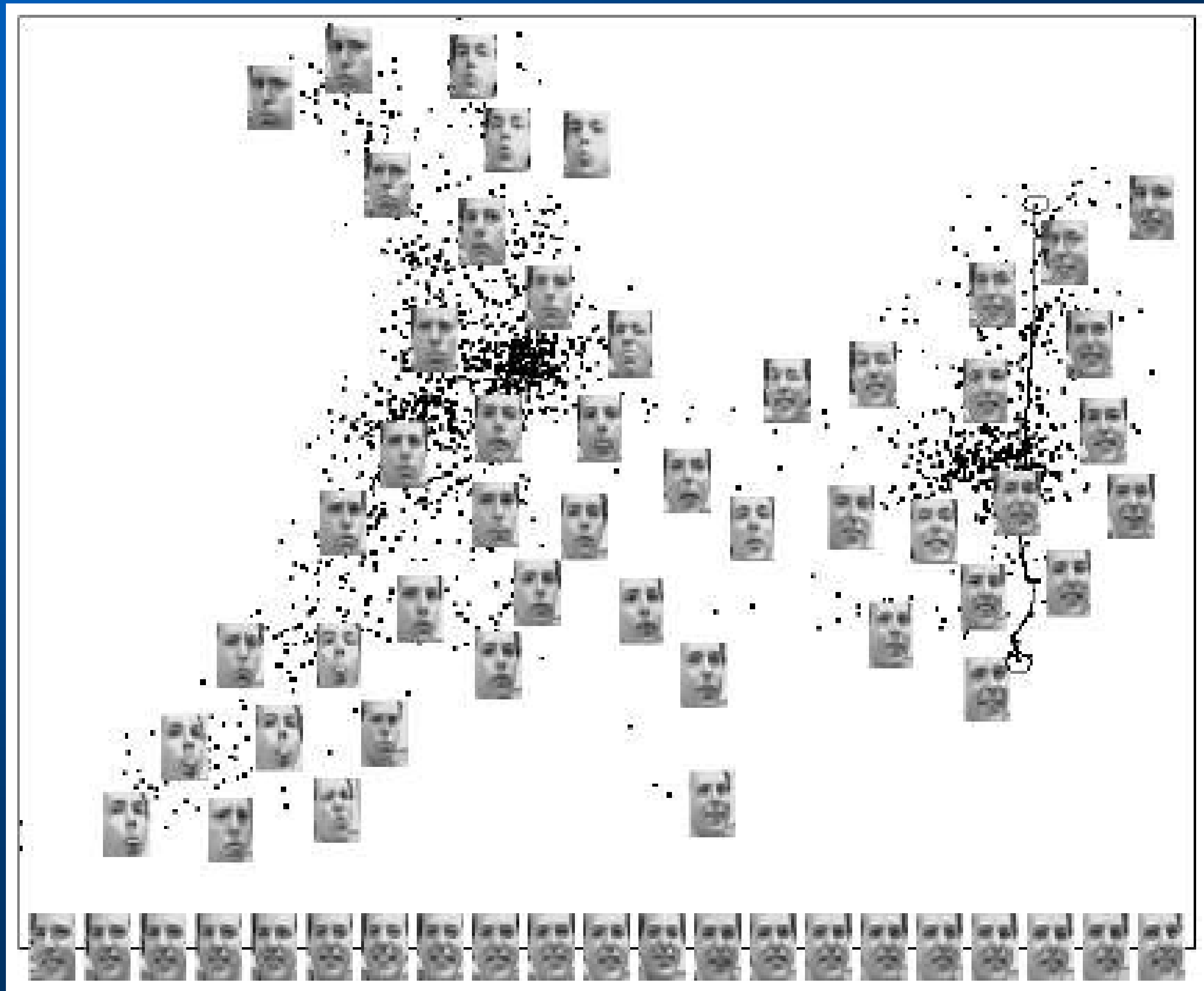
Visualization

Left:

closed mouth

Right:

open mouth



Pose changes consecutively. [HN2004]

Locality Pursuit Embedding (LPE)

PCA $\max \sum_i \left\| y_i - \bar{y} \right\|^2$ where $y_i = a^T x_i$, $\bar{y}$ is the *total* mean

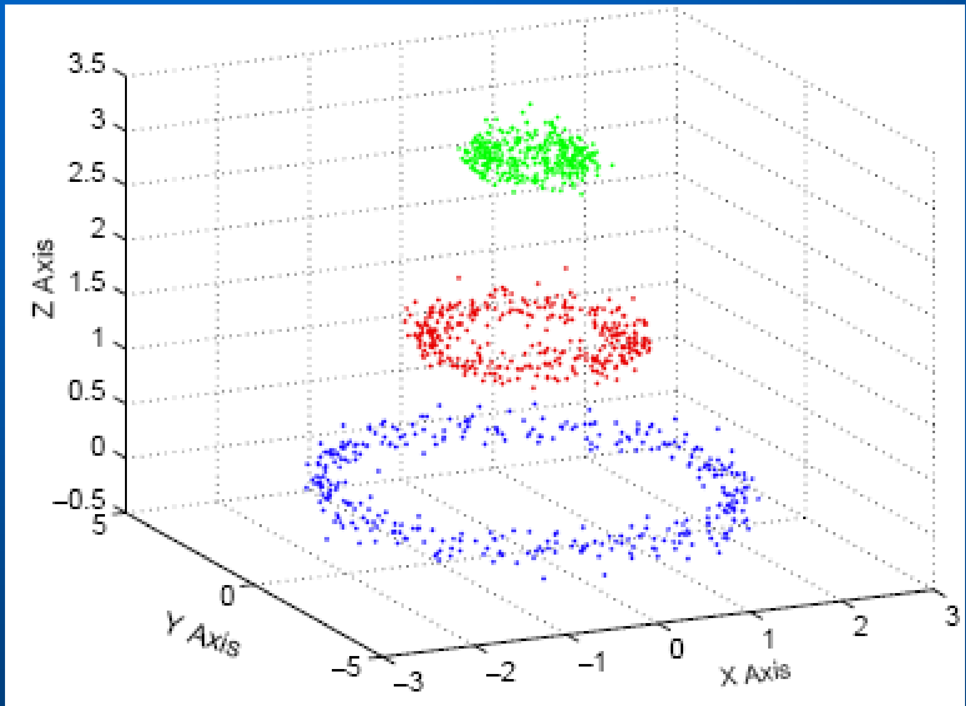
LPE $\max \sum_{i=1}^{n} \sum_{k \in ne(i)} \left\| y_k - \bar{y}^{(i)} \right\|^2$ where $\bar{y}^{(i)}$ is (***local***) mean vector of the neighbors of y_i.

PCA maximize the global variance,

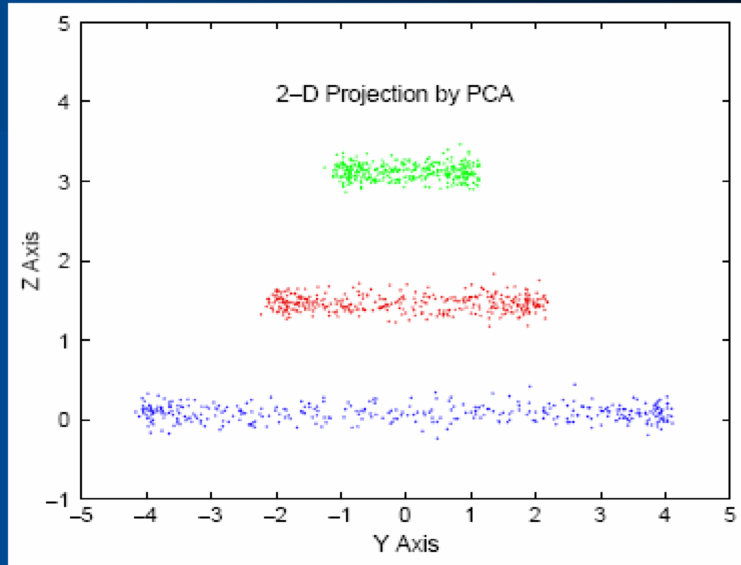
while LPE maximize the sum of the local variance.

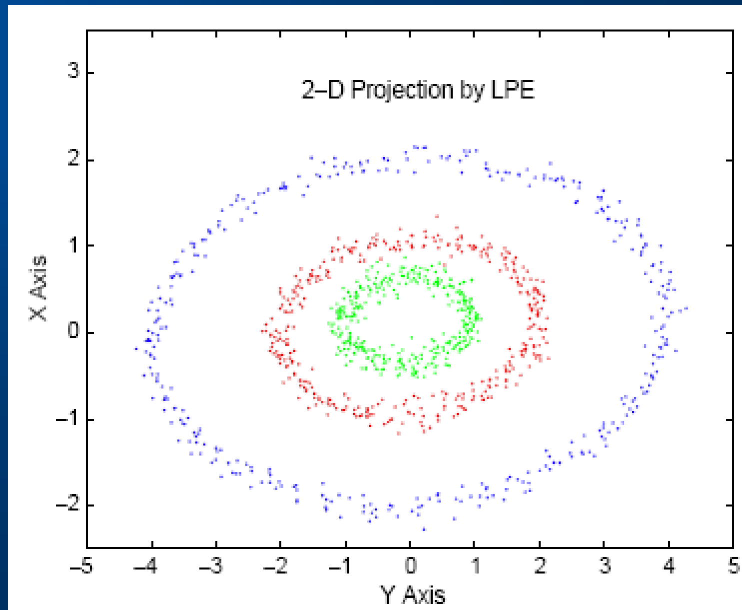
LPE is another generalization version of PCA.[MLH2004]

Application of LPE



3-D plot of samples [MLH2004]





A review of some locality based methods

1) Linear methods, e.g. PCA, LDA, CCA, etc
   advantages: easy computation
   disadvantages: weaker ability for nonlinear cases
2) Some nonlinear methods, e.g. LLE, Isomap
   advantages: able to deal with nonlinear cases
   disadvantages: relatively complex in computation
3) Advantages of some nonlinear methods, e.g. LPP and LPE
   advantages: easy computation
   disadvantages: insufficiently strong nonlinear processing ability, etc.

perform nonlinear dimensionality reduction via linear style.

Globally nonlinear, but locally linear.

3.2 Locality Preserving CCA (LP-CCA)

Our works:

1) **Introduction of locality into CCA**
2) **Deviation of LP-CCA**
3) **Experiments**
   - i) **Data visualization**
   - ii) **Pose estimation using LP-CCA**

4) **Kernel LP-CCA (KLP-CCA)**
   - i) **Formula of kernel LP-CCA**
   - ii) **Pose estimation using KLP-CCA**

Introduction of locality into CCA

CCA

$$\max \ \boldsymbol{w}_x^T \cdot \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)^T \cdot \boldsymbol{w}_y$$

$$\text{s.t.} \quad \boldsymbol{w}_x^T \cdot \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^T \cdot \boldsymbol{w}_x = 1$$

$$\boldsymbol{w}_y^T \cdot \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)^T \cdot \boldsymbol{w}_y = 1$$

Recall that a linear relationship only holds approximately in local region. So large $\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|$ or $\left\|\boldsymbol{y}_i - \boldsymbol{y}_j\right\|$ should be avoided for locality preserving.

Introduction of locality into CCA (2)

Define the ***local*** correlation in the neighborhood of $(\boldsymbol{x}_i, \boldsymbol{y}_i)$

as $$\boldsymbol{w}_x^T \cdot \sum_{j \in \mathrm{ne}(i)} \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)^T \cdot \boldsymbol{w}_y$$ or

$$\boldsymbol{w}_x^T \cdot \sum_{j=1}^{n} S_{ij}^x \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) S_{ij}^y \left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)^T \cdot \boldsymbol{w}_y$$

where local information factor $\boldsymbol{S_{ij}}$ can be define as

$$S_{ij}^x = \begin{cases} \exp\left(-\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2 / t_x\right), \text{if } \boldsymbol{x}_j \in \mathrm{LN}\left(\boldsymbol{x}_i\right) \text{ or } \boldsymbol{x}_i \in \mathrm{LN}\left(\boldsymbol{x}_j\right) \\ 0, \ \text{otherwise} \end{cases}$$

$$S_{ij}^y = \begin{cases} \exp\left(-\left\|\boldsymbol{y}_i - \boldsymbol{y}_j\right\|^2 / t_y\right), \text{if } \boldsymbol{x}_j \in \mathrm{LN}\left(\boldsymbol{x}_i\right) \text{ or } \boldsymbol{x}_i \in \mathrm{LN}\left(\boldsymbol{x}_j\right) \\ 0, \ \text{otherwise} \end{cases}$$

Note: “LN” denotes local neighborhoods.

Introduction of locality into CCA (3)

LP-CCA

$$\max\ \boldsymbol{w}_x^T \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}^{x}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) S_{ij}^{y}\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)^T \cdot \boldsymbol{w}_y$$

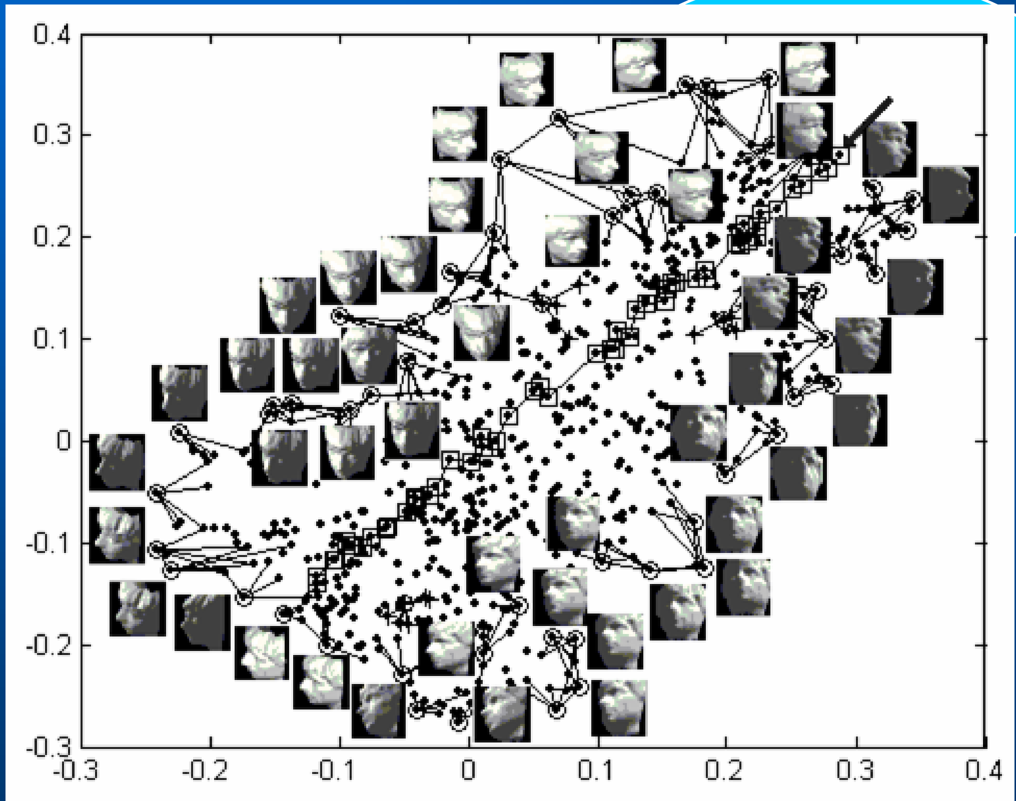
$$\text{s.t.}\quad \boldsymbol{w}_x^T \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}^{x2}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^T \cdot \boldsymbol{w}_x = 1$$

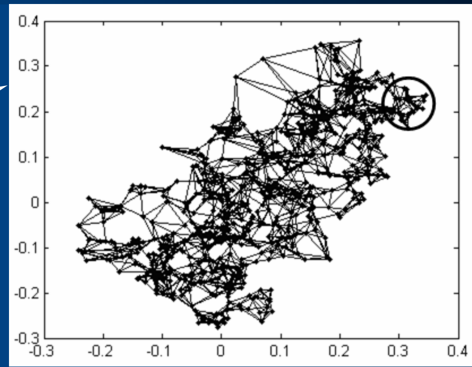
$$\boldsymbol{w}_y^T \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}^{y2}\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)^T \cdot \boldsymbol{w}_y = 1$$

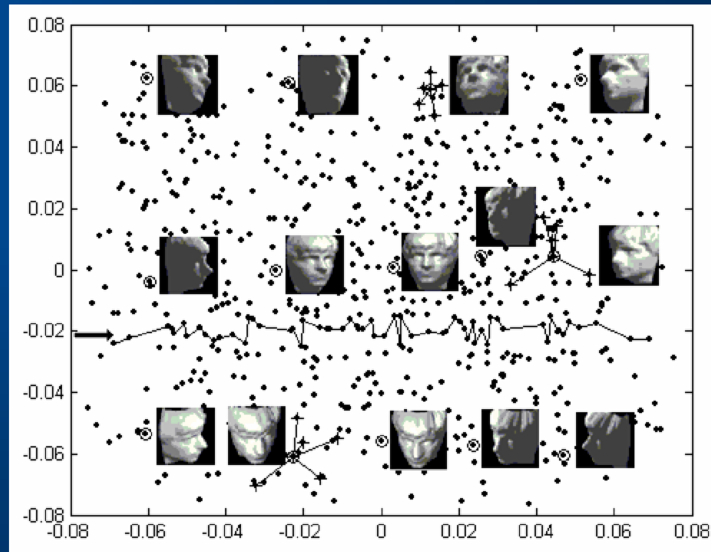
$$\Rightarrow \begin{pmatrix} & \boldsymbol{X}\boldsymbol{S}_{xy}\boldsymbol{Y}^T \\ \boldsymbol{Y}\boldsymbol{S}_{xy}\boldsymbol{X}^T & \end{pmatrix}\begin{pmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{X}\boldsymbol{S}_{xx}\boldsymbol{X}^T & \\ & \boldsymbol{Y}\boldsymbol{S}_{yy}\boldsymbol{Y}^T \end{pmatrix}\begin{pmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{pmatrix}$$

where $X=[x_1,\ldots,x_n]$, $Y=[y_1,\ldots,y_n]$, $\boldsymbol{S}_{xx}$, $\boldsymbol{S}_{yy}$ and $\boldsymbol{S}_{xy}$ are Laplacian matrices containing local information.

Experiment: data visualization

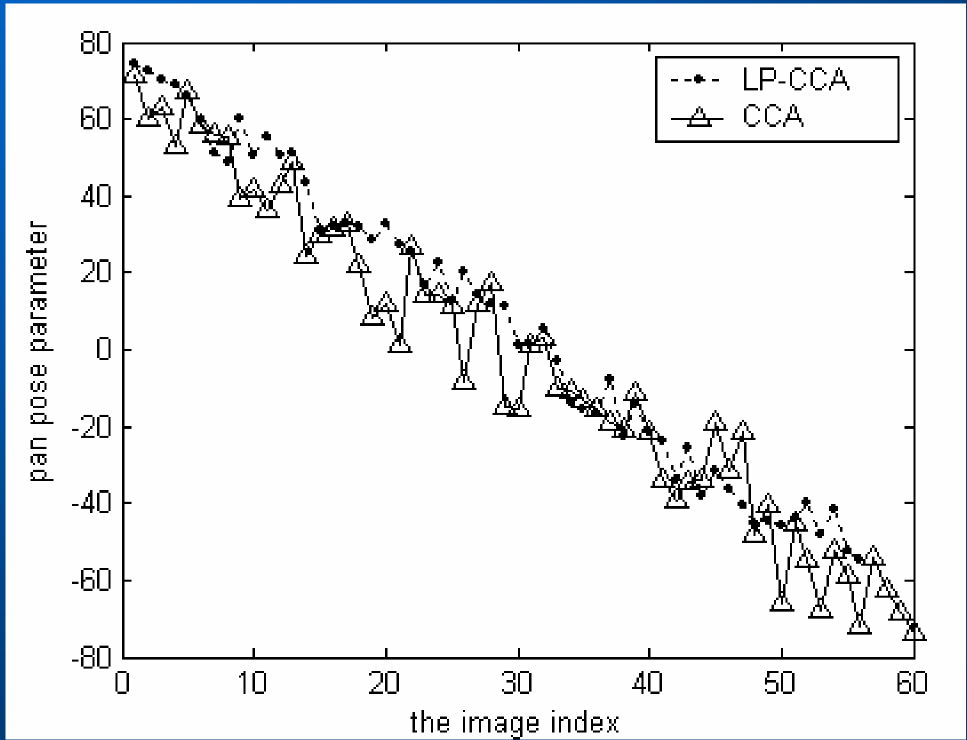






LP-CCA **CCA**

Experiment: data visualization



CCA ***locally*** fluctuates more heavily than LP-CCA, indicating that LP-CCA is possibly more suitable for pose estimation.

Pose estimation - introduction

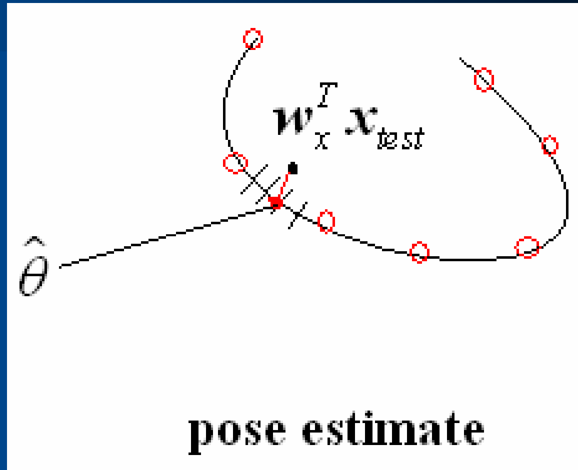
One of the main goals of an intelligent vision system is to recognize objects in an image and compute their poses in the three-dimensional scene.

Its applications include:

visual inspection, robot vision to autonomous navigation [MN1995, NNM1996a].

So pose estimation has become one of the active research topics in computer vision field [MRB2003 , MN1995, RYS2004].

Pose estimation: strategy



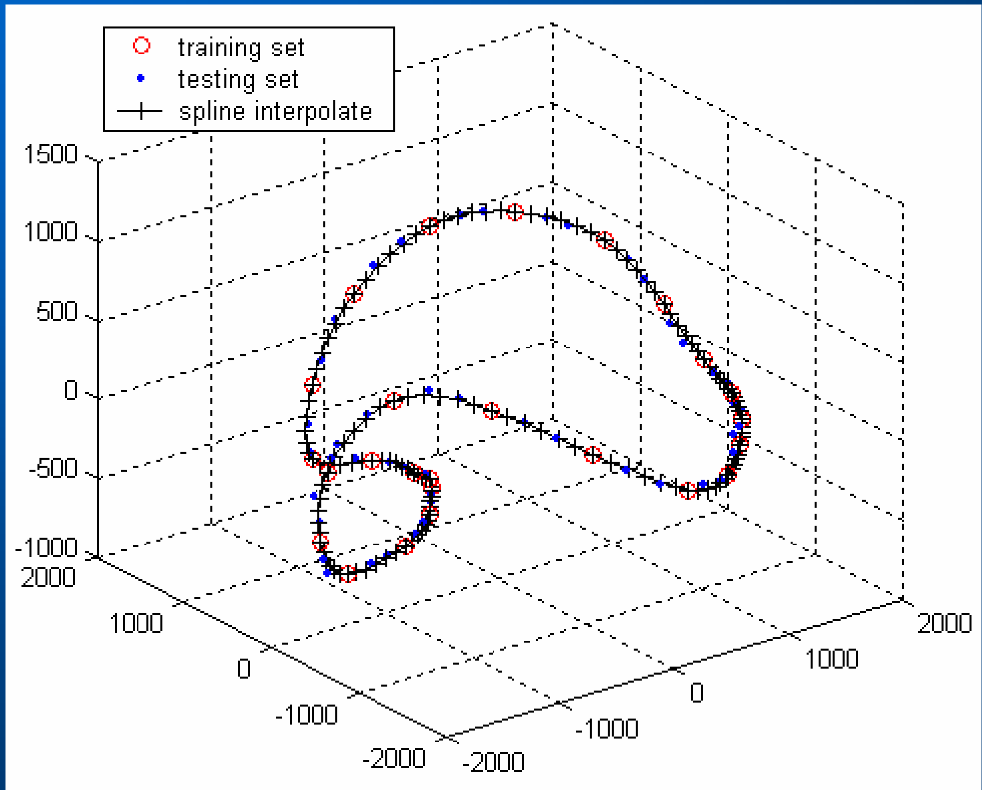
Step 1: perform dimensionality reduction for training samples "O" .

Step 2: perform resampling (using e.g. cubic spline interpolation) to obtain many "+" points to construct the parametric manifold. The pose parameters of "+" points are known.

Step 3: perform dimensionality reduction for testing samples (dark point), searching for its nearest neighbor, and the corresponding pose parameter is just that to be estimated. [MN1995, MRB2003]

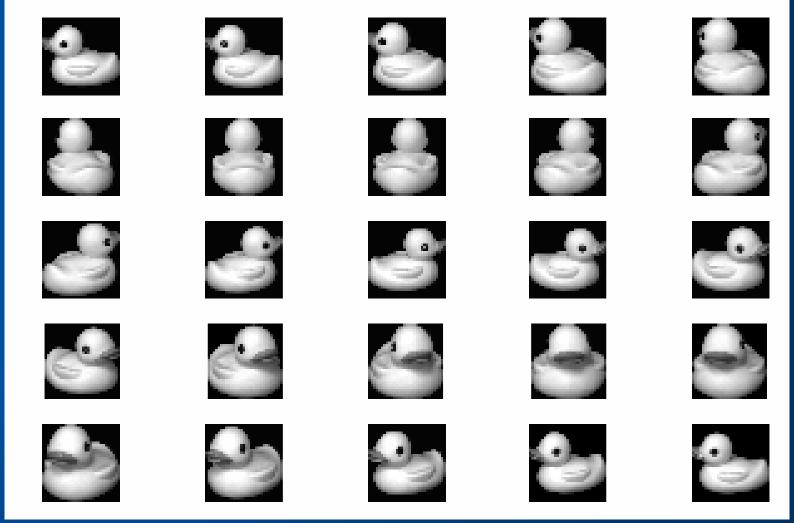
The essence of pose estimation is regression and prediction!

An example of manifold



an examples of manifold for pose estimation

Pose estimation

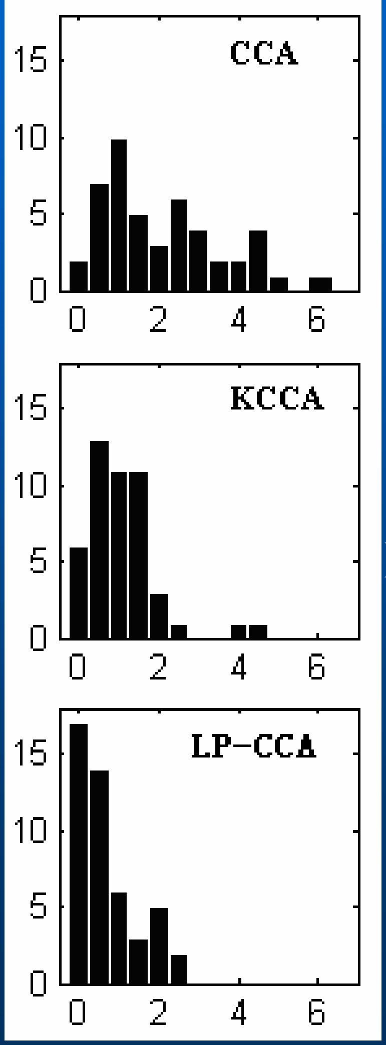


Subjects 1-20 (left) and training samples of subject 1 (right)

[NNM1996b]

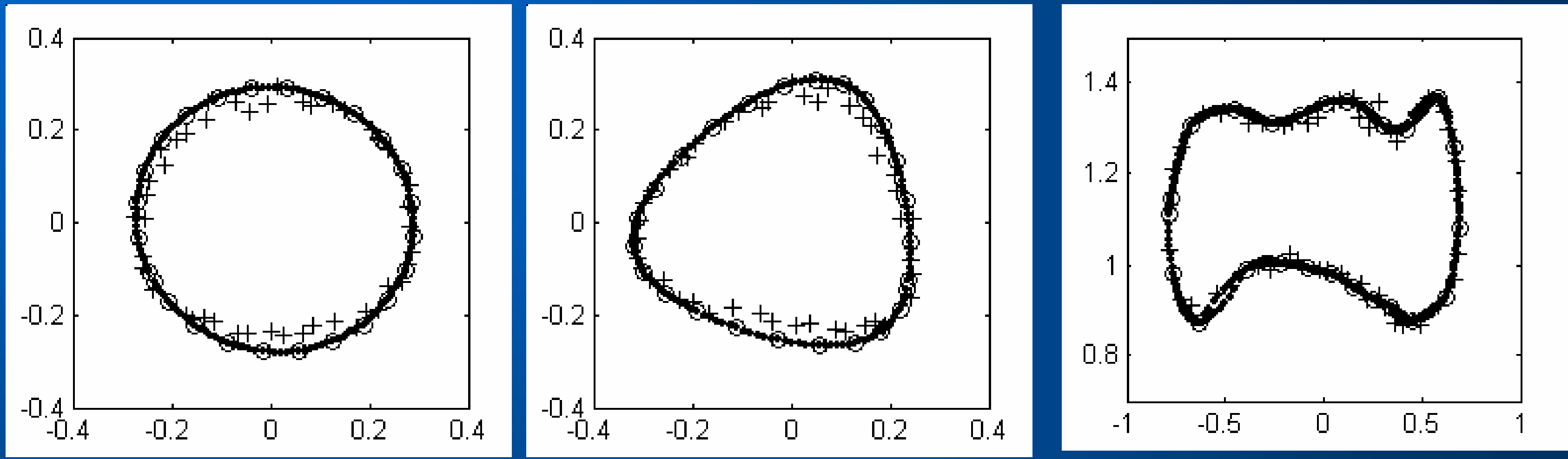
25 frames (every 15 deg) each subject are used for training, the remaining 47 frames for testing.

Pose estimation: results



subject	CCA		KCCA		LP-CCA	
	mean	std	mean	std	mean	std
1	0.59	0.60	0.65	0.59	**0.52**	**0.57**
2	1.78	1.75	**1.17**	**1.19**	1.35	1.32
3	4.62	3.60	**2.82**	**4.06**	3.51	3.94
4	1.19	1.13	0.95	0.70	**0.85**	**0.82**
5	3.17	2.64	2.15	1.42	**1.98**	**1.56**
6	15.03	31.77	13.16	41.03	**5.86**	**5.15**
7	2.10	1.53	2.40	3.38	**1.40**	**1.01**
8	2.61	2.56	1.77	1.79	**1.06**	**1.10**
9	12.12	37.21	8.29	35.50	**4.57**	**3.75**
10	*2.08*	*1.47*	*1.10*	*0.91*	***0.70***	***0.71***
11	4.07	3.47	**2.19**	**2.13**	2.47	2.45
12	2.28	2.22	2.17	2.53	**1.99**	**1.75**
13	2.79	2.68	**1.27**	**1.19**	2.02	1.87
14	14.83	50.48	18.19	70.78	**5.73**	**7.08**
15	2.45	2.26	2.88	2.49	**1.41**	**1.60**
16	1.31	1.02	2.78	2.14	**1.08**	**1.07**
17	3.14	2.53	4.12	4.11	**2.06**	**1.89**
18	4.23	3.78	**3.80**	**3.71**	3.83	3.86
19	14.64	50.11	9.66	49.47	**6.57**	**6.95**
20	2.18	1.83	2.28	1.90	**2.04**	**1.68**

Pose estimation: manifold



Parametric manifolds for pose estimation, respectively obtained by **CCA (left), KCCA (middle)** and **LP-CCA (right).**

The nature of pose estimation is regression and prediction, so the large deviation should be avoided as much as possible to improve the estimation accuracy.

Further study – Kernel LP-CCA

Dimensionality reduction of LP-CCA

is still linear although it achieve the effect of nonlinear dimensionality reduction to some degree.

LP-CCA can be non-linearized by , e.g., “kernel trick” .

Goal: fuse the advantages of locality- and kernel-based methods together to further improve the estimation accuracy.

Kernel LP-CCA

LP-CCA
$$\begin{pmatrix} & XS_{xy}Y^T \\ YS_{xy}X^T & \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} XS_{xx}X^T & \\ & YS_{yy}Y^T \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix}$$

where $w_x = X\alpha$ and $w_y = Y\beta$

(dual representation), so

LP-CCA can be described as

$$\max\ \alpha^T K_x S_{xy} K_y \beta$$
$$\text{s.t.}\quad \alpha^T K_x S_{xx} K_x \alpha = 1$$
$$\beta^T K_y S_{yy} K_y \beta = 1$$

Finally we obtain

$$\begin{cases} K_x S_{xy} K_y \beta = \lambda K_x S_{xx} K_x \alpha \\ K_y S_{xy} K_x \alpha = \lambda K_y S_{yy} K_y \beta \end{cases}$$

as the formula of **KLP-CCA**.

KLP-CCA: pose estimation results

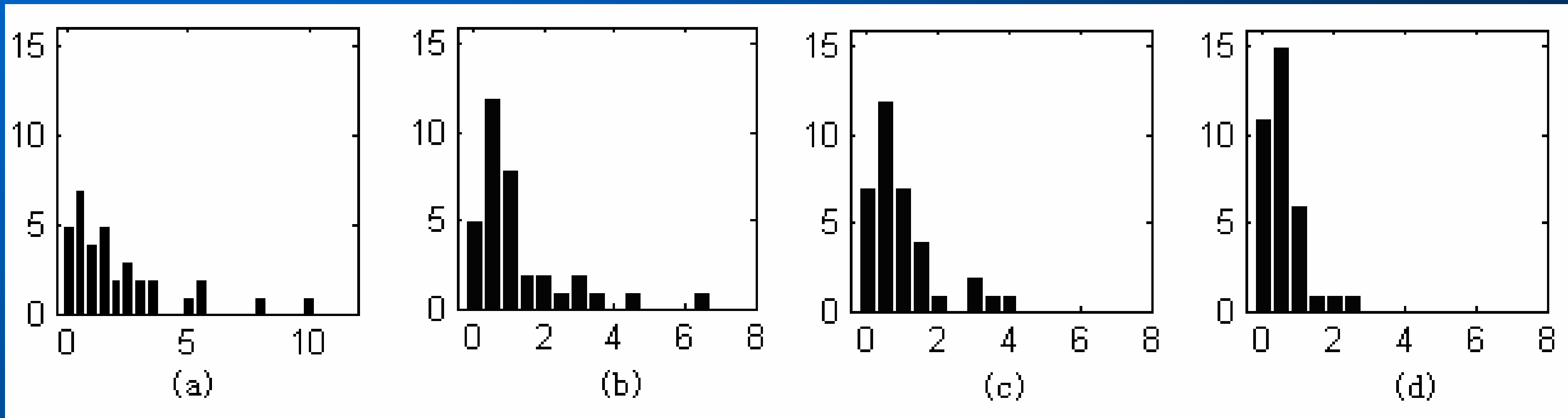
subject	CCA		KCCA		LP-CCA		KLP-CCA	
	mean	std	mean	std	mean	std	mean	std
1	**0.44**	**0.38**	0.55	0.55	**0.42**	**0.36**	0.46	0.43
2	1.78	1.73	**1.17**	**1.17**	1.32	1.23	**0.55**	**0.66**
3	3.93	3.31	2.04	2.28	**1.84**	**1.40**	**0.60**	**0.64**
4	1.01	0.84	0.90	0.65	**0.81**	**0.74**	**0.56**	**0.46**
5	1.87	1.42	1.25	0.98	**0.76**	**0.91**	**0.46**	**0.42**
6	3.86	5.72	2.37	4.70	**2.10**	**1.87**	**0.76**	**0.75**
7	1.36	1.01	1.35	1.74	**0.93**	**0.81**	**0.43**	**0.38**
8	2.12	2.31	1.24	1.38	**1.00**	**1.02**	**0.55**	**0.56**
9	6.69	9.28	9.53	41.02	**2.32**	**1.70**	**0.96**	**0.61**
10	1.06	1.04	0.73	0.55	**0.49**	**0.37**	**0.47**	**0.45**
11	3.22	2.80	**1.57**	**1.21**	2.26	1.87	**0.34**	**0.38**
12	1.86	1.76	1.97	2.24	**1.52**	**1.39**	**1.78**	**1.66**
13	1.73	1.57	**0.91**	**0.79**	0.94	0.98	**0.62**	**0.74**
14	4.02	3.77	1.91	2.35	**1.94**	**1.73**	**1.01**	**1.27**
15	2.26	2.38	2.11	1.72	**0.62**	**0.53**	**0.85**	**1.03**
16	1.43	1.36	3.46	3.46	**1.01**	**0.66**	**1.27**	**1.20**
17	2.54	2.27	3.08	3.35	**1.10**	**1.33**	**1.97**	**1.78**
18	2.10	1.77	2.34	2.31	**1.64**	**1.32**	**1.31**	**1.16**
19	14.41	57.47	11.85	57.34	**3.20**	**2.36**	**0.50**	**0.50**
20	1.75	1.40	1.74	1.74	**1.33**	**1.10**	**1.65**	**1.50**

Optimal

Ilustration on
next slide

Sub-optimal

pose estimation results (cont'd)



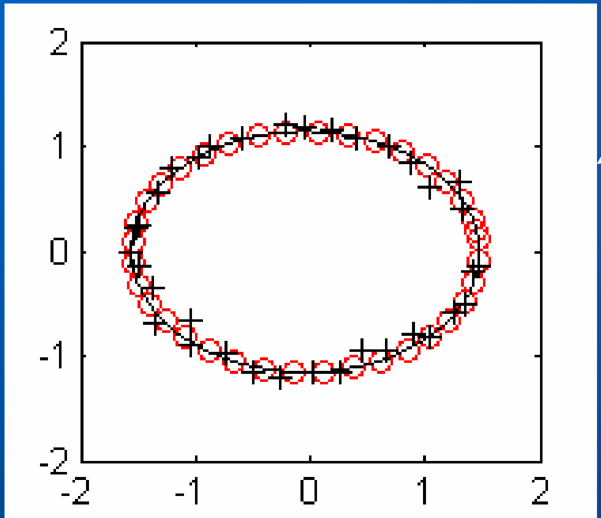
Histogram of error magnitude .

(a).CCA (b).KCCA (Gaussian) (c).LP-CCA (d).KLP-CCA

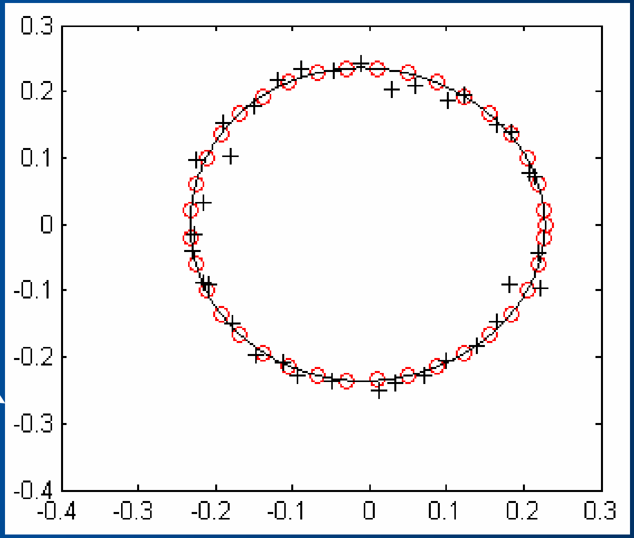
For error magnitude and its standard variance, **(a) > (b) > (c) > (d)** holds.

Smaller errors will result in smaller standard variance, which indicates the more stable estimation.

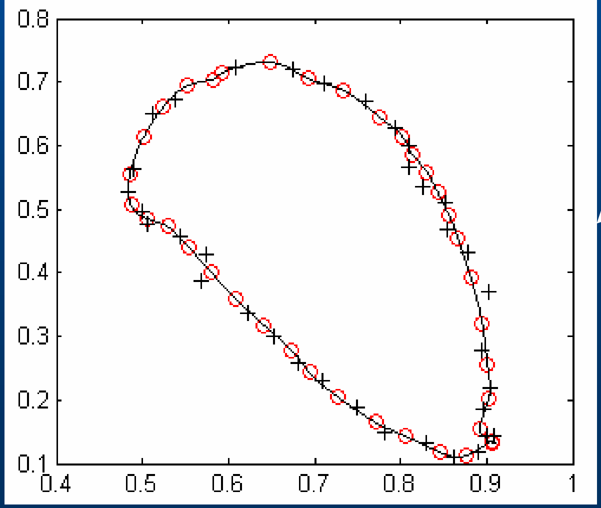
pose estimation results (cont'd)



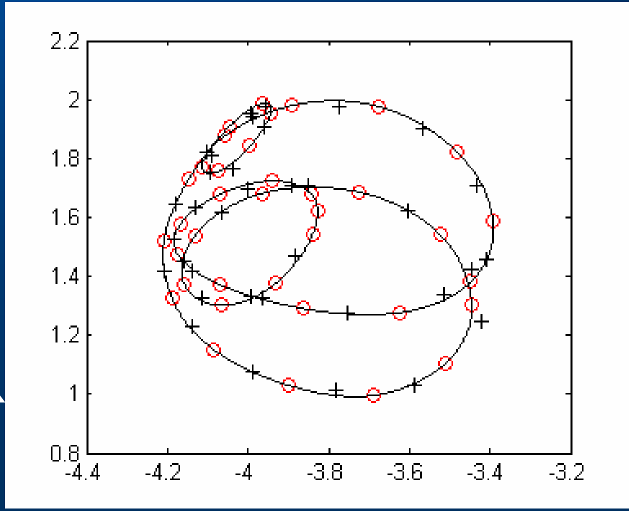
CCA



KCCA

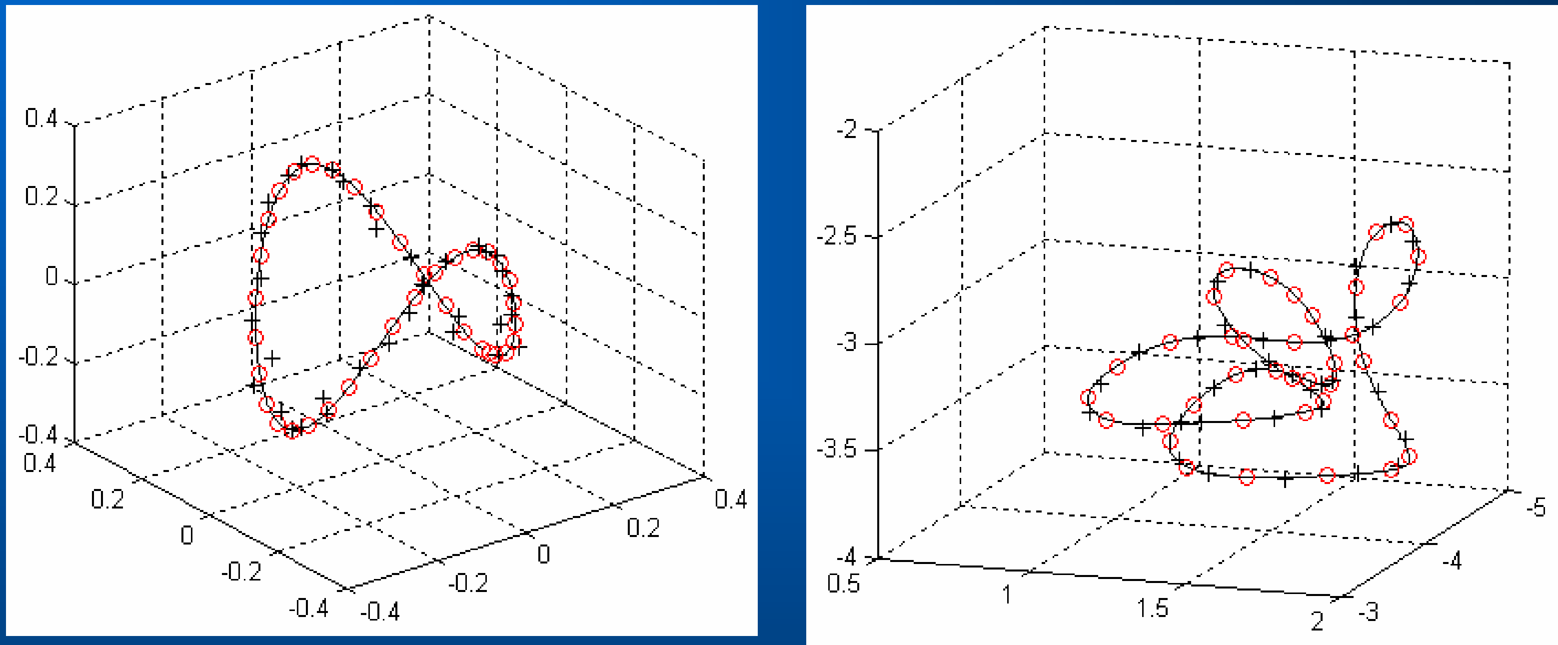


LP-CCA



KLP-CCA

pose estimation results (cont'd)



3D plot of parametric manifold for KCCA (left) and KLP-CCA (right), the symbol O denotes training samples, + the testing samples, solid line the points of interpolation.

Smaller deviation of testing points from the manifold will benefit to more stable estimation.

Reference (1)

[ACS2005] N.E. Ayat, M. Cheriet, C.Y. Suen, Automatic model selection for the optimization of SVM kernels, Pattern Recognition, 38(2005), 1733 – 1745.
[BF1985] L. Breiman and J. H. Friedman, Estimating optimal transformations for multiple regression and correlation, J. Amer. Statist. Assoc., 80(1985) 580-619.
[BK2001] M. Borga & H. Knutsson, A canonicial correlation approach to blind source separation. Tech. Report LiU-IMT-EX-0062, dept. of Biomedical Engineering, Linkoping Univ., Sweden.
[Bor1999] Canonical correlation: a tutorial, at http://people.imt.liu.se/~magnus/cca/tutorial/, 1999.
[BL1983] E. van der Burg and J. de Leeuw, Non-linear canonical correlation, Brit. J. Math. Statist. Psychol., 36(1983), 54-80.
[BLK1992] M Borga, T Landelius, H Knutsson - , A Unified Approach to PCA, PLS, MLR and CCA, Report LiTH-ISY-R-1992, ISY, SE-581
[BR2003] Barker M., Rayens W.: 'Partial least squares for discrimination', Journal of Chemometrics, 2003, 17, pp.166-173
[Bor1998] M. Borga, Learning Multidimensional signal processing, ph.D thesis, Department of Electrical Engineering, Linköping University, Linköping, Sweden,1998.
[BN2001] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Proc. Conf. Advances in Neural Information Processing System 15, 2001.
[Buj1990] A. Buja, Remarks on functional canonical variates, alternating least squares methods and ACE, Ann. Statist., 18(1990)1032-1069.

Reference (2)

[ESE2005] E. Kidron, Y.Y. Schechner, M. Elad, Pixels that Sound, IEEE Proc. of Computer Vision and Pattern Recognition, 1, 88-95, June 2005

[Fis1936] Fisher R.A., The use of multiple measurements in taxonomic problems, Annals of Eugenics, 1936

[For2004] Blaz Fortuna, Kernel canonical correlation analysis with applications, In: SIKDD 2004 at Multiconference IS 2004, 12-15 Oct 2004, Ljubljana, Slovenia.

[FBL2001] O. Friman, M, Borga, P. Lundberg, H. Knutsson, Canonical correlation as a tool in functional MRI data analysis, SSAB 2001, Proceedings of the SSAB Symposium on Image Analysis, March, Norrköping, Sweden, 2001

[Fri1989] J.H. Friedman, Regularized discriminant analysis, Journal of the American Statistics Association, 84, pp.165-175, 1989.

[Fri2001] O. Friman, Detection of neural activation in functional MRI using CCA, Magnetic Resonance in Medicine, 45(2001), 323-330.

[GF2004] Zhenkun Gou, Colin Fyfe, A canonical correlation neural network for multicollinearity and functional data, Neural Networks 17 (2004) 285–293

[GSB2001] T.V.Gestel, J.A.K.Suykens, J. De Brabanter, B. De Moor, J.Vandewalle, Kernel canonical correlation analysis and least squares support vector machines，Proc. of the International Conference on Artificial Neural Networks (ICANN 2001) 384-389.

Reference (3)

[Hil2001] K. B. Hilger, Expolaratory analysis of multivariate data: Unsupervised image segmentation and data driven linear and nonlinear decomposition, Phd dissertation, Dept. Inform. Math. Modeling, Tech. Univ. Demark, Lyngby, Demark, http://www.imm.dtu.dk/~kbh, 2001.

[HN2004] X. He, P. Niyogi, Locality preserving projections, Advances in Neural Information Processing Systems, 2004

[HSS2004] D.R. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: An overview with application to learning methods, Neural Computation, 16(2004), 2639-2664.

[Hot1936] Hotelling H., Relations between two sets of variates, Biometrika, 28 (3-4) pp.312-377, 1936

[IHA2004] H. Ichihashi, K. Honda & S. Araki, Fuzzy canonical correlation and cluster analysis for brain mapping on long term memory consolidated by mnemonics, IEEE conf. FUZZ, 2004,pp. 155-160.

[KK2005] Tae-Kyun Kim, J. Kittler, Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image, IEEE Trans. Pattern Analysis and Machine Intelligence, 27(2005) :318 – 327.

[KL1997] N. Kambhatla, T. K. Leen, Dimension reduction by local principal component analysis. Neural Computation, 9(1997), 1493-1516.

Reference (4)

[Ket1971] J.R. Kettenring, Canonical analysis of several sets of variables, Biometrika, 58(1971), 433-451.

[Lai2000] P.L. Lai, Neural implementations of canonical correlation analysis, ph.D thesis, Dept. of computing and information systems, University of Paisley, Scotland, March 2000.

[LGD2005] M. Loog, B. van Ginneken, R. P.W. Duin, Dimensionality reduction of image features using the canonical contextual correlation projection, Pattern Recognition, 38(2005), 2409-2418.

[Mat2005] Mathworks Inc., Matlab 7.0 Release 14 help: Statistics toolbox, Jan. 2005

[MLH2004] W. Min, K. Lu, X. He, Locality pursuit embedding, Pattern Recognition, 37(2004), 781-788.

[MN1995] H. Murase, S.K. Nayar, Visual learning and recognition of 3-D objects from appearance, International J. Comp. Vision, 14(1995), 5-24.

[MRB2003] T.Melzer, M.Reiter, H.Bischof, Appearance models based on kernel canonical correlation analyusis, Pattern Recognition, 36(2003), 1961-1971.

[NCK1998] I. Niemeyer, M. Canty and D. Klaus, Possibilities and limits of remote sensing for the verification of international agreements: algorithms to detect changes in nuclear plants, in Proc. Int. Geoscience and remote sensing Symp. (IGASS), Seattle, WA, July 6-10, 1998, pp. 819-821.

[NCK1999] I. Niemeyer, M. Canty and D. Klaus, Unsupervised change detection techniques using multispectral satellite images, in Proc. Int. Geoscience and remote sensing Symp. (IGASS), Hamburge, Germany, 1999, pp. 327-329.

Reference (5)

[NGK2001] A. Nakaya, S. Goto and M. Kanehisa, Extraction of correlated gene clusters by multiple graph comparison, Genome Informatics, 12 (2001), 44-53.

[Nie1994] A.A. Nielsen, Analysis of regularly and irregularly sampled spatial, multivariate and multi-temporal data, Phd dissertation, Dept. Inform. Math. Modeling, Tech. Univ. Demark, Lyngby, Demark, http://www.imm.dtu.dk/~aa/phd/.

[Nie2002] A.A. Nielsen, Multiset Canonical correlations analysis and multispectral, truly multitemporal remote sensing data, IEEE Transactions on Image Processing 11(2002), 293-305.

[NNM1996a] S.K. Nayar, S.A. Nene, H. Murase, Subspace methods for robot vision, IEEE Trans. Robotics and Automation,12(1996), 750-758.

[NNM1996b] S. A. Nene, S. K. Nayar, H. Murase, Columbia Object Image Library (COIL-20), Technical Report CUCS-005-96, February 1996

[PAM2003] E. Parrado-Hernández, J. Arenas-García, I. Mora-Jiménez, A. Navia-Vázquez, On problem-oriented kernel refining, Neurocomputing, 55(2003), 135-150.

[RYS2004] B. Raytchev, I. Yoda, K. Sakaue, Head pose estimation by nonlinear manifold learning, IEEE Proc. of the 17th International Conf. on Pattern Recognition, 2004.

[RS2000] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science, 290(2000), 2323-2326.

[ST1992] S. G. Shi and W. Taam, Non-linear CCA with a simulated annealing solution, J. Appl. Statist., 19(2)(1992) 155-165.

Reference (6)

[SZL2005] Quan-Sen Sun, Sheng-Gen Zeng, Yan Liu, Pheng-Ann Heng, De-Shen Xia, A new method of feature fusion and its application in image recognition, Pattern Recognition, 38(2005), 2437-2448.

[TSL2000] J.B.Tenenbaum, Vin de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science, 290(2000), 2319-2323.

[TC2004] J.Shawe-Taylor, N.Cristianini, Kernel methods for pattern analysis, Cambridge University Press, U.K., 2004.

[Wan1999] H.W. Wang, Partial least squares regression method and applications, pp134-135, (in Chinese), National defense industry press, Beijing, 1999.

[Win1992] K. Windfeld, Application of computer intensive data analysis: Methods to the analysis of digital images and spatial data, PhD dissertation, Inst. Math. Statist. Oper. Res., Tech. Univ. Demark, Lyngby, Demark, 1992.

[VK2003] J. P., Vert and M. Kanehisa, Graph-driven features extraction from microarray data uding diffusion kernels and kernel CCA, NIPS 15, 2003.

[Yac2004] Yacov Hel-Or, The canonical correlations of color images and their use for demosaicing, HP Labs Technical Report, HPL-2003-164(R.1), Feb. 2004

[Yo2004] Yo Horikawa, Use of autocorrelation kernels in kernel canonical correlation analysis for texture classification, in N.R. Pal et al. (Eds.): ICONIP 2004, LNCS 3316, pp. 1235.1240, 2004. Springer-Verlag Berlin Heidelberg

[YVN2003] Y. Yamanishi, J.P Vert, A. Nakaya, M. Kanehisa , Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis, Bioinformatics, 19(2003), i323-330

Thanks!