Classifying Facial Actions

by G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski

IEEE Trans. On Pattern Analysis and Machine Intelligence,
Vol. 21, No. 10, October 1999

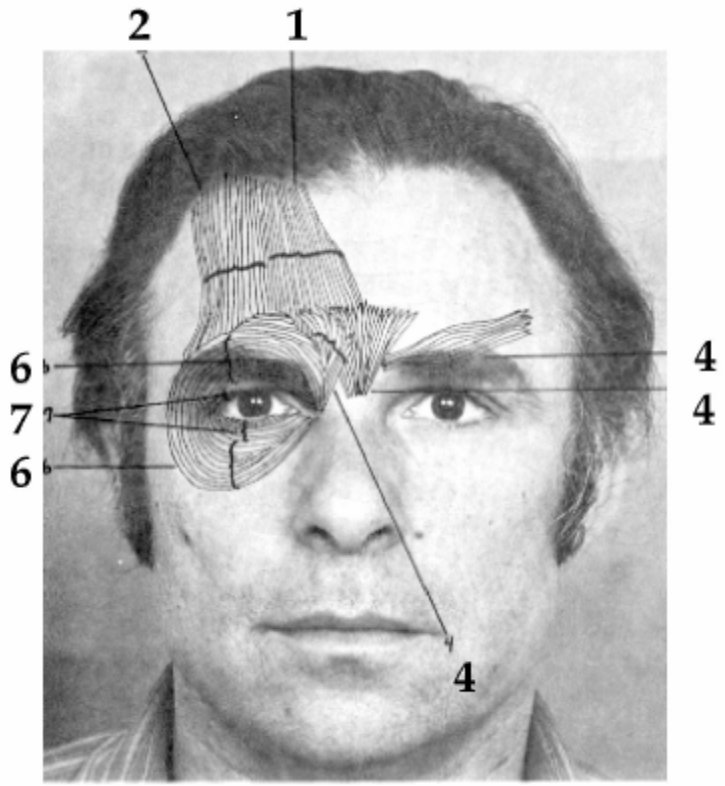
Heeyoul (Henry) Choi
Dept. of Computer Science
Texas A&M University
hchoi@cs.tamu.edu

Outline

- Facial Action Coding System
- Overview
- Optic Flow Analysis
  - Local Velocity Extraction
  - Local Smoothing
- Holistic Analysis
  - PCA, LFA, FLD, ICA
- Local Analysis
  - Local PCA, Gabor wavelets, PCA Jets
- Discussion

Facial Action Coding Systems (FACS)

- Objective description of facial signals in terms of component motions or “facial actions”



Facial Action Coding Systems (FACS)

- Developed by Ekman and Friesne in 1978
  - For behavioral science investigations of the face
  - 46 Action Units (AUs) : independent motions
- Manual FACS: a major impediment is the time.
  - to train human experts
  - to score the video tape

Facial Action Coding Systems (FACS)

- It provides
  - a tool for basic science research
  - Strong basis for human-computer interaction systems
  - More detailed information about facial behavior

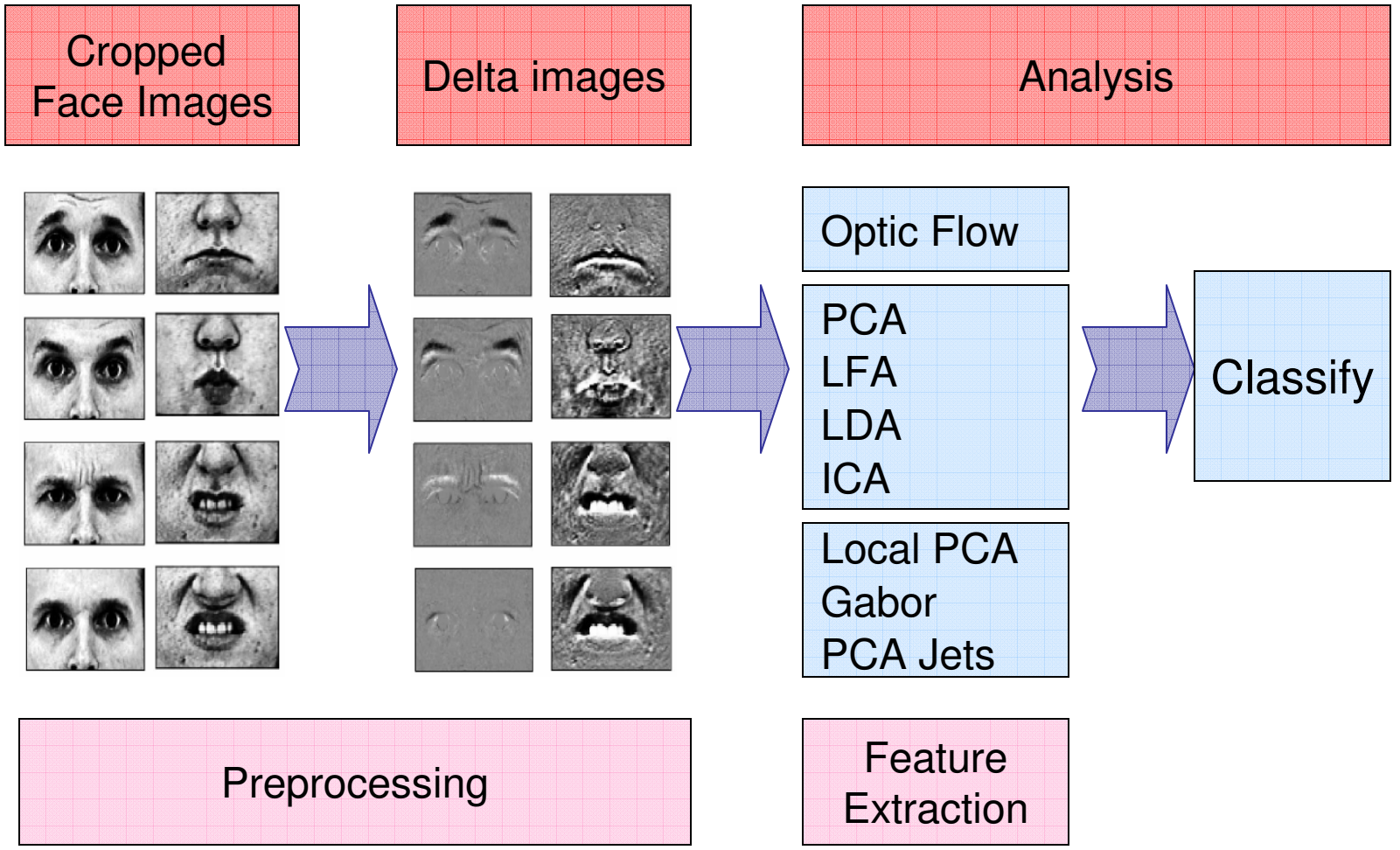
- So, we need an automated FACS.

Facial Action Coding Systems (FACS)

- Previous methods involved tracking the positions of dots attached to the face.
  - Detecting facial actions without dots to the subjects face have broader utility.

- Full-field representation of image textures and image motion
  - More reliable than task-specific feature such as the increase of facial wrinkles.

Overview

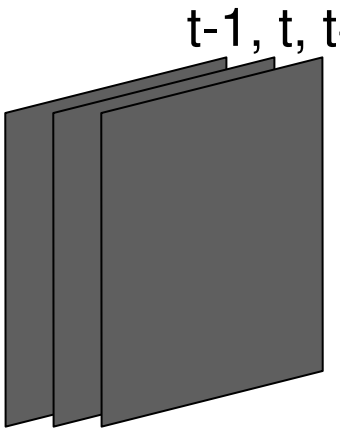

Cropped Face Images
Delta images
Analysis
Optic Flow
PCA
LFA
LDA
ICA
Local PCA
Gabor
PCA Jets
Classify
Preprocessing
Feature Extraction

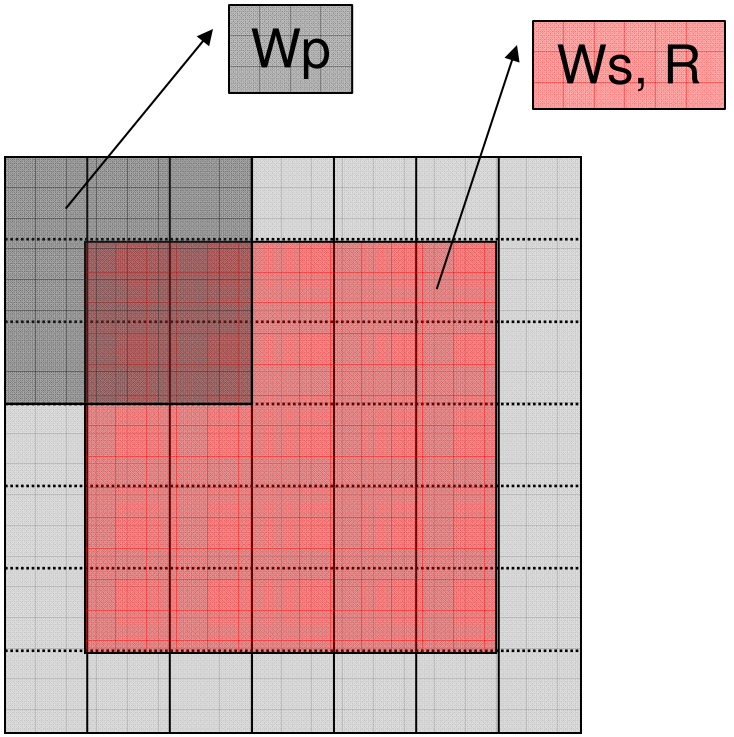
Database

- 1100 sequences, 150 actions, 24 subjects
- Each sequence contains 6 images.
- No head motion
- Alignment with manual mouse click.
  - Rotation and Scale

Optic Flow Analysis

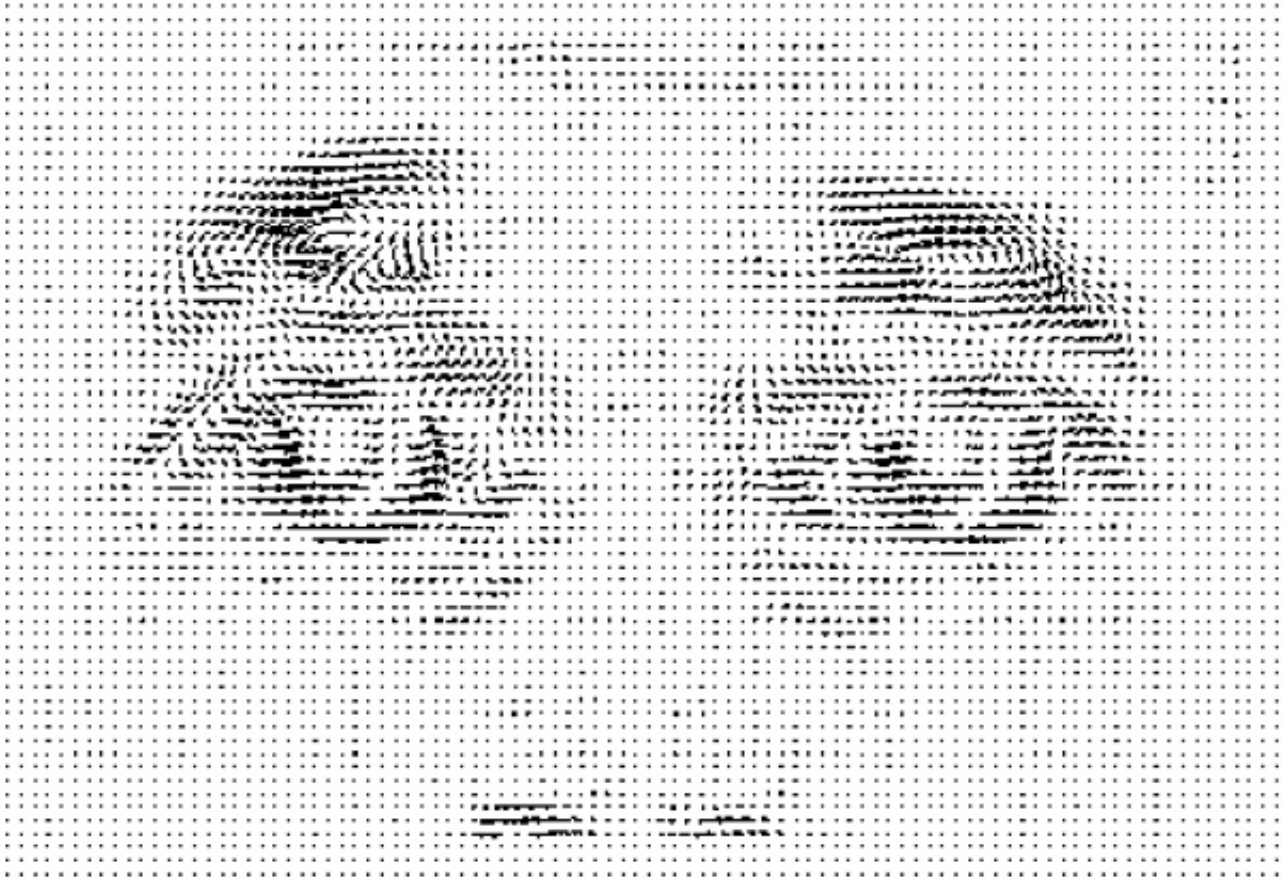
- Local Velocity Extraction





$$\hat{u} = \frac{\sum_u \sum_v \mathcal{R}(u,v)u}{\sum_u \sum_v \mathcal{R}(u,v)} \qquad \hat{v} = \frac{\sum_u \sum_v \mathcal{R}(u,v)v}{\sum_u \sum_v \mathcal{R}(u,v)} \qquad u, v \in [-2, 2]$$

Optic Flow Analysis



Optic Flow Analysis

- Local Smoothing
  - Optic Flow is a noise measure.
  - Optimal estimate U is a combination of U_{cc} and U_{mean}
  - Dist(U, U_{cc})
    - the error due to conservation constraints
  - Dist(U, U_{mean})
    - the error in the smoothness constraint estimate
  - Updating rule

$$\mathcal{U}^0 = \mathcal{U}_{cc}$$

$$\mathcal{U}^{k+1} = \left[\mathcal{S}_{cc}^{-1} + \overline{\mathcal{S}}^{-1}\right]^{-1}\left[\mathcal{S}_{cc}^{-1}\mathcal{U}_{cc} + \overline{\mathcal{S}}^{-1}\overline{\mathcal{U}}^{k}\right]$$

Optic Flow Analysis

- Classification Procedure
  - Each image analysis algorithm produced a feature vector, f.
  - Classifier: Simple nearest neighbor or template matching
  - Similarity: cosine or Euclidean between two feature vectors.
  - Leave-one-out cross-validation

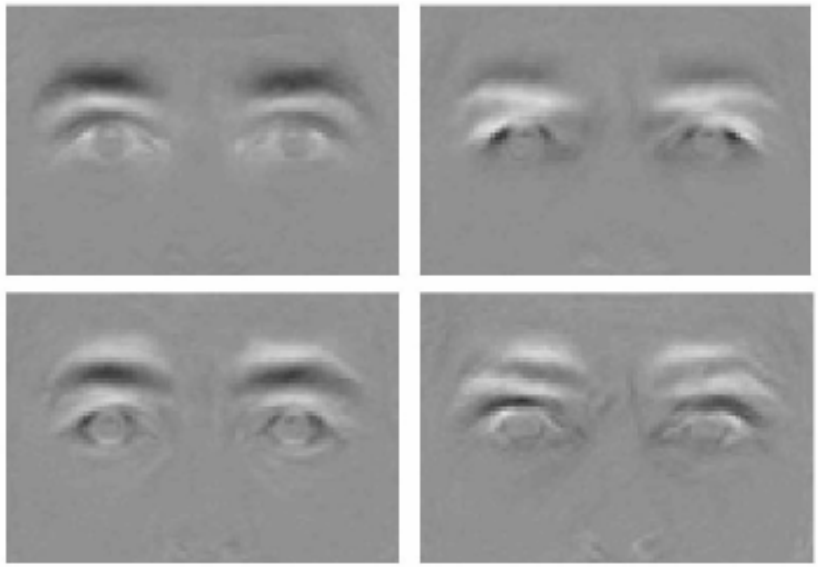
  - Smoothing degraded the performance.
    - High spatial resolution optic flow is important.

Holistic Analysis

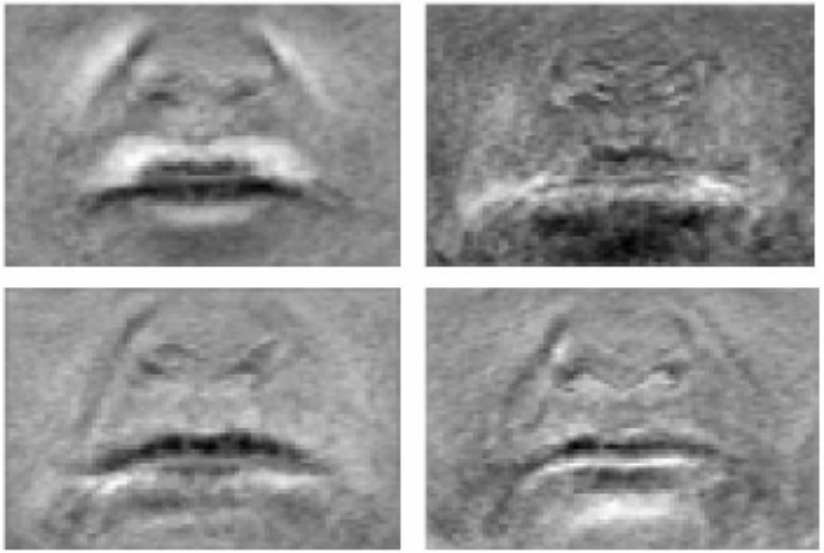
- PCA
- Local Feature Analysis
- LDA
- ICA

PCA

- Eigenactions
- 79.3% with 30 principal components
- Euclidean distance and template matching.



(a) upper face actions



(b) lower face actions.

Local Feature Analysis (LFA)

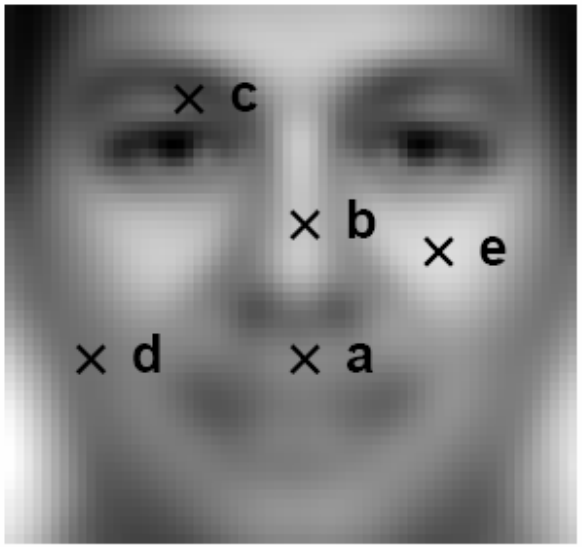
- Output O

$$S = PDP^T$$
$$K = PVP^T \quad \text{where} \quad V = D^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right)$$
$$O = KX^T$$

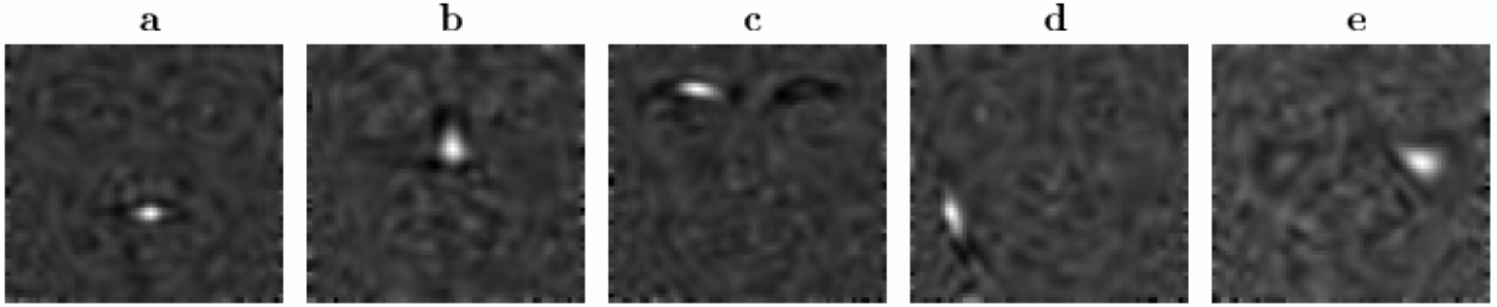
- Adding kernels (filters or basis)

$$\arg\max_x \langle \| O_t(x) - O_t^{rec}(x) \|^2 \rangle$$
$$O^{rec} = Z^T O(\mathcal{M}, \mathcal{N})$$
$$Z = \frac{(O^{rec})^T O(\mathcal{M}, \mathcal{N})}{O(\mathcal{M}, \mathcal{N})^T O(\mathcal{M}, \mathcal{N})}$$

LFA



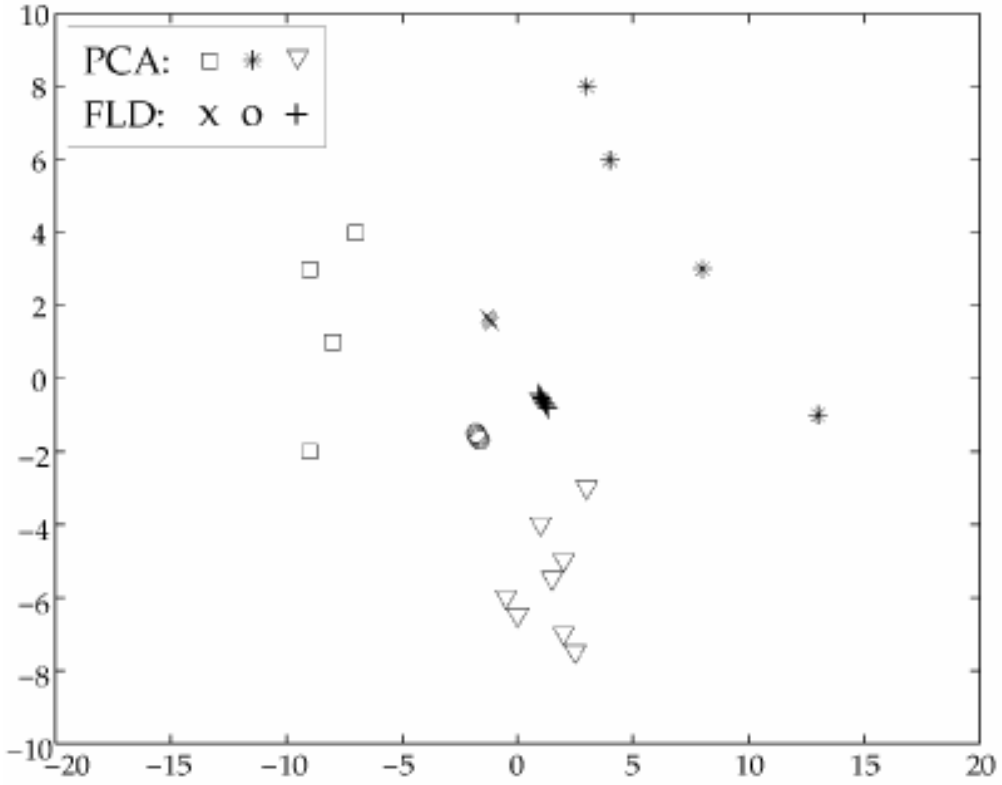
positions



From P.S. Penev and J.J. Atick 1996

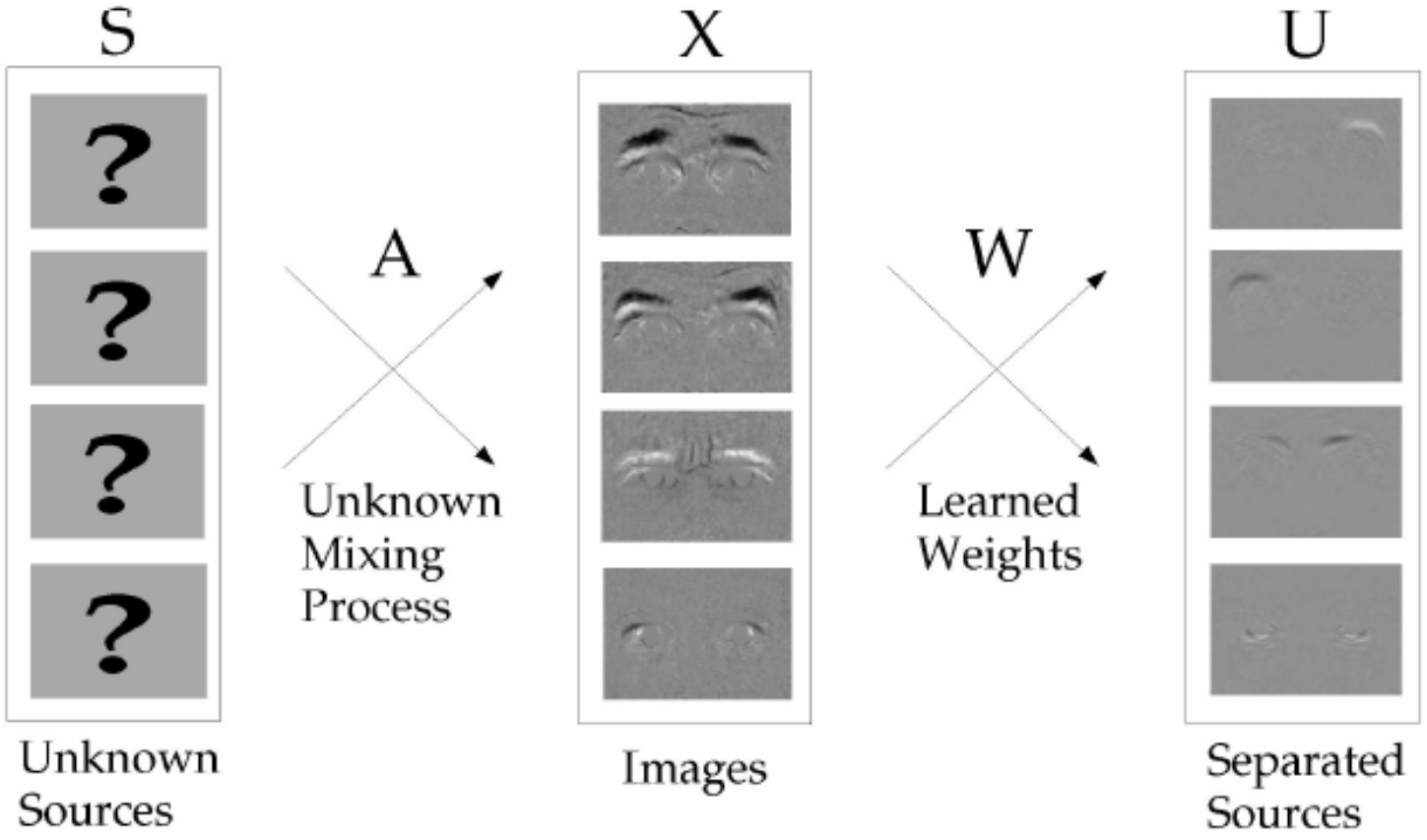
FisherActions (FLD or LDA)

- FLD does not improve classification over PCA.
  - Class information is in the variance.



Independent Component Analysis (ICA)

- High-order statistics, Non-gaussianity.
  - More informative and more sparse



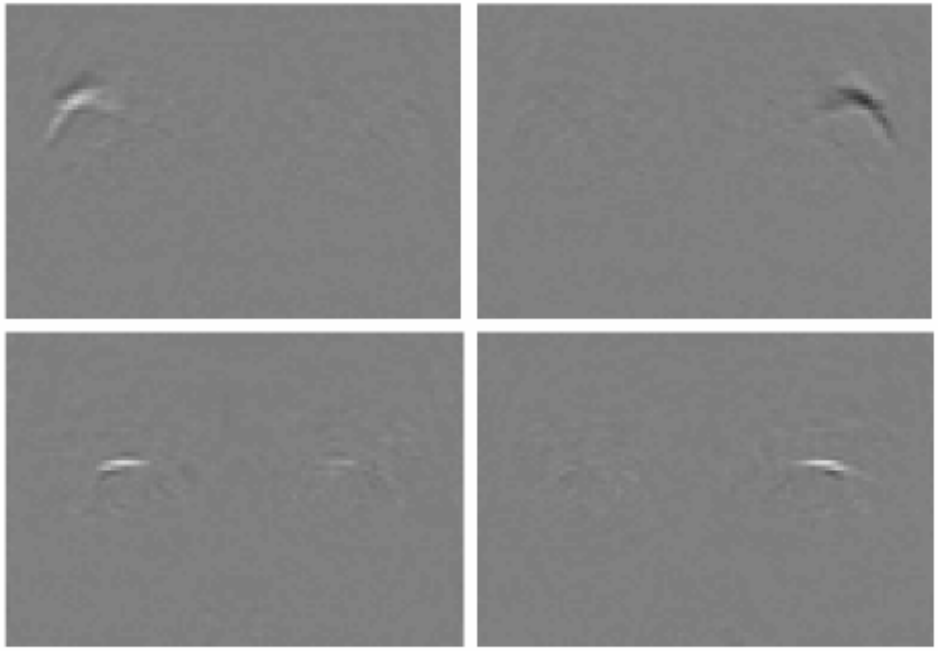
ICA

- Update rule

$$\Delta W = (I + y' u^T)W$$

$$\text{where } y' = \frac{\partial}{\partial y_i}\frac{\partial y_i}{\partial u_i} = \frac{\partial}{\partial u_i} ln \frac{\partial y_i}{\partial u_i} = (1 - 2y_i).$$

- After sphering
  - faster



ICA

- Ambiguities
  - Scale and order

- Ordering based on class discriminability

$$r = \frac{\sigma_{between}}{\sigma_{within}}$$

$$\sigma_{between} = \sum_j (\overline{a}_{jk} - \overline{a}_k)^2$$

$$\sigma_{within} = \sum_j \sum_i (a_{ijk} - \overline{a}_{jk})^2$$

- 95.5% with 75 components
- Best among all of the holistic classifiers
- As in LFA, basis is local.

Local Representations

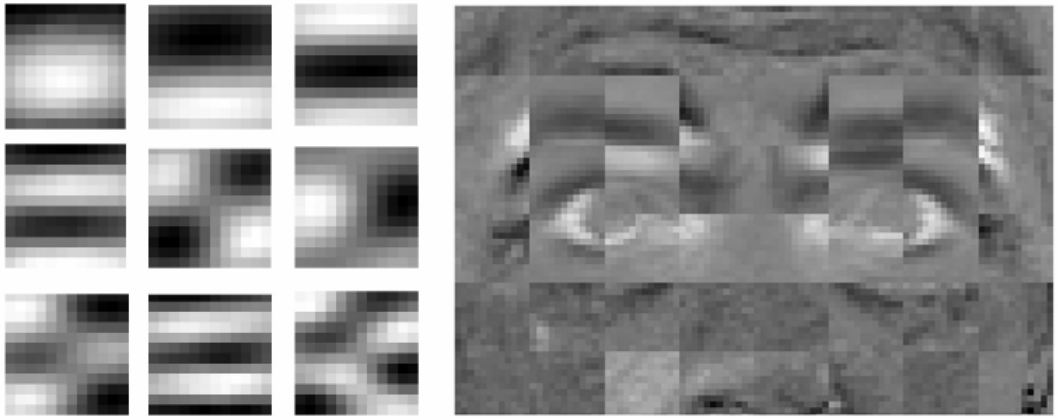
- Local PCA
- Gabor Wavelet
- PCA Jets

Local spatial filters may be superior to global spatial filters for facial expression classification.

Single Gaussian kernel for benchmark.

Local PCA

- Shift-invariant
  - PCA on randomly sampled image patches.
  - Convolution kernels to filter
- Shift-variant
  - Divide images into m fixed regions and PCA on each region.
- Contrary to our expectation, it is not better than PCA.
  - Delta image has no identification information.

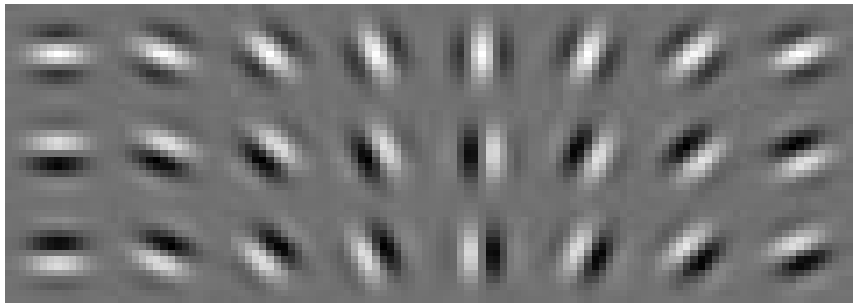


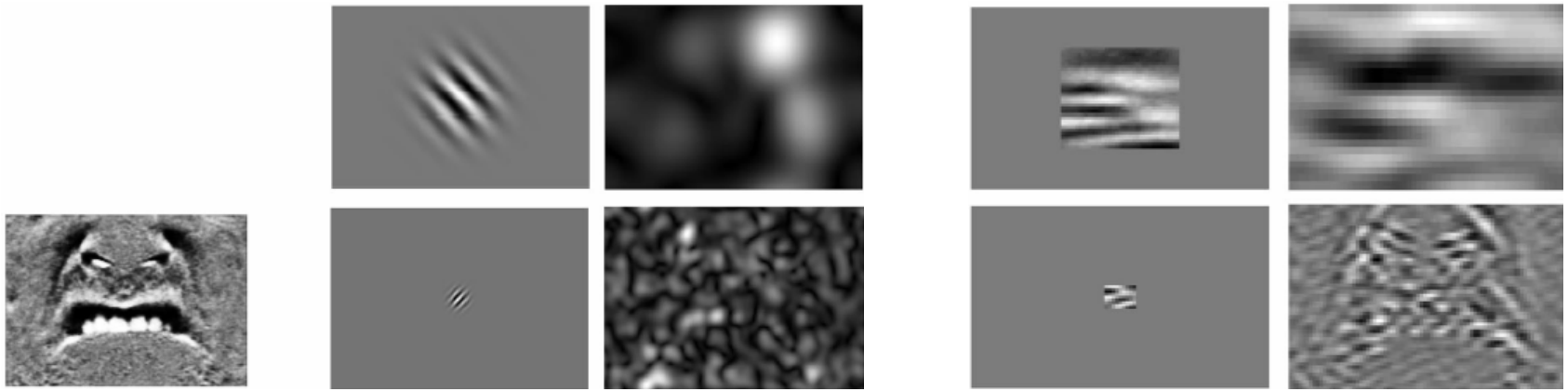
(a) First nine components

(b) Shift-variant local PCA kernels.

Gabor Wavelets

- Predefined local filters. (95.5%)





PCA Jets

- Multiscale version of the local PCA
  - Random image patches at five different scales
  - Principal components at each scale.
- It did not improve performance over the single scale local PCA.
  - The multiscale property of the Gabor wavelets does not account for the improvement.

Discussion

Optic Flow	Correlation	85.6% ± 3.3
	Smoothed	53.1% ± 4.7
Holistic Spatial Analysis	PCA	79.3% ± 3.9
	LFA	81.1% ± 3.7
	FLD	75.7% ± 4.1
	ICA	95.5% ± 2.0
Local Spatial Analysis	Gaussian Kernel	70.3 ± 4.
	PCA Shift-inv	73.4% ± 4.2
	PCA Shift-var	78.3% ± 3.9
	PCA Jets	72.1% ± 4.2
	Gabor Jets	95.5% ± 2.0

Human Subjects	Naive	77.9% ± 2.5
	Expert	94.1% ±2.1

Discussion

- Human subjects
  - Naïve subjects : 77.9%
  - Expert coders : 94.1%
- Best performance
  - Gabor filter and ICA: 96%
  - As well as experts
- Second order statistics
  - As well as naïve human subjects
- High spatial frequencies are important
  - In Optic flow, smoothing does not work.
  - In ICA and Gabor, sparse coding needs high frequencies.
  - Facial Action vs Expression.
    - Actions are more detailed (need high frequencies)
- Motion information + spatial texture information?

Discussion (HC)

- Facial actions are independent and sparse factors.
- ICA could be used locally like local PCA.
- Nonnegative Matrix Factorization (NMF) is a sparser coding method than ICA.
- There is a LFA method with class information.
  - More compact than LFA.
- With original image instead of delta image?
- However, I guess, human does not use the action coding to recognize expressions.