A Unifying View of Component Analysis Methods (from a Computer Vision Perspective)

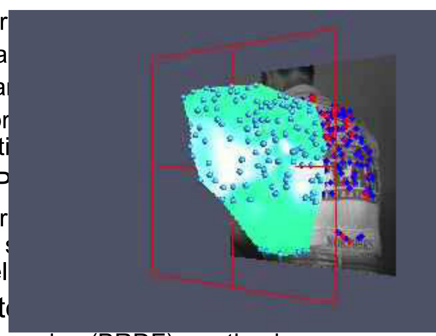
Fernando De la Torre

Carnegie Mellon
THE ROBOTICS INSTITUTE

Outline

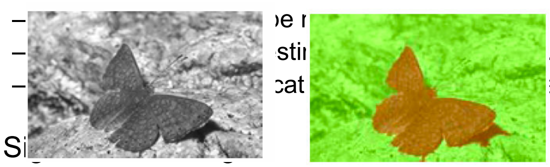
Component Analysis

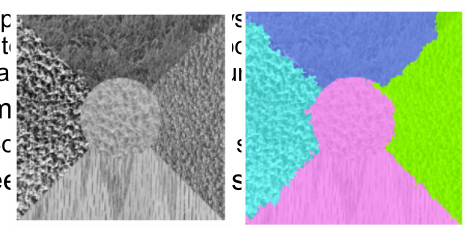
- Computer Vision & Image Processing
  - Structure from motion.
  - Spectr
  - Appea
  - Funda n.
  - Compr ensionality reducti
- Signal P
  - Spectr . Kalman filter), roblem, eco cancel
- Comput
  - Compression (BRDF), synthesis,…
- Speech, bioinformatics, combinatorial problems.



Component Analysis

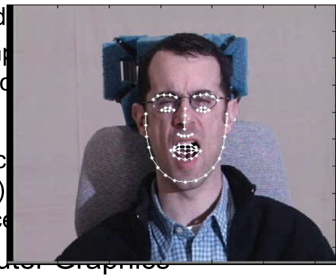
- Computer Vision & Image Processing
  - Structure from motion.
  - Spectral graph methods for segmentation.
  -)e r
  - sti .
  - cat ensionality
- Si
  - Sp .g. Kalman filt problem, eco ca
- Com
  - Cc
- Spe roblems.





Component Analysis

- Computer Vision & Image Processing
  - Structure from motion.
  - Spectral graph methods for segmentation.
  - Appearance and shape models.
  - Fund calibration.
  - Comp zation, dimensionality reduc
- Signal
  - Spec cation (e.g. Kalman filter) . cocktail problem, eco canc on, …
- Compu
  - Compression (BRDF), synthesis,…
- Speech, bioinformatics, combinatorial problems.



Why Component Analysis?

- Learn from very high dimensional data and few samples.
  - Useful for dimensionality reduction.
- Easy to incorporate
  - Robustness to noise, missing data, outliers (Torre & Black '03)
  - Invariance to geometric transformations (Torre et al. '07, Torre & Black '02)
  - Non-linear (kernel methods) (Vapnik '95, Scholkopf '99,Cristiani & Taylor '04)
  - Probabilistic (latent variable models) (Everitt '84, Bishop et al 99, Benter '80)
  - Multi-factorial (tensors) (Ratiu '88, Vasilescu & Terzopolous 02, Avidan & Shashua '97)
  - Exponential family PCA (Gordon '02)
- Efficient methods O($\underbrace{d}_{\text{features}}$ $\underbrace{n< <n^2}_{\text{samples}}$)

Are CA methods popular/useful/used?

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

Generative Models

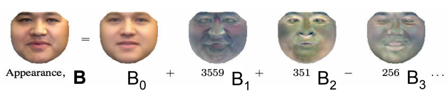
$$D \approx BC$$

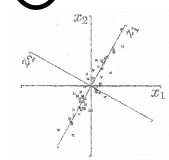
Principal Component Analysis (CA)

- SVD (Beltrami, 1873;Jordan 1874). PCA (Pearson, 1901; Hotelling, 1933)

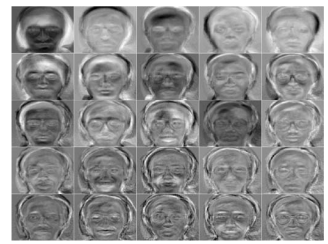
d= pixels $\mathbf{D} = [\mathbf{d}_1 \; \mathbf{d}_2 \; \dots \; \mathbf{d}_n]$ n= images

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



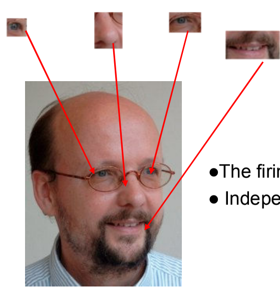


"Intercorrelations among variables are the bane of the multivariate researcher's struggle for meaning"



Cooley and Lohnes, 1971

Part-based representation



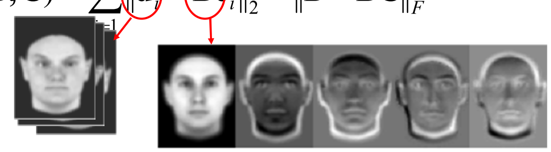
- The firing rates of neurons are never negative.
- Independent representations.

NMF & ICA

Optimization for PCA

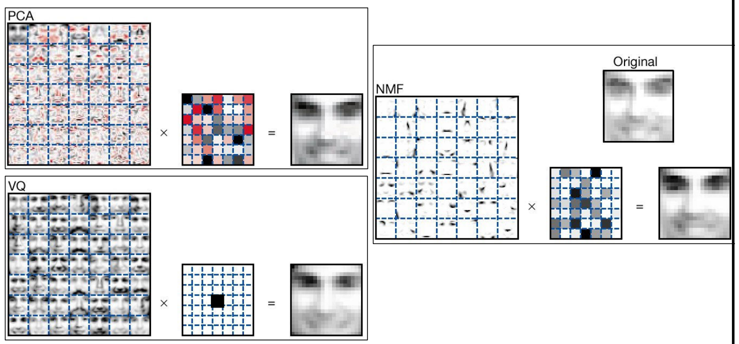
- PCA minimizes the following **CONVEX** function.

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



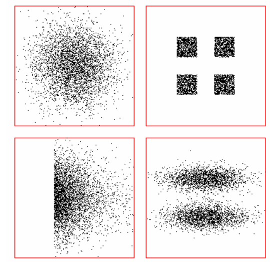
Non-negative Matrix Factorization

- Positive factorization.
  $E(\mathbf{B},\mathbf{C}) = \| \mathbf{D} - \mathbf{BC} \|_F \quad \mathbf{B},\mathbf{C} \geq 0$
- Leads to part-based representation.



Independent Component Analysis

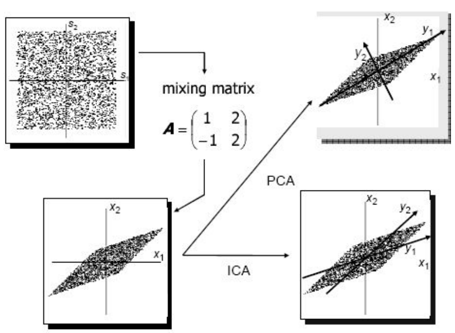
- We need more than second order statistics to represent the signal.



ICA vs PCA

(Herault and Jutten 83; Olhausen & Field, 1996,Hyvrinen et al., 2001)

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



Many optimization criteria

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

$$\sum_{i=1}^{n} \|\mathbf{d}_i - \mathbf{Bc}_i\| + \sum_{i=1}^{n} S(\mathbf{c}_i)$$ Sparseness (e.g. S=| |)

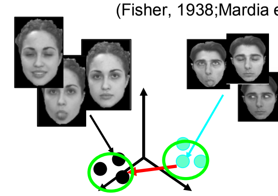
- Other sparse PCA.
  (Chennubhotla & Jepson, 2001b; Zou et al., 2005; dAspremont et al., 2004;)

Discriminative Models

- Linear Discriminant Analysis (LDA)
- Canonical Correlation Analysis (CCA)

Linear Discriminant Analysis (LDA)

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



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

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

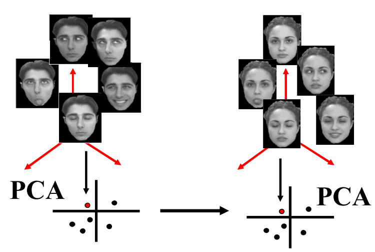
$$J(\mathbf{B}) = \frac{|\mathbf{B}^T\mathbf{S_b}\mathbf{B}|}{|\mathbf{B}^T\mathbf{S}_t\mathbf{B}|} \qquad \mathbf{S_b}\mathbf{B} = \mathbf{S}_t\mathbf{B}\boldsymbol{\Lambda}$$

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

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

Canonical Correlation Analysis

- PCA independently and general mapping



- Signals dependent signals with small energy can be lost.

Canonical Correlation Analysis (CCA)

(Mardia et al., 1979; Borga)

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

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

- Several ways of optimizing it:

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

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

$$\mathbf{A}\mathbf{w} = \lambda\mathbf{B}\mathbf{w}$$

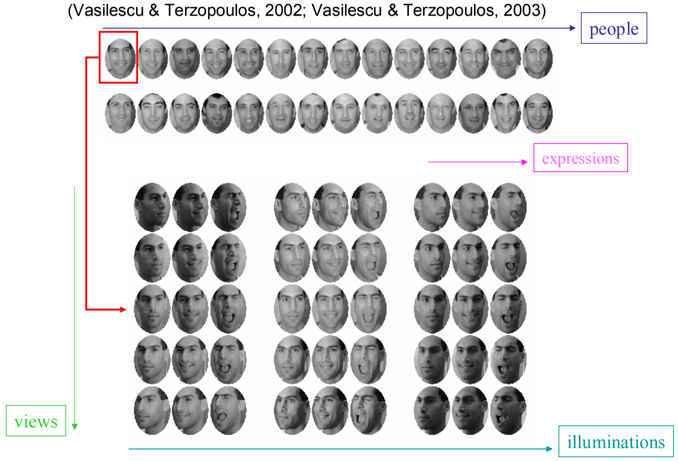
Standard extensions

- Tensor Factorization.
- Latent Variable Models
- Kernel Methods

Tensor Factorization

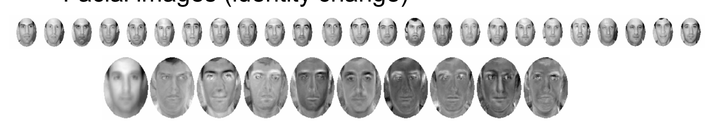
Tensor faces

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



Eigenfaces

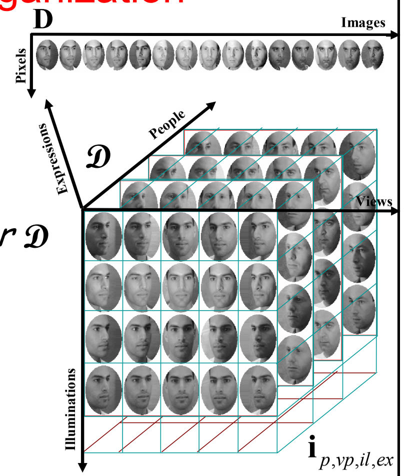
- Facial images (identity change)



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

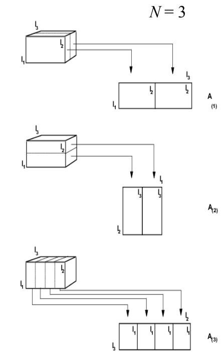
Data Organization

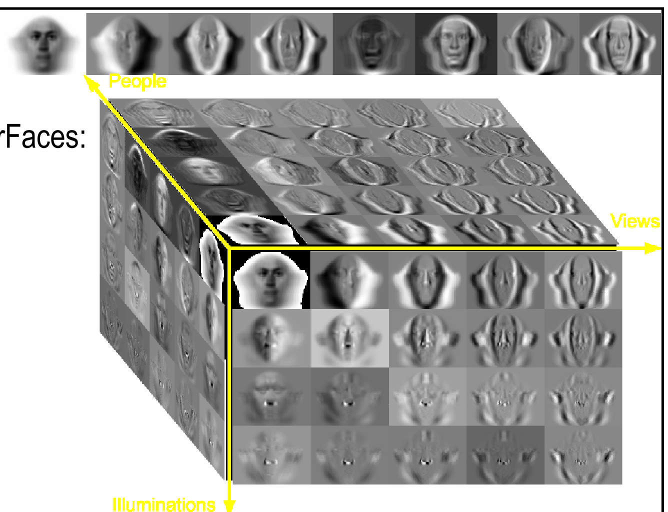
- Linear/PCA: **Data Matrix**
  - $R^{\text{pixels x images}}$
  - a matrix of image vectors
- Multilinear: *Data Tensor* $\mathcal{D}$
  - $R^{\text{people x views x illums x express x pixels}}$
  - N-dimensional matrix
  - 28 people, 45 images/person
  - 5 views, 3 illuminations, 3 expressions per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



N-Mode SVD Algorithm

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





Strategic Data Compression = Perceptual Quality

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

Original	TensorFaces	TensorFaces	PCA
		Mean Sq. Err. = **409.15**	Mean Sq. Err. = **85.75**
	6 illum + 11 people param.	3 illum + 11 people param.	33 parameters
176 basis vectors	66 basis vectors	33 basis vectors	33 basis vectors

Latent Variable Models

Factor Analysis

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

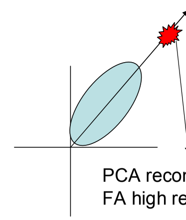
$$\mathbf{d} = \boldsymbol{\mu} + \mathbf{B}\mathbf{c} + \boldsymbol{\eta}$$

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

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

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

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



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

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

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

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

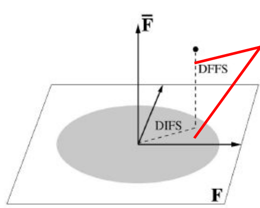
PCA reconstruction low error.
FA high reconstruction error (low likelihood).

Ppca

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

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

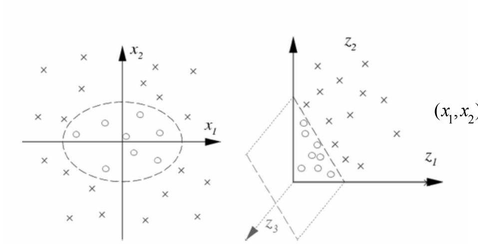
$$\mathbf{c}_i = \mathbf{B}^T\mathbf{d}_i$$



Other non-linear

- Mixture of Factor Analyzers
- Generative topographic mapping

Kernel methods

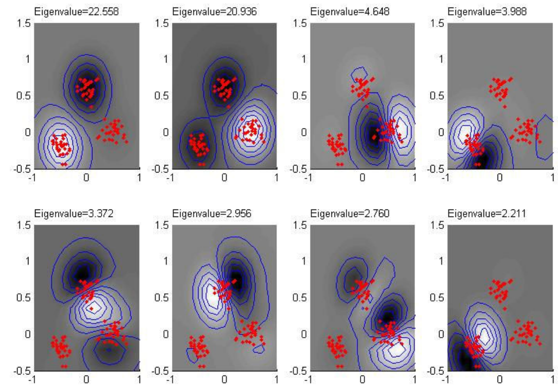


$$(x_1, x_2) \rightarrow (x_1, x_2, x_1^2 + x_2^2) = (z_1, z_2, z_3)$$

- Non-linear high dimensional mapping of the data so it becomes linearly separable in the feature space
- This is equivalent to finding a non-linear decision boundary in the input space
- Computation in the feature space can be costly because it is high dimensional
  - The feature space is typically infinite-dimensional!

Examples

$$\langle \mathbf{b}, \Phi(\mathbf{d}) \rangle = \sum_{i=1}^{n} \alpha_i K(\mathbf{d}_i, \mathbf{d})$$



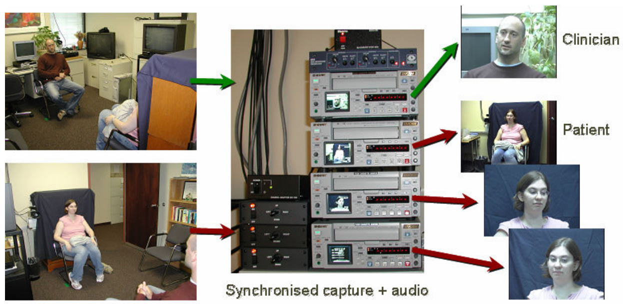
Second Part

Clinical Depression

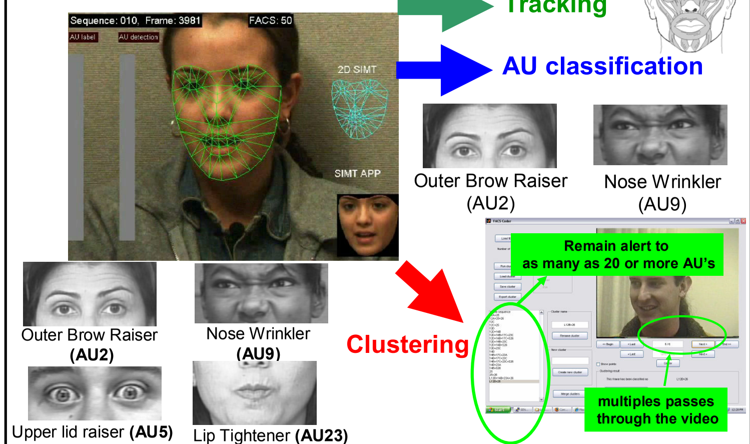
- Point Prevalence
  - 10% of the population
- High personal and social costs
- Current standards for diagnosis or therapeutic effectiveness of treatment relies on:
  - Interview with the physician.
  - Self-report measures.

1. **DEPRESSED MOOD** (Sadness, hopeless, helpless, worthless)
   - **0=** Absent
   - **1=** These feeling states indicated only on questioning
   - **2=** These feeling states spontaneously reported verbally
   - **3=** Communicates feeling states non-verbally—i.e., through facial expression, posture, voice, and tendency to weep
   - **4=** Patient reports VIRTUALLY ONLY these feeling states in his spontaneous verbal and non-verbal communication

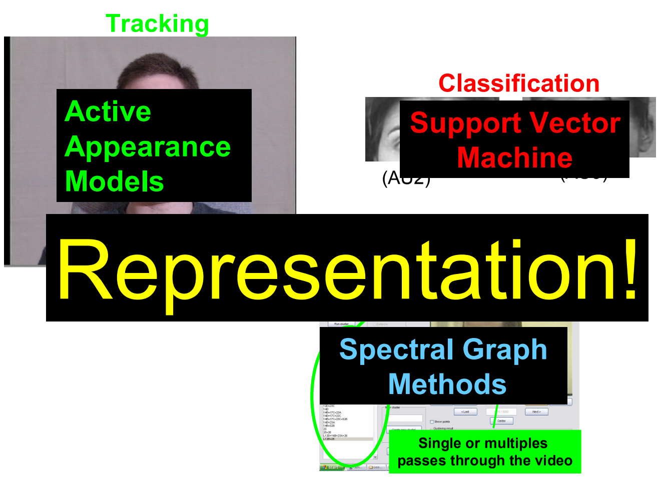
Behavioral Analysis from Audio and Video



Automatic FACS coding

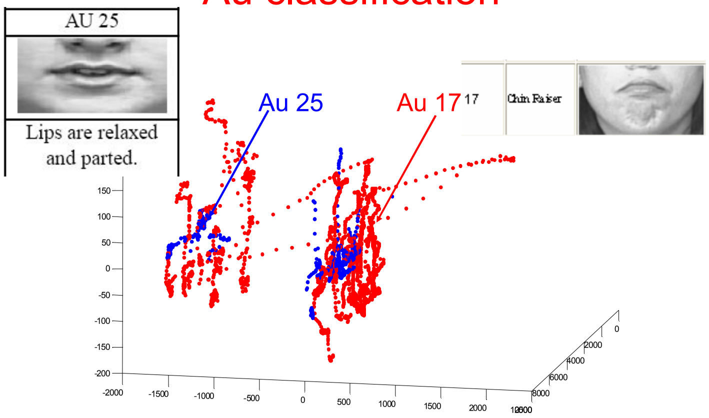


Problems to Solve



Classification

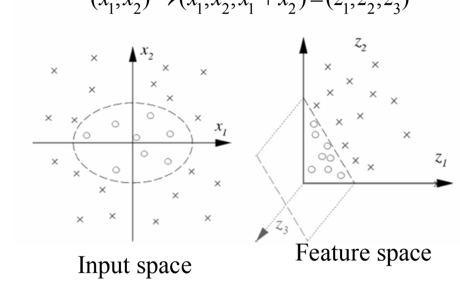
Au classification



Kernels

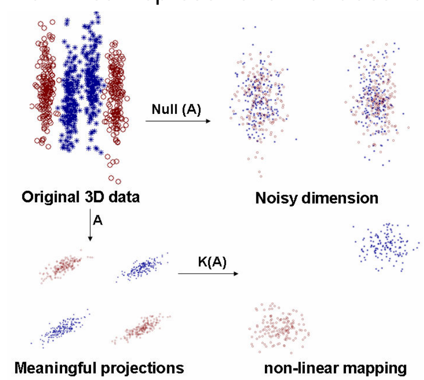
- Support Vector Machines
  - Large margin (QP- no local minima).
  - **Kernels.**

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



Problem

- Learning a non-linear representation for classification



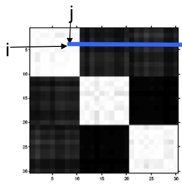
Previous work

- Neural networks
  - Hard to train, local minima, difficult interpretation of hidden units.
- Metric learning (Xing et al. '02, Goldberger et al. '05 , He et al. '03, Weinberger '06, Lebanon '03, Shental '02)
- Kernels (Cristiani et al '01, Lanchkriet et al. '04, Chapelle et al. '02, Kwok et al. '03)

- Learning kernel expansions
- Kernel with dimensionality reduction
- Computational cost 0(dkn²) samples features

Learning Kernel Expansions

$$\mathbf{K}(\mathbf{B}_1,\ldots,\mathbf{B}_p,\lambda_1,\ldots,\lambda_p,\boldsymbol{\alpha}) = \sum_{t=0}^{p} \alpha_t \hat{\mathbf{M}}_t^{\circ t} \quad \alpha_t \succ 0$$



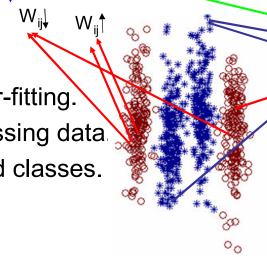
$$k_{ij} = \sum_{t=0}^{p} \alpha_t (\hat{m}_{tij})^t \quad \alpha_t \succ 0$$

$$\hat{m}_{tij}^{\;t} = \left[\frac{\mathbf{d}_i^T \,(\mathbf{B}_t\mathbf{B}_t^T\mathbf{A} + \lambda_t \mathbf{I})\,\mathbf{d}_j}{\sqrt{\mathbf{d}_i^T(\mathbf{B}_t\mathbf{B}_t^T\mathbf{A} + \lambda_t\mathbf{I})\mathbf{d}_i\,\mathbf{d}_j^T(\mathbf{B}_t\mathbf{B}_t^T\mathbf{A} + \lambda_t\mathbf{I})\,\mathbf{d}_j}} \right]^t$$

Learning from an Ideal Kernel

$$\sum_{t=0}^{p} \alpha_t \hat{\mathbf{T}}_t^{\circ t} \quad \alpha_t \succ 0$$

$$E = \| \mathbf{W} \circ (\mathbf{F} - \mathbf{K}(\mathbf{B}_1,\ldots,\mathbf{B}_p,\lambda_1,\ldots,\lambda_p,\boldsymbol{\alpha})) \|_F$$



- Avoids over-fitting.
- Outliers/missing data.
- Unbalanced classes.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Ideal kernel

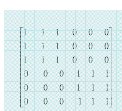
- First order (t=1) $E = \| \mathbf{F} - \mathbf{D}^T\mathbf{B}\mathbf{B}^T\mathbf{D} \|_F$ ➡ **LDA**

Experiments

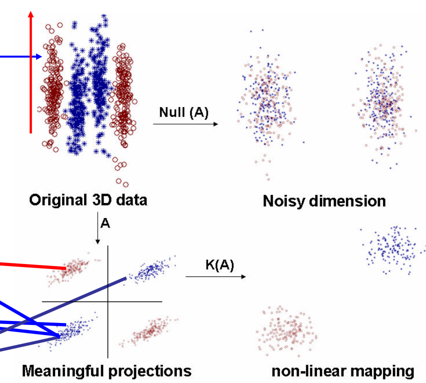
$\mathbf{A} \in \Re^{3\times 3}$

$\lambda_1 = 1.98 \quad \lambda_2 = 0.68 \quad \lambda_3 = 0.00$

$\alpha_1 = 0 \quad \alpha_2 = 0.2 \quad \alpha_3 = 0 \quad \alpha_4 = 0.1$



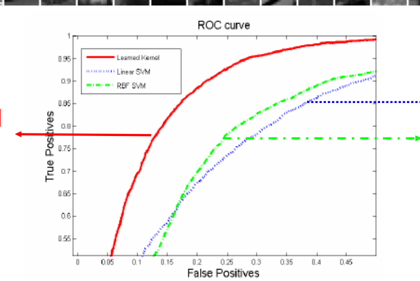
$$\begin{bmatrix} 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}$$



Pedestrian Detection

- **Training** (20,000) **Cross validation** (10,000) **Testing** (20,000)

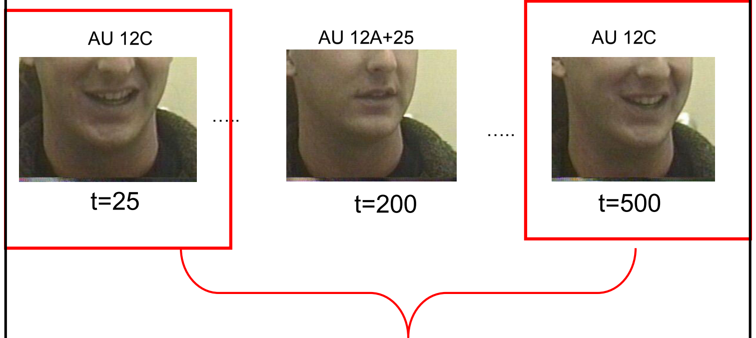
(Munder and Gavrila 06)



Clustering

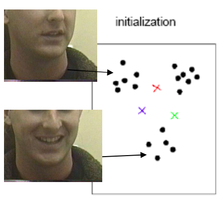
Clustering AUs

- How to cluster invariantly to geometric transformations to capture subtle facial behavior?



The Clustering Problem

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



- Number of possible partitions

$$S(n,c)=\frac{1}{c}\sum_{i=1}^{c}(-1)^{c}\binom{c}{i}i^{n} \qquad \begin{cases} n=21 \\ c=4 \end{cases} \qquad \approx 10^{12}$$

- NP-hard and approximate algorithms (k-means, hierarchical clustering, mog, …)

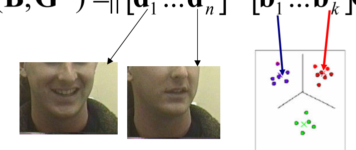
K-means

(MacQueen 67)

- K-means criterion:

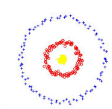
$$g_{ij} \in \{0,1\}$$
$$\mathbf{G}\mathbf{1}_c = \mathbf{1}_n$$

$$E(\mathbf{B},\mathbf{G}^T) = \| \underbrace{[\mathbf{d}_1 \ldots \mathbf{d}_n]}_{\mathbf{D}} - \underbrace{[\mathbf{b}_1 \ldots \mathbf{b}_k]}_{\mathbf{B}} \mathbf{G}^T \|_F$$



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

- Problems:
  - Local minima and sensitive to initial conditions
  - Optimal for spherical clusters

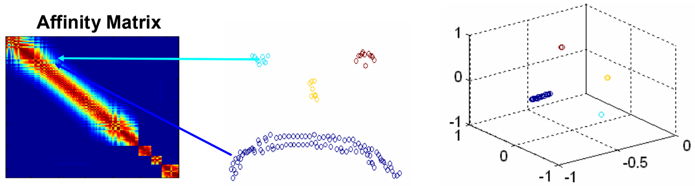


A Unifying View of Clustering

- K-means (Gu et al. 01') $E(\mathbf{G},\mathbf{B}) = \| \mathbf{D} - \mathbf{B}\mathbf{G}^T \|_F$
- Energy-based Spectral Clustering (Dhillon et al., '04, Torre et al '06).

$$E(\mathbf{G},\mathbf{B}) = \| (\boldsymbol{\Gamma} - \mathbf{B}\mathbf{G}^T)\mathbf{W} \|_F \rightarrow$$ Normalized Cuts (Shi & Malik '00) Ratio-cuts (Hagen & Kahng '02)

$$\boldsymbol{\Gamma} = [\varphi(\mathbf{d}_1)\ \varphi(\mathbf{d}_2)\ \ldots\ \varphi(\mathbf{d}_n)] \quad \boldsymbol{\Gamma} = \varphi(\mathbf{D})$$

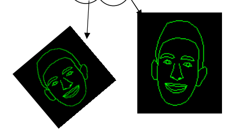


Parameterized Cluster Analysis

$$E_1(\mathbf{G},\mathbf{B}) = \| \boldsymbol{\Gamma} - \mathbf{B}\mathbf{G}^T \|_F \quad \boldsymbol{\Gamma} = [\varphi(\mathbf{d}_1)\ \varphi(\mathbf{d}_2)\ \ldots\ \varphi(\mathbf{d}_n)]$$

- Parameterize the shape.

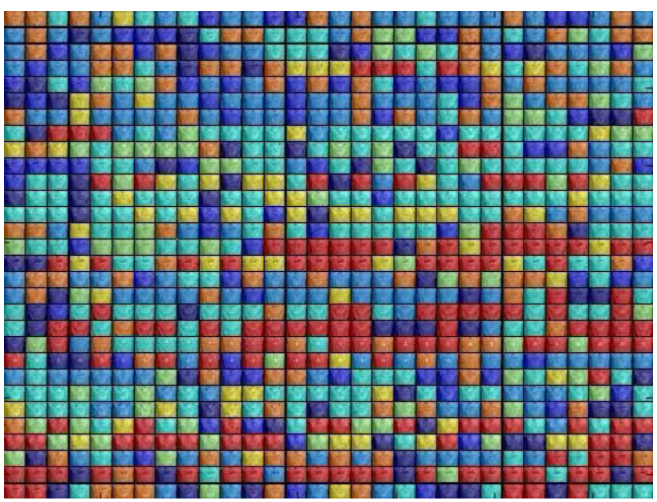
$$\boldsymbol{\Gamma} = [\varphi(\mathbf{A}_1\mathbf{d}_1)\ \varphi(\mathbf{A}_2\mathbf{d}_2)\ \ldots\ \varphi(\mathbf{A}_n\mathbf{d}_n)]$$



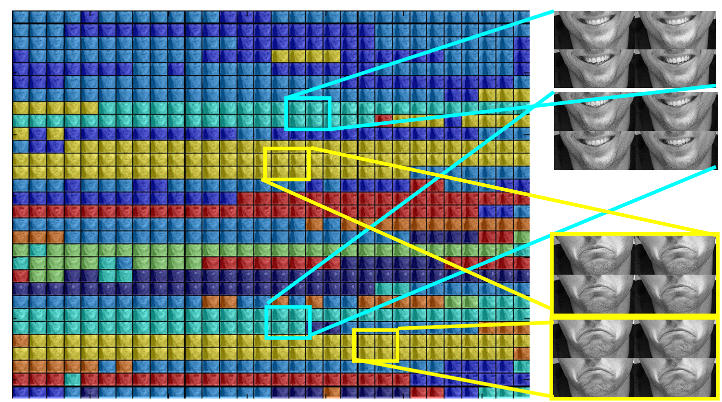
parameterized kernel

$$E_2(\mathbf{A},\mathbf{G}) = tr(\boldsymbol{\Gamma}(\mathbf{A})^T\boldsymbol{\Gamma}(\mathbf{A})(\mathbf{I} - \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T))$$

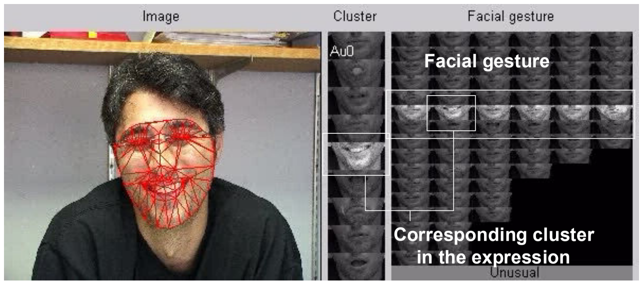
$$\mathbf{G} \geq \mathbf{0} \quad \mathbf{G}\mathbf{1_c} = \mathbf{1_n}$$



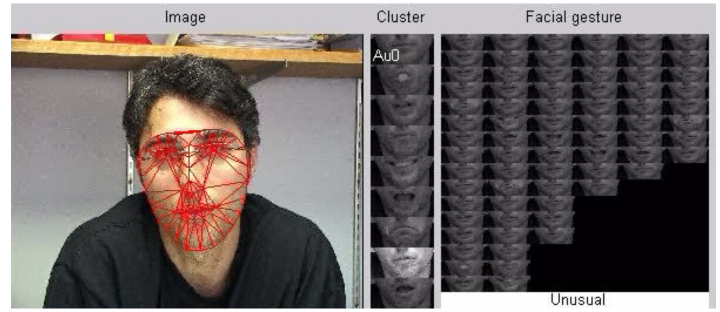
Example



Experiments



Experiments



An Energy-based Framework for Component Analysis

Modeling

- Filtered CA [CVPR-Torre et al. 07]
- Robust PCA [IJCV-Torre & Black 03]
- Convolutional CA [Torre et al. In prep.]
- Parameterized CA [IVC-Torre & Black 03]
- Unsupervised bilinear [ICCV- Gross et al. 07]

$$\left\| \begin{bmatrix} \mathbf{D} \otimes \mathbf{f}_1 \\ \dots \\ \mathbf{D} \otimes \mathbf{f}_k \end{bmatrix} \right\|_F$$

Classification

- Learning kernel [CVPR-Torre &Vinyals 07]
- Multimodal Oriented DA [ICML-Torre et al. 05]
- Representational OCA [CVPR-Torre et al. 05]

$$E(\mathbf{B}, \boldsymbol{\alpha}) = \left\| \sum_{i=1}^{n} \alpha_i \left[\frac{\mathbf{D}^T \mathbf{B} \mathbf{B}^T \mathbf{D}}{normalization} \right]^{l} - \mathbf{F} \right\|_F$$

$$E(\mathbf{B}, \mathbf{C}) = \left\| \mathbf{D} - \mathbf{BC} \right\|_F$$

$$\left\| (\varphi(\mathbf{D}(\mathbf{f}(\mathbf{x}, \mathbf{a}))) - \mathbf{BG})\mathbf{W} \right\|_F$$

Clustering

- Parameterized CA [CVPR-Torre et al. 07]
- Discriminative CA [ICML-Torre & Kanade 06]

Visualization

- DCCA [CVPR- Torre & Black 02]
- Dynamic MDS [ICML-Cabero et al. 07]

Thanks

1st Workshop on **Component Analysis**

http://www.cs.cmu.edu/~ftorre/ca/