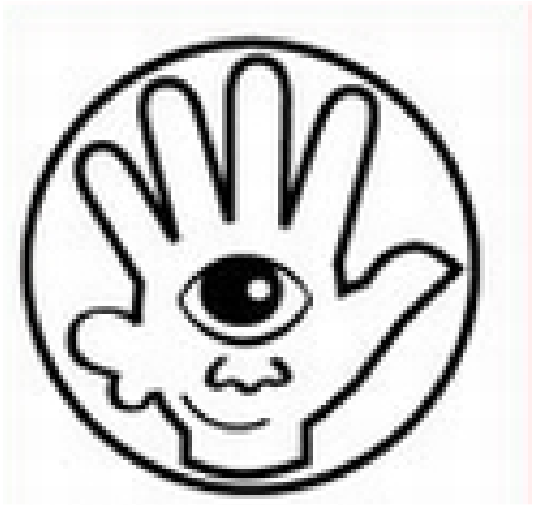
Human Sensing



Fernando De la Torre

Carnegie Mellon
THE ROBOTICS INSTITUTE

Human Sensing Laboratory

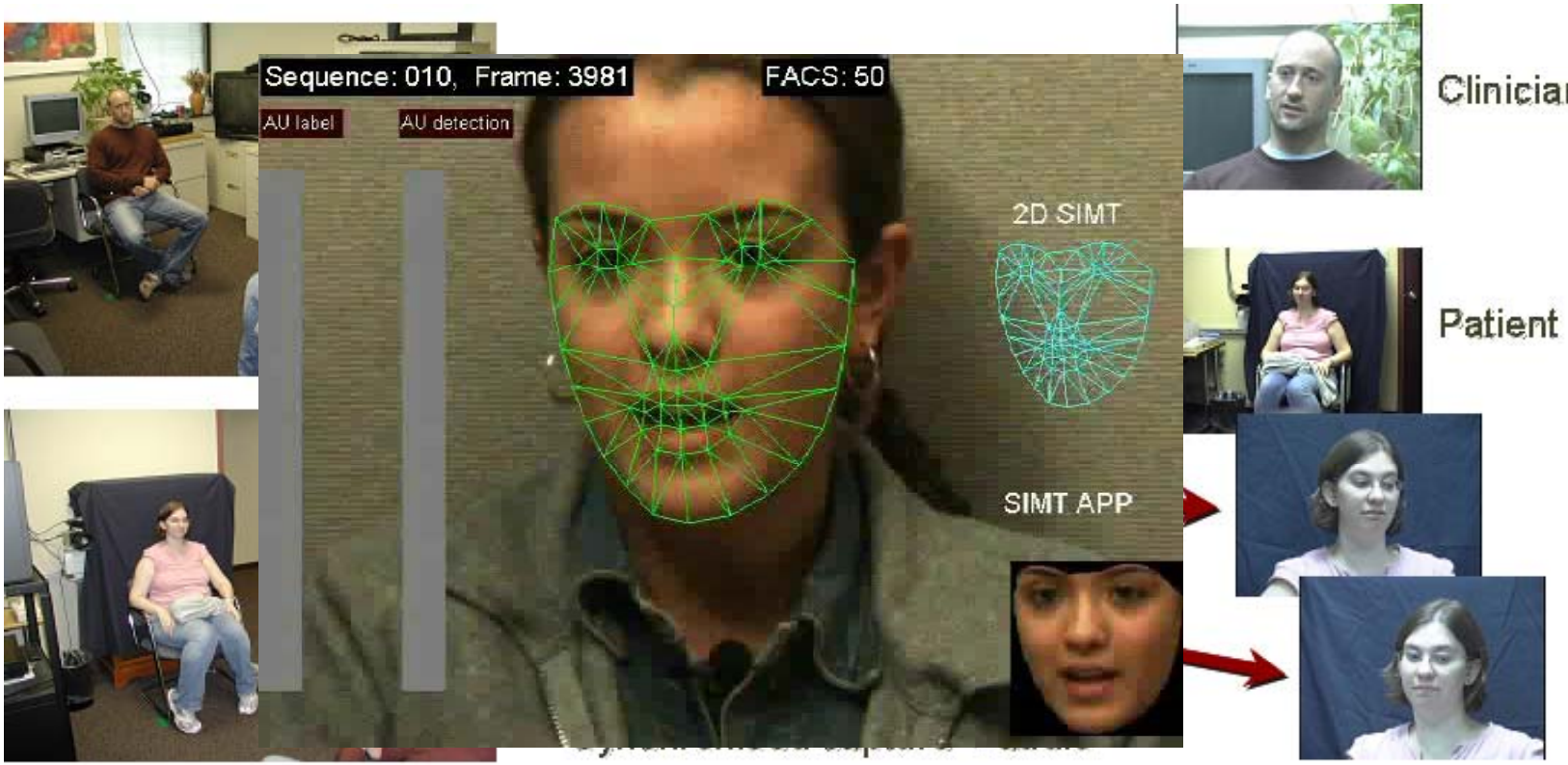
A dream

Outline

- Introduction
- Facial image analysis (FIA)
  - Applications
  - Facial feature detection and tracking
  - Supervised facial expression recognition
  - Unsupervised discovery of facial events
- Conclusions

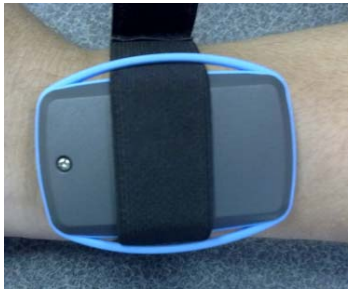
Human health

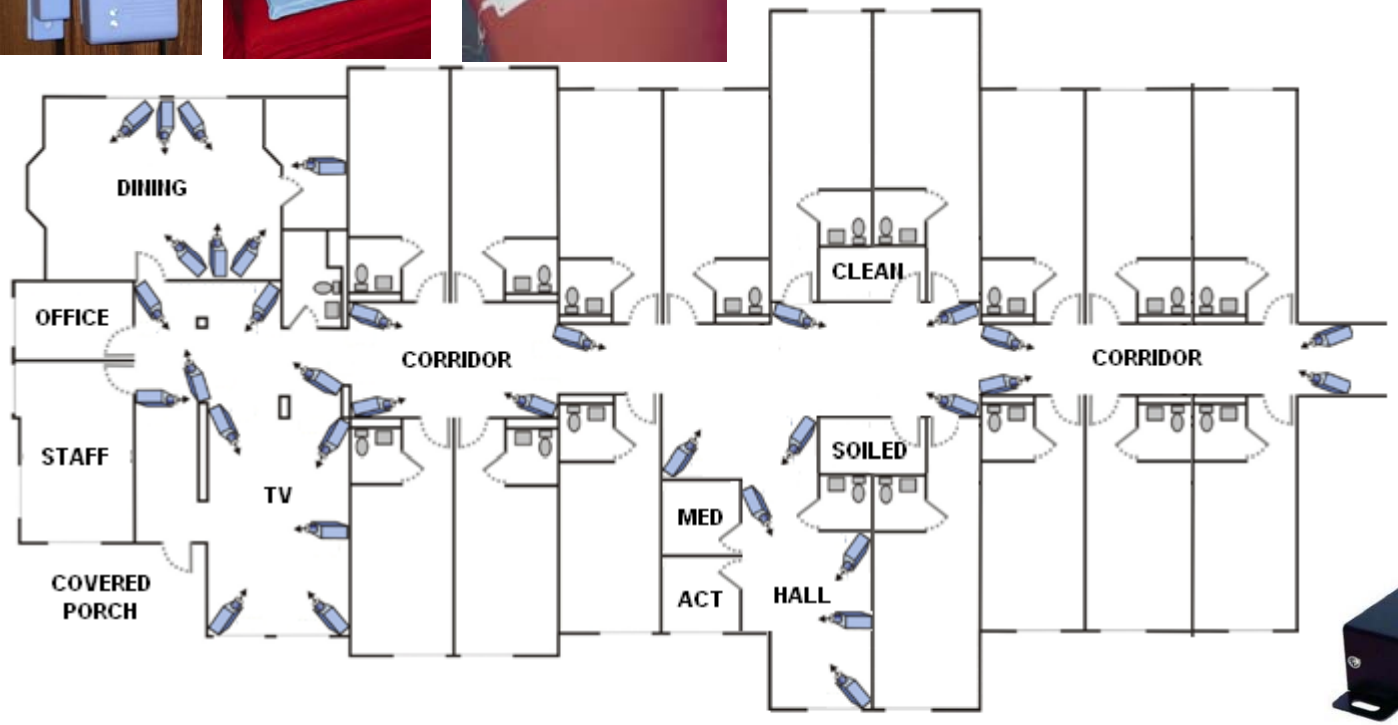
- Depression



Human health

- Depression





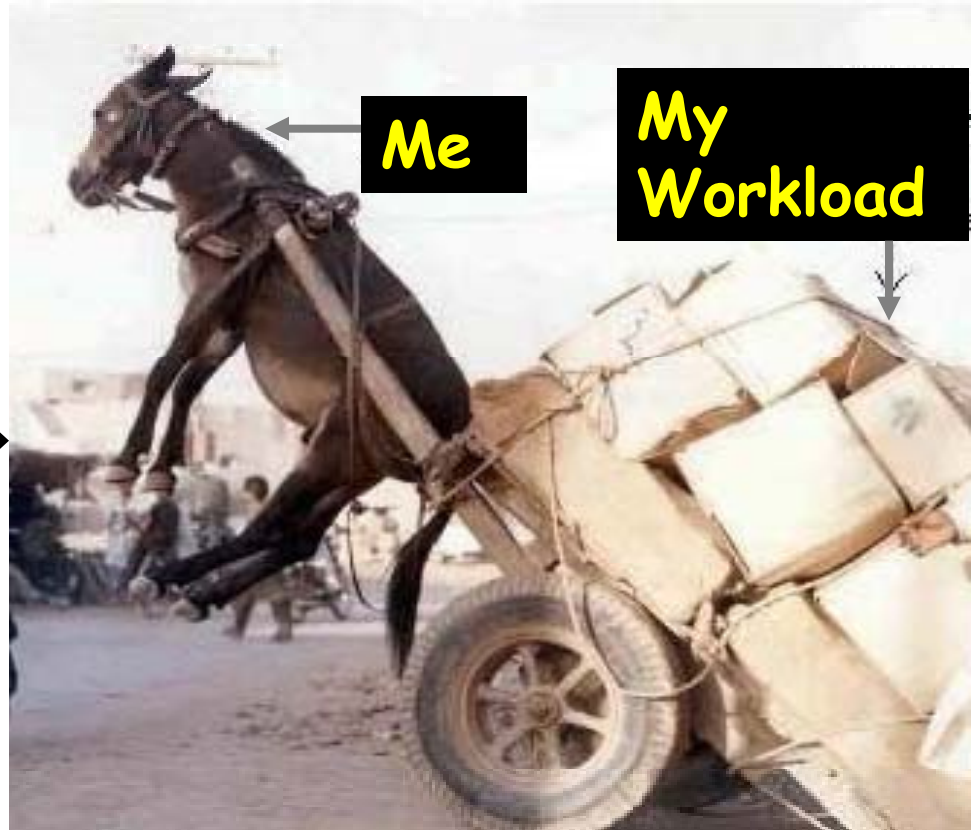
Human health

- Depression
- Deception detection
- Hot flash detection

Time management

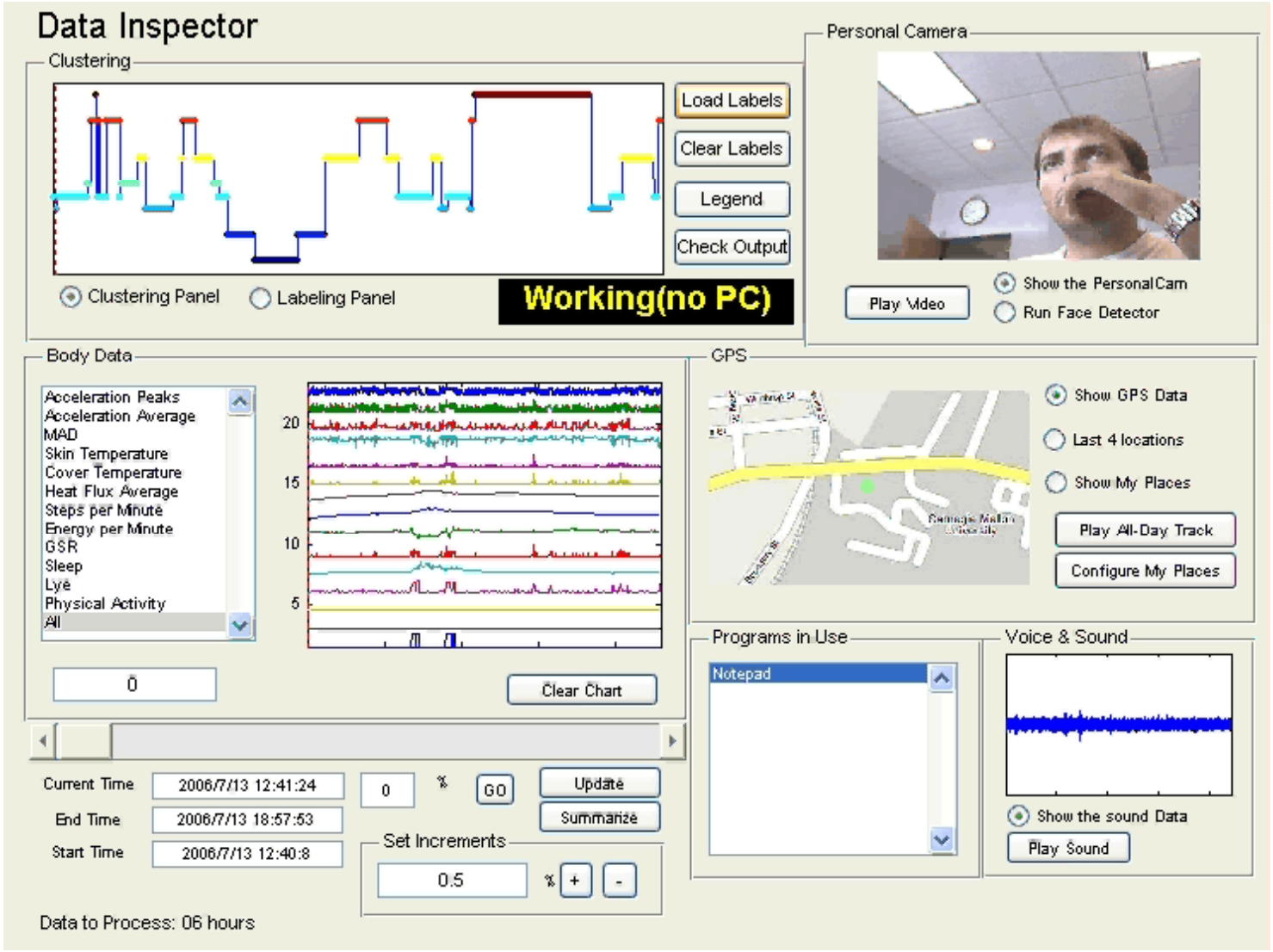
8:00 am

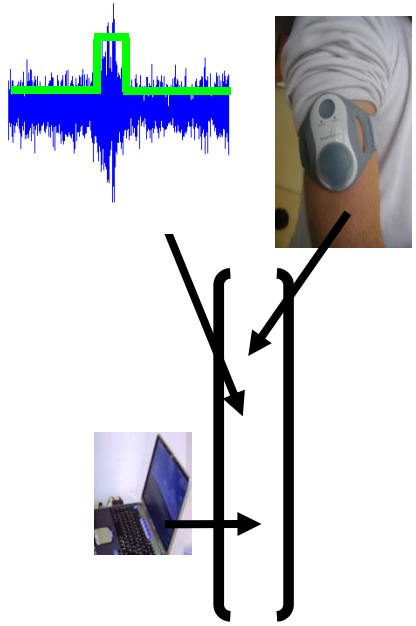
Working day



8:00 pm

Where my time goes?





Outline

- Introduction
- Facial image analysis (FIA)
  - Applications
  - Facial feature detection and tracking
  - Supervised facial expression recognition
  - Unsupervised discovery of facial events
- Conclusions

What does the face reveal?

Facial Image Analysis (FIA)

- Tasks in FIA:
  - Facial feature detection
  - Facial feature tracking (registration)
  - Facial image classification (e.g. expression)
  - Facial synthesis (e.g. animation)

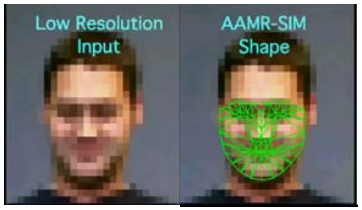
- Challenges:

Person Variability

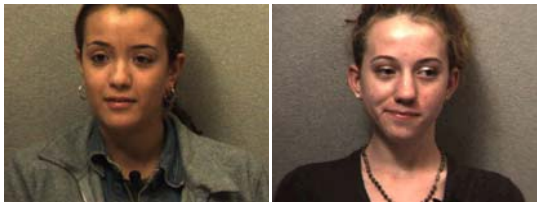
Pose

Occlusion

Resolution



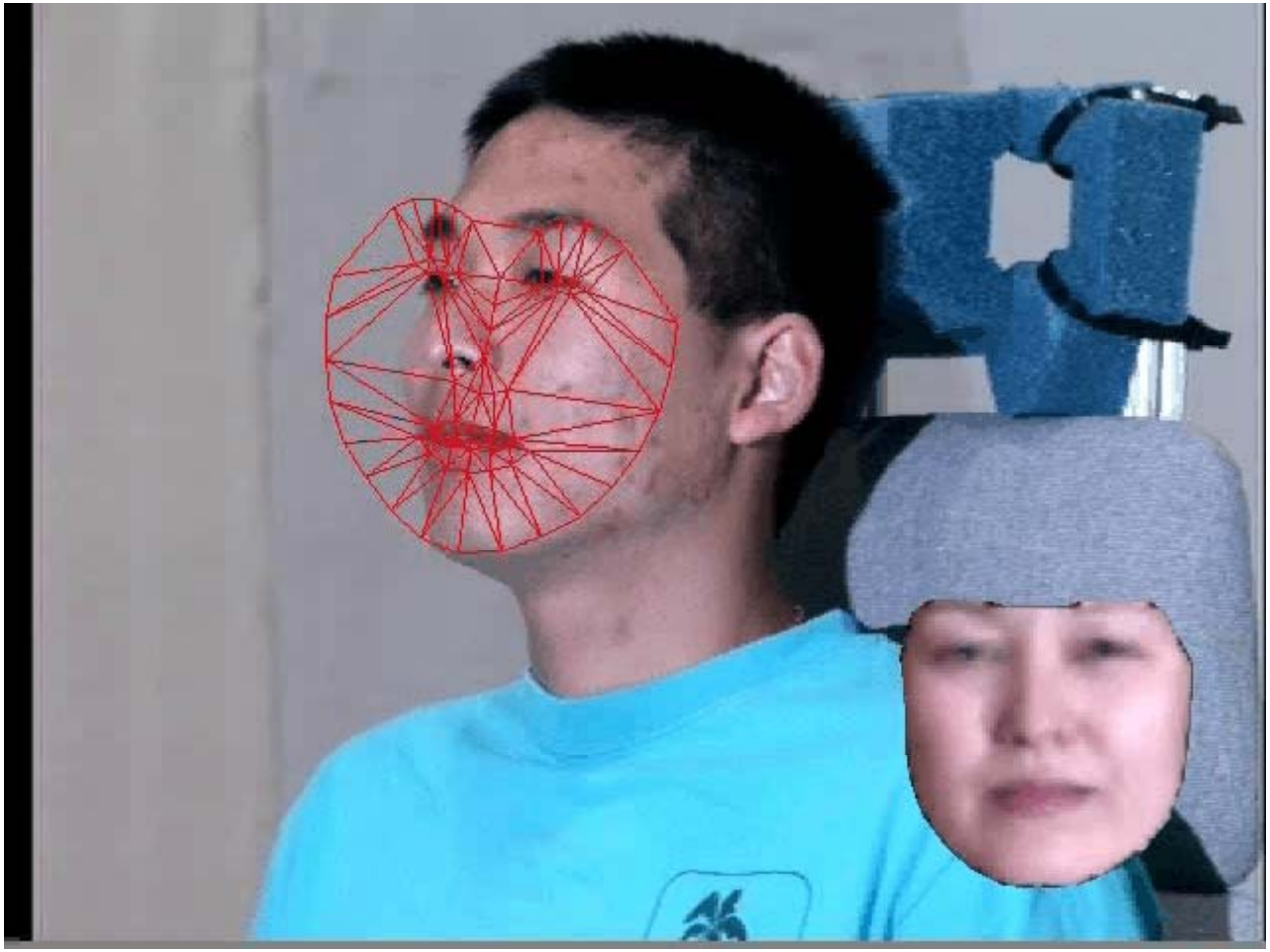
Subtle facial behavior



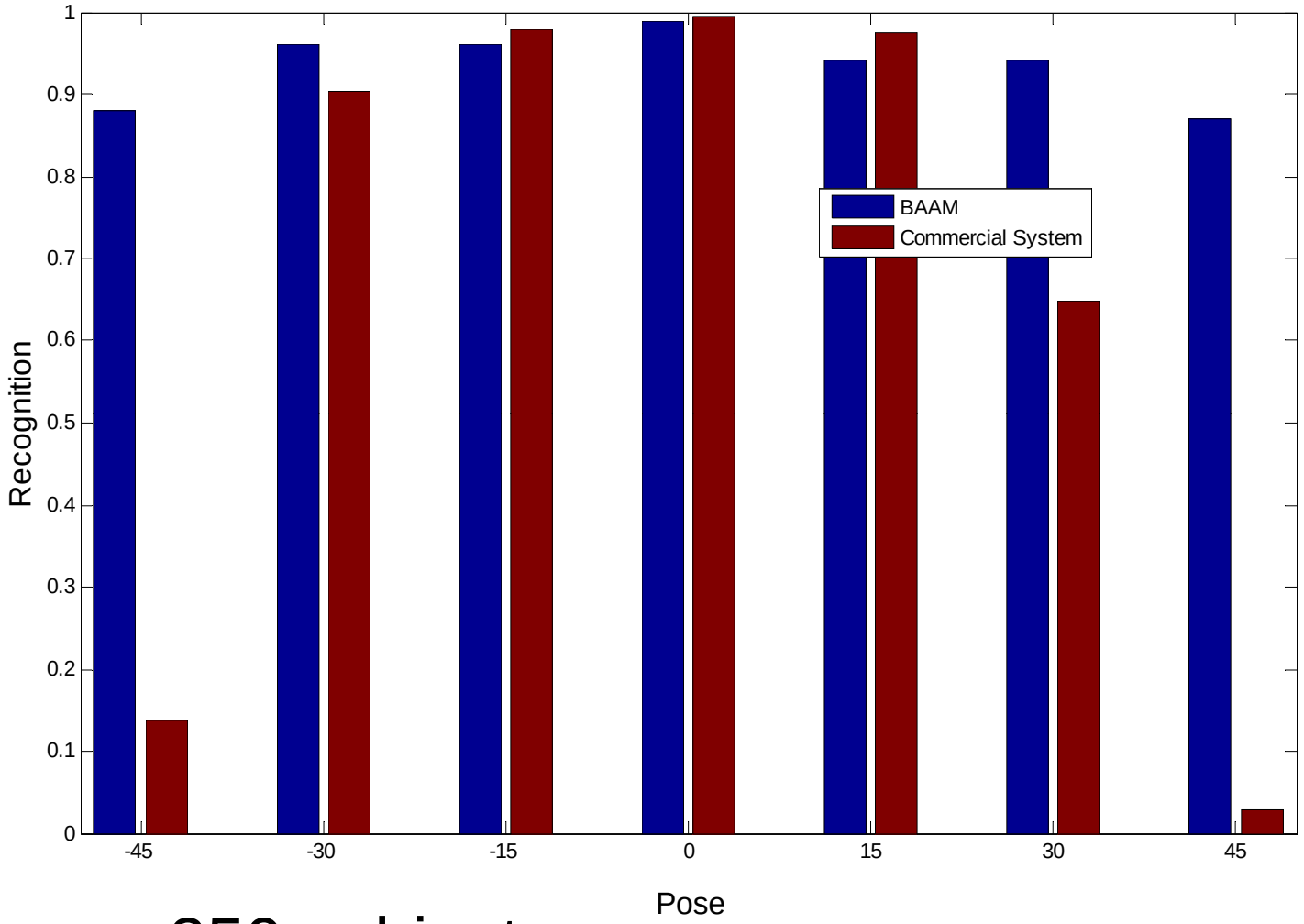
Illumination

FIA: face recognition across pose

Fitting BAAMs



FIA: experiment



- Gallery – 250 subjects
- Probe – 200 subjects different sessions/expressions

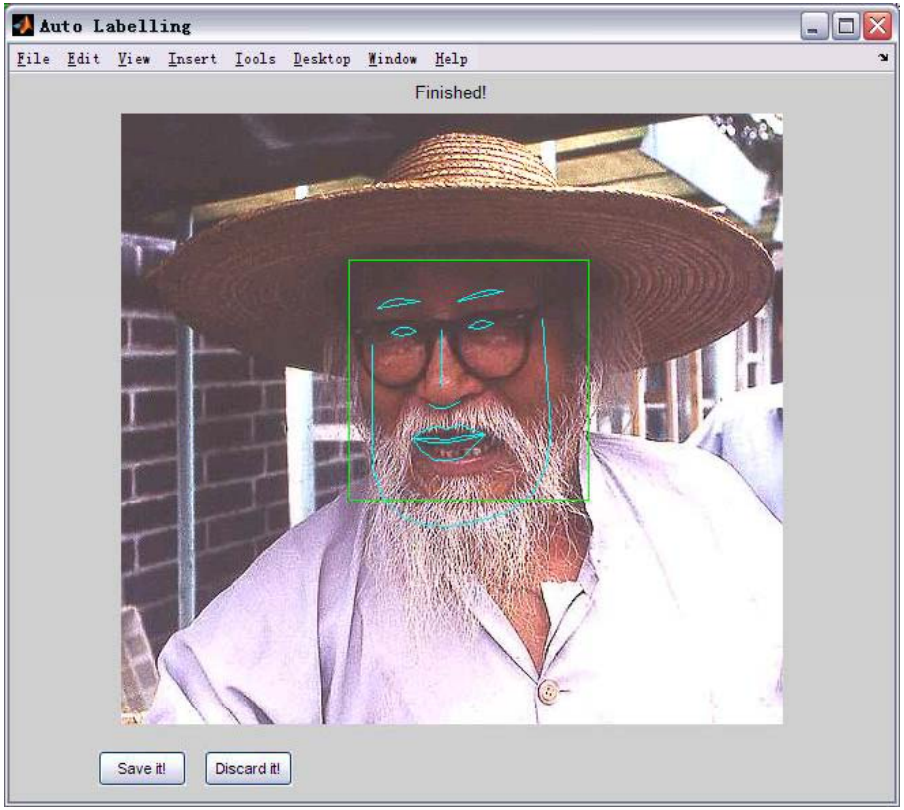
FIA: 3D face models

Fitting

3D face model

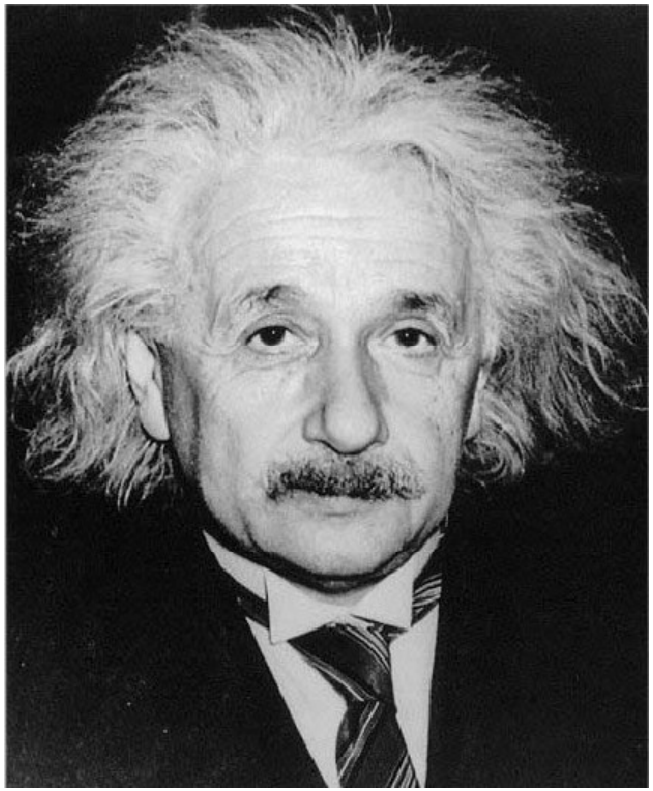
FIA: classifying subjects' characteristics

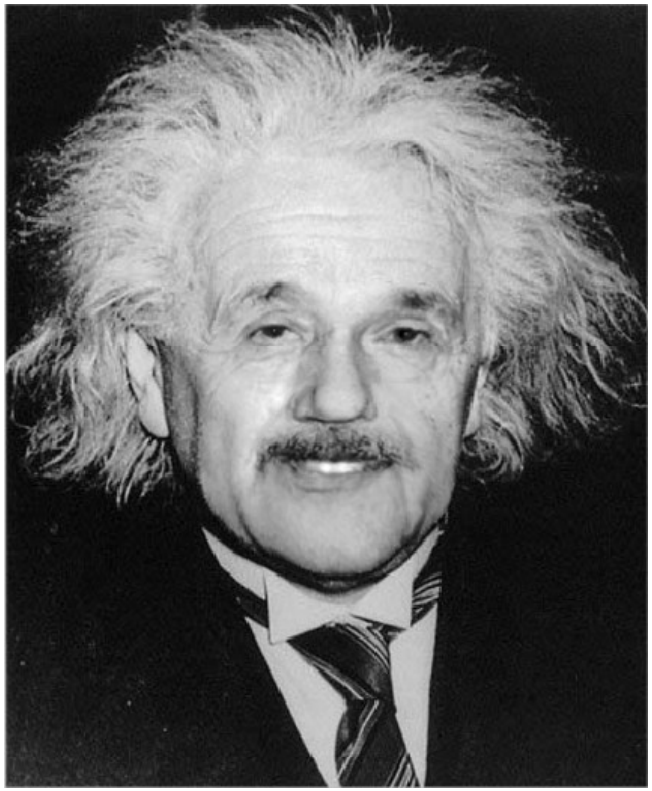
- Classify gender, age, ethnicity, subject attributes (e.g. beard) from images?



Ethnicity: Chinese
Gender: Male
Age: around 60
Other subject's attributes:
•Beard
•Glasses

FIA: facial expression synthesis

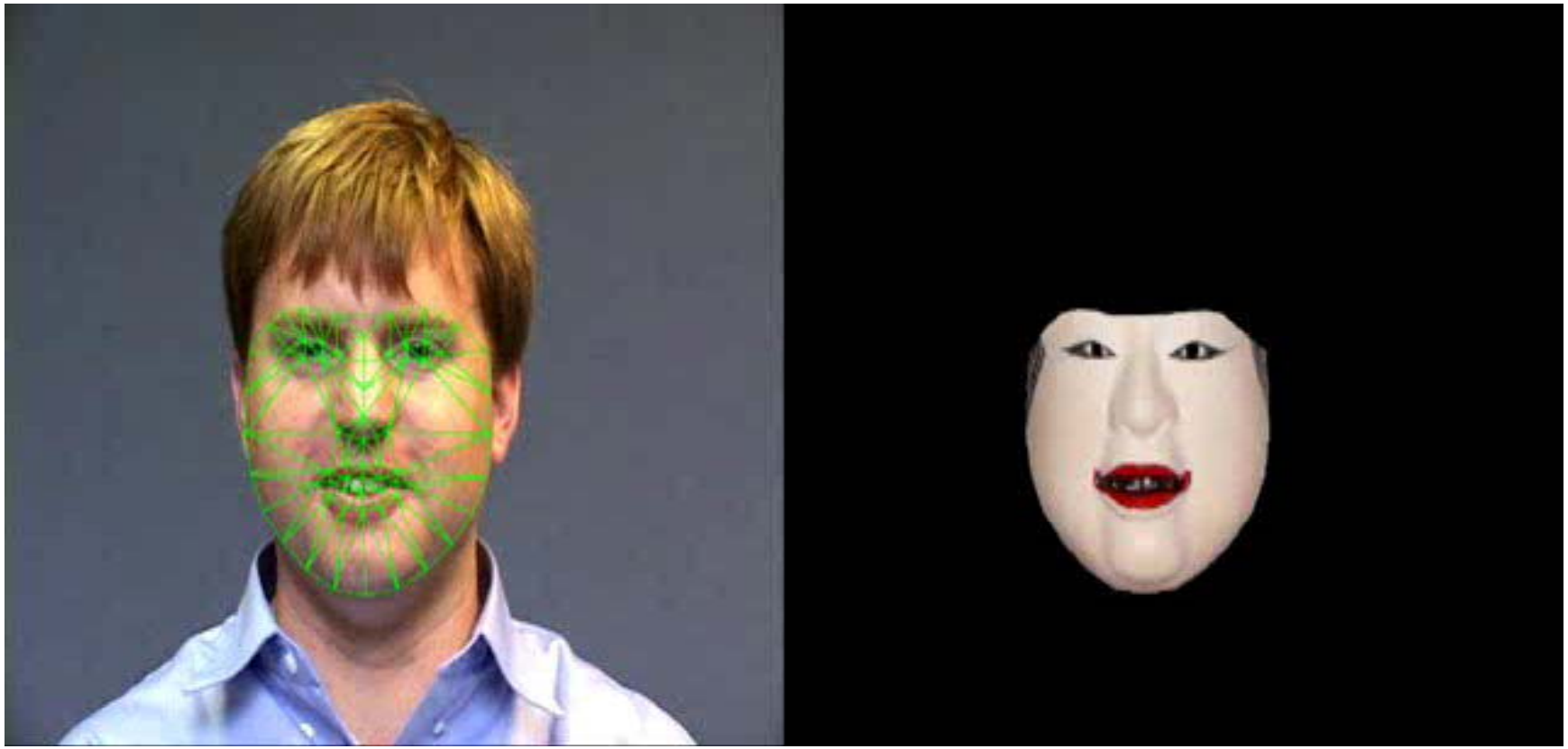




FIA: editing faces

FIA: glasses removal

FIA: Virtual avatar

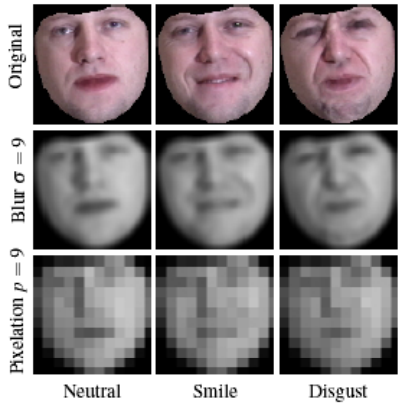


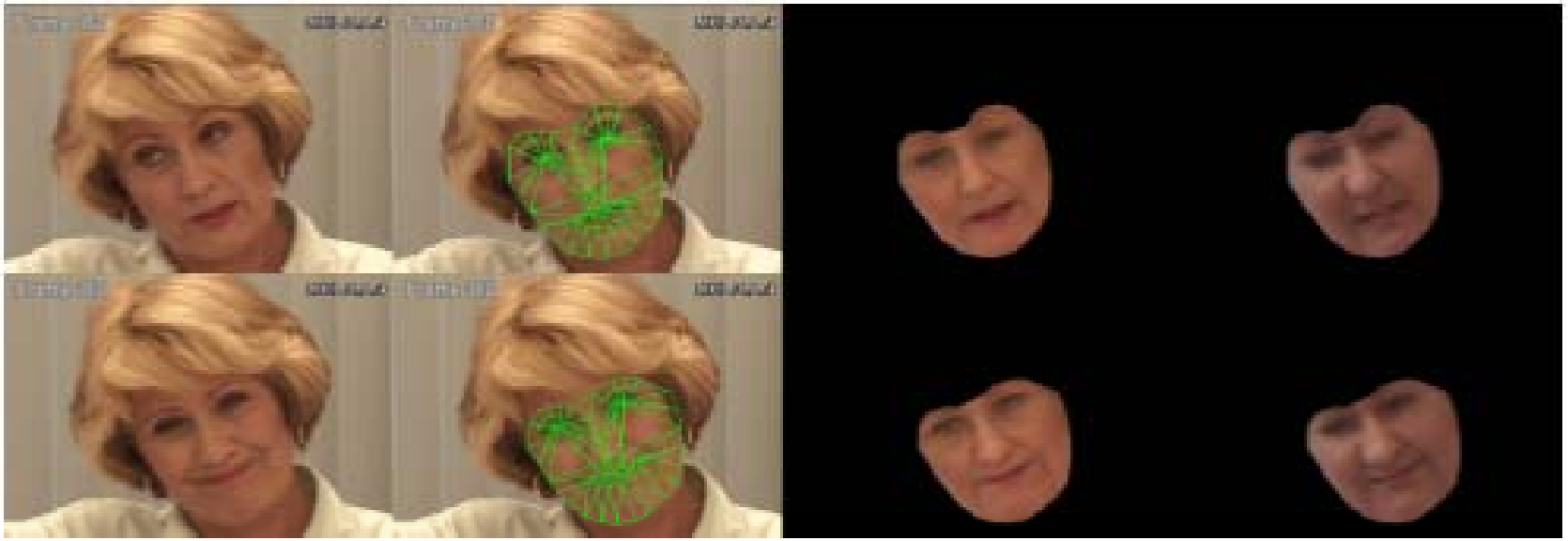
[Thanks S. Baker and I. Matthews]

FIA: speech-driven facial animation

[Thanks I. Matthews]

FIA: face de-identification



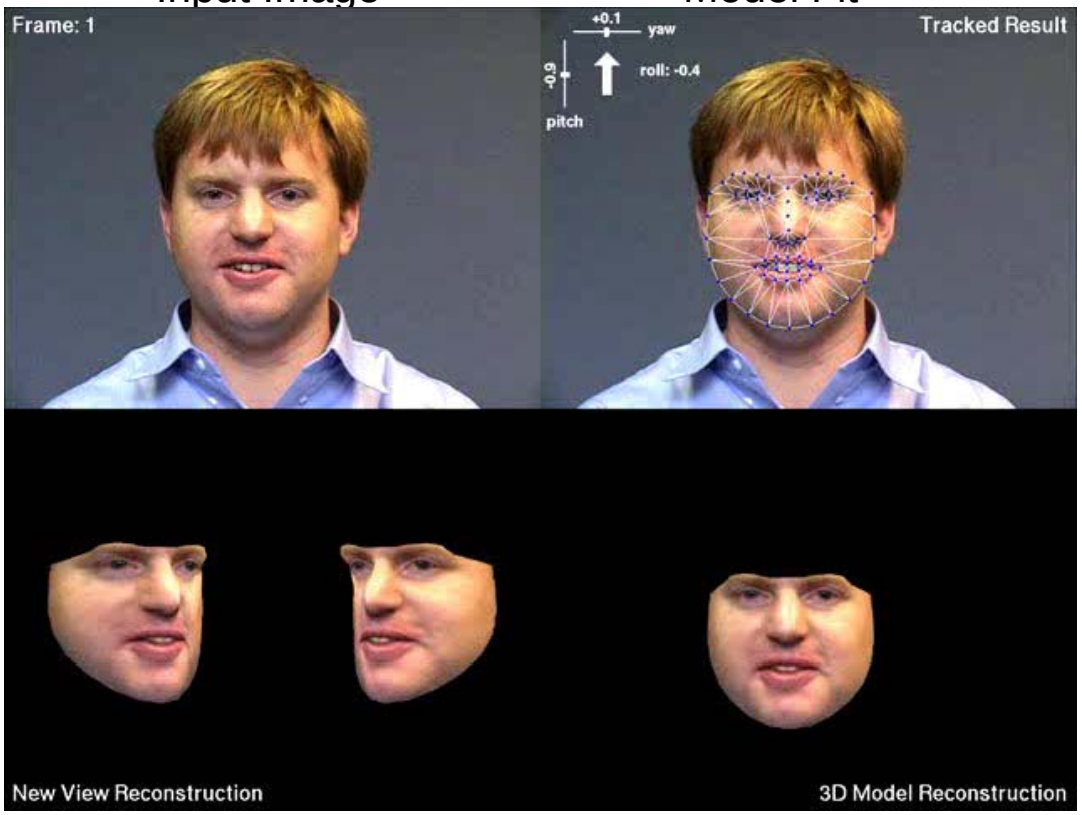


Input AAM Track k-Same Multi-Factor

FIA: low-bandwidth video conferencing

Input Image

Model Fit



10's of Params

Fit Model

Code and Transmit

10's of Params

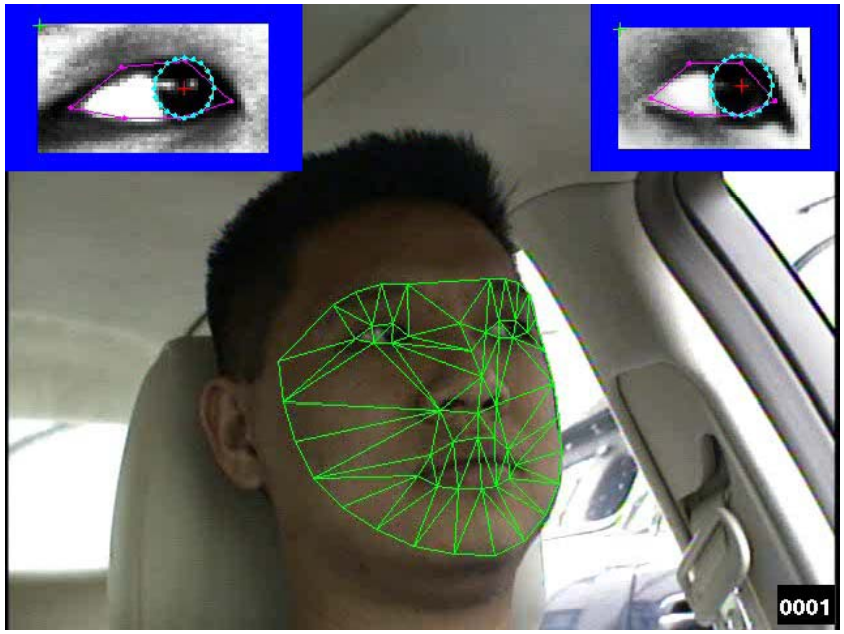
Render Model

Reconstructed Novel Poses

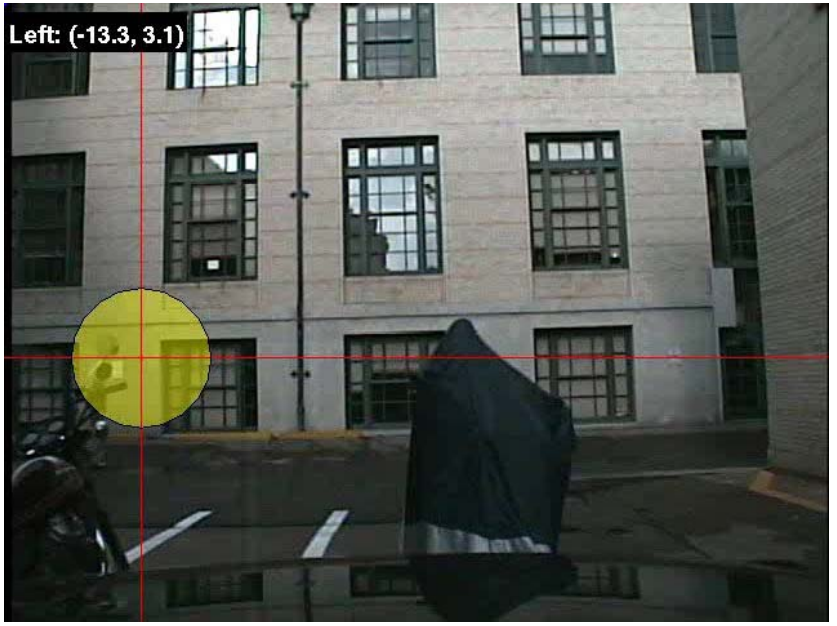
Reconstructed From Parameters

[Thanks S. Baker and I. Matthews]

FIA: driver's gaze estimation



Driver Camera



Exterior View Camera

[Thanks S. Baker and I. Matthews]

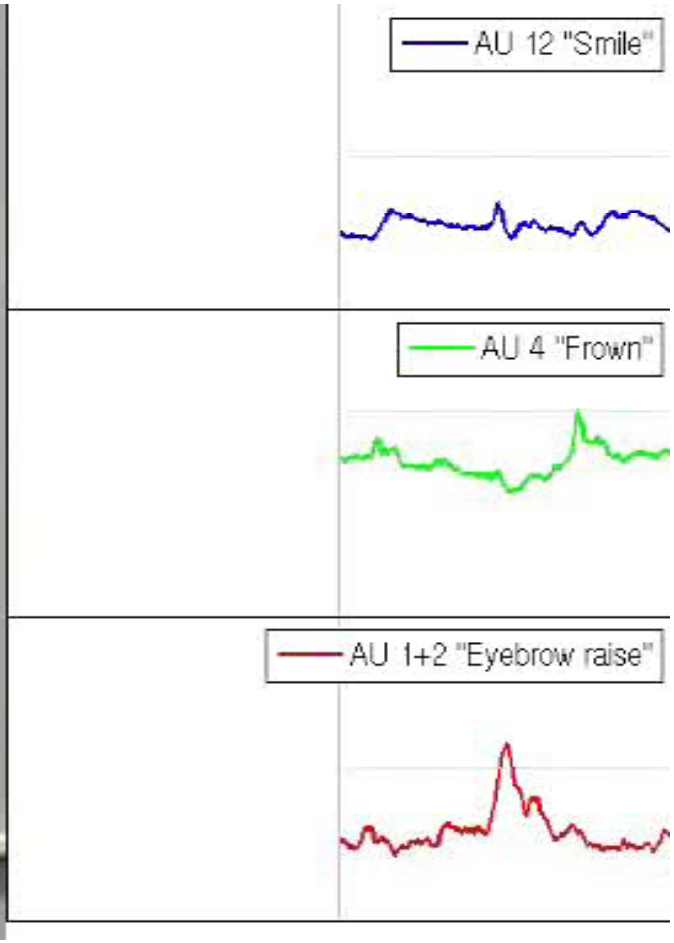
FIA: automatic FACS coding

Outer Brow Raiser **(AU2)**

Upper lid raiser **(AU5)**

Lip Tightener **(AU23)**

Nose Wrinkler **(AU9)**

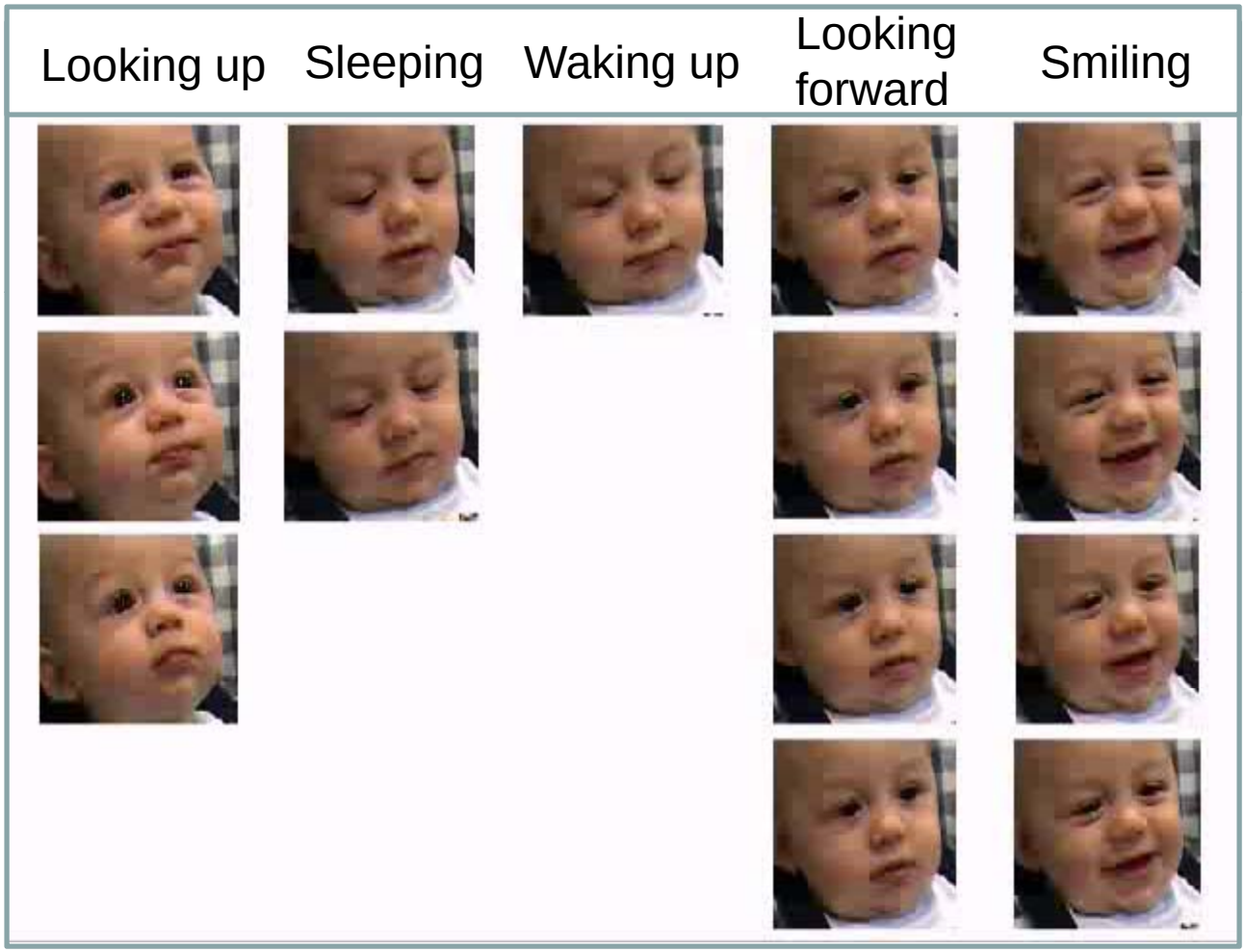


FIA: mining facial expression

FIA: Mining facial expression

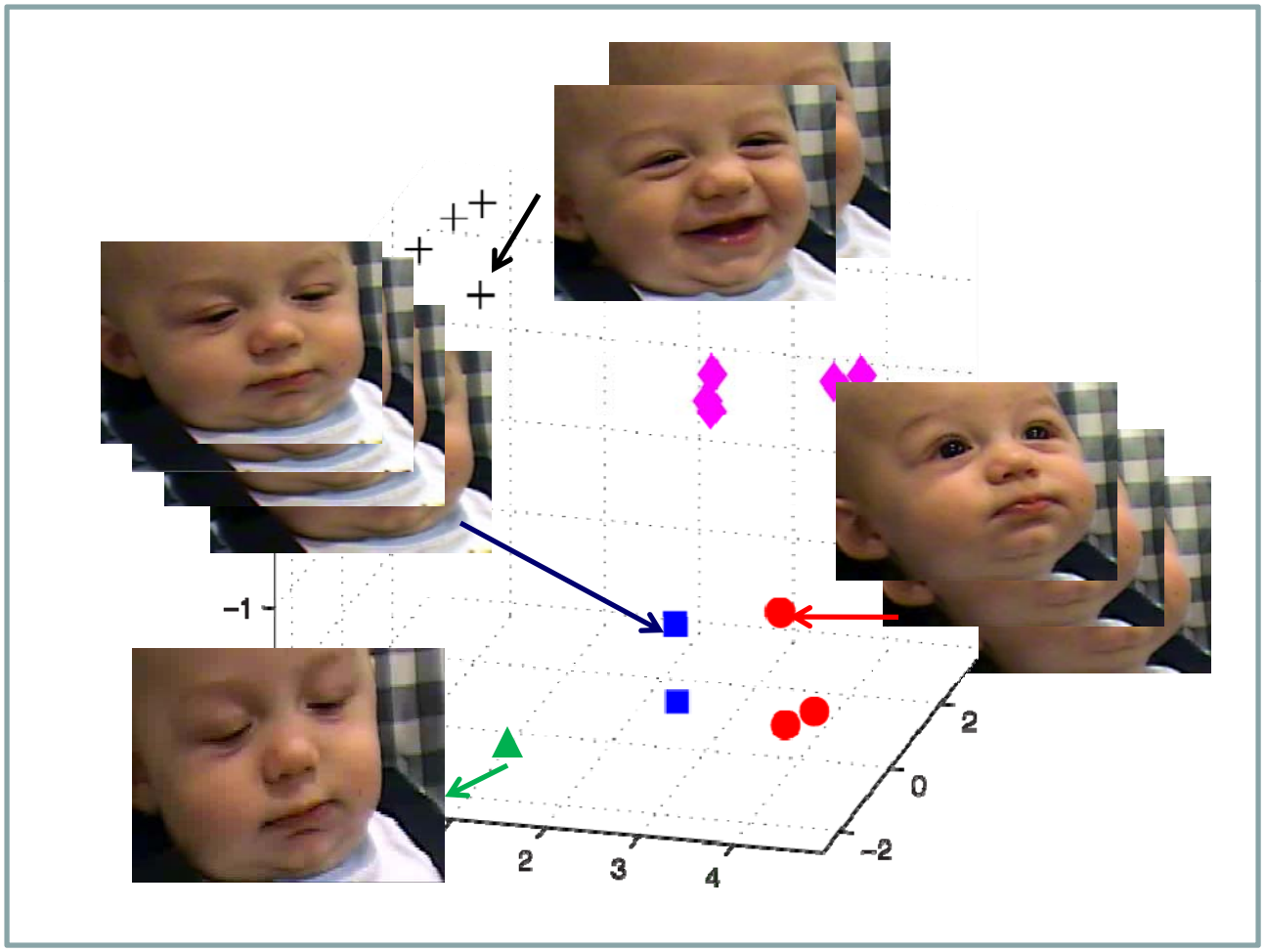
- **Summarization**

FIA: Mining facial expression



- **Summarization**

FIA: Mining facial expression



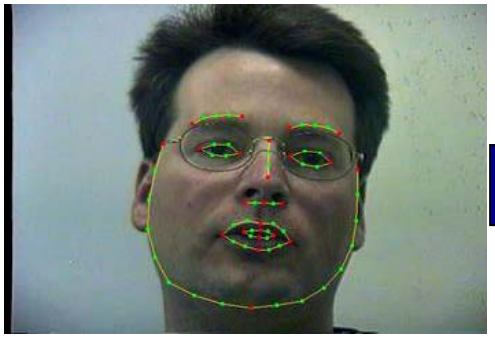
- ☑ **Summarization**
- **Embedding**

Outline

- Introduction
- Facial image analysis (FIA)
  - Applications
  - Facial feature detection and tracking
  - Supervised facial expression recognition
  - Unsupervised discovery of facial events
- Conclusions

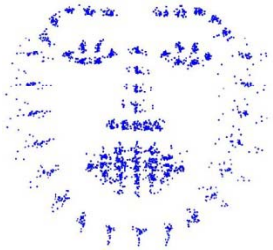
Active Appearance Models (AAMs)

(Cootes et al ‘96, De la Torre et al ’00, Matthews and Baker ‘02)



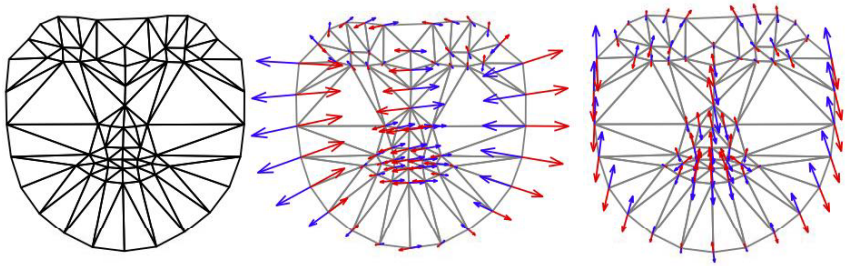
Hand-Labeled Training Data

Procrustes



PCA

Shape modes



s_0 s_1 s_2

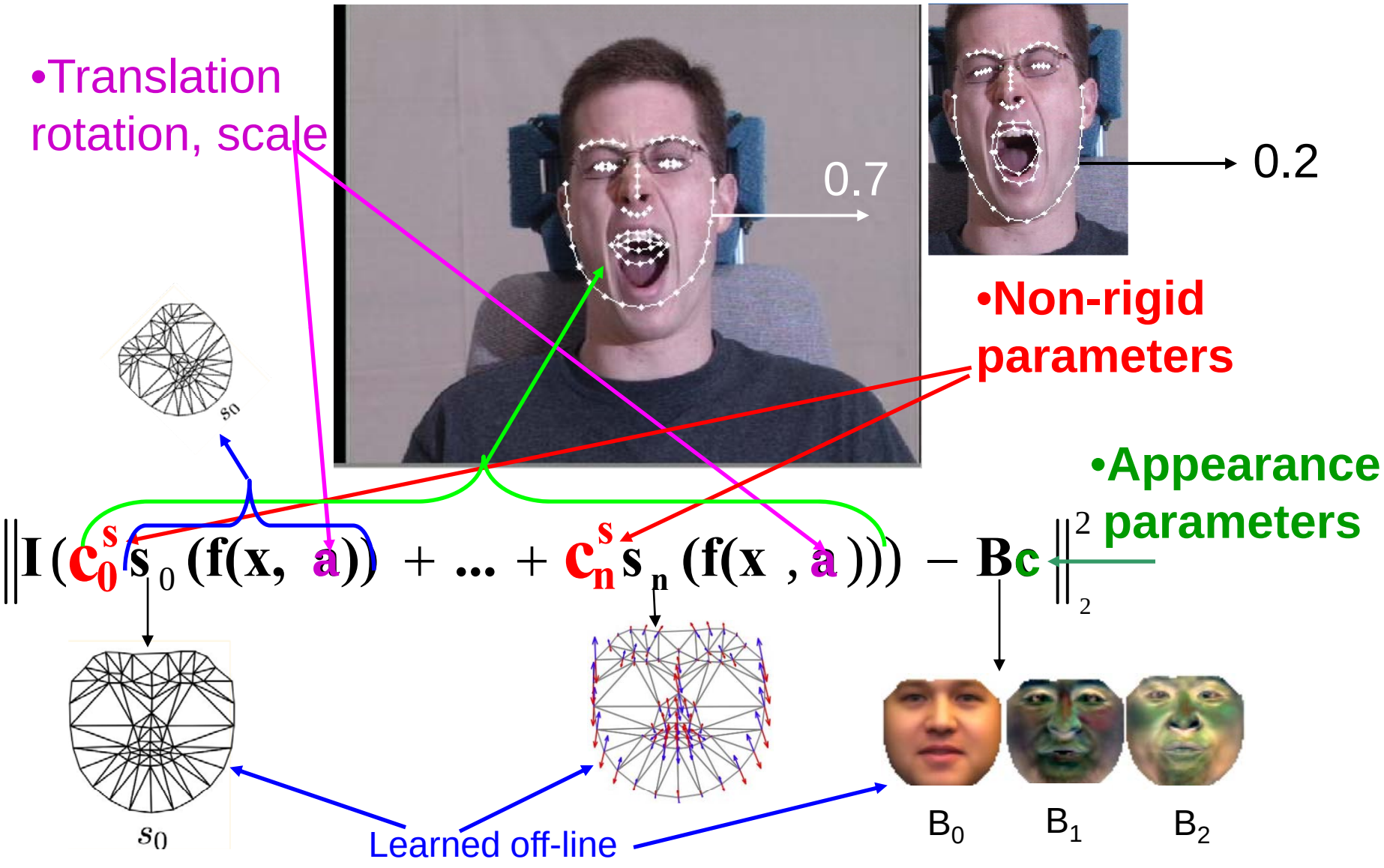
Shape normalised images

PCA

Appearance modes

B_0 B_1 B_2

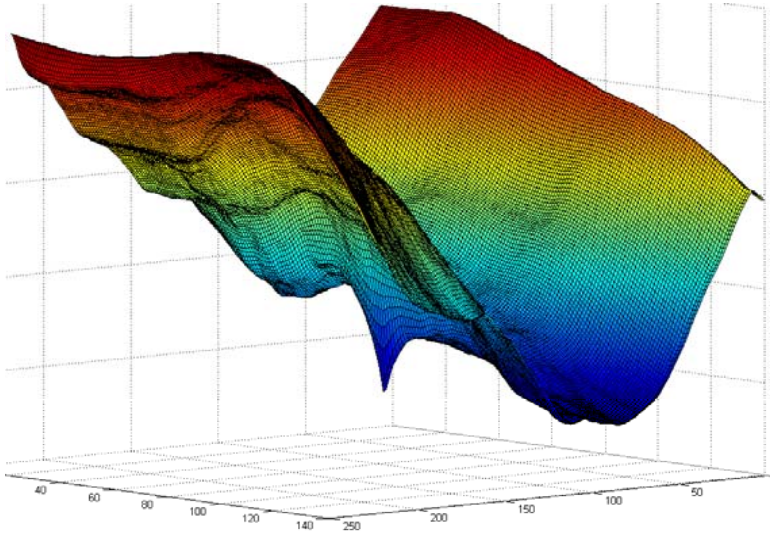
Tracking as an optimization problem

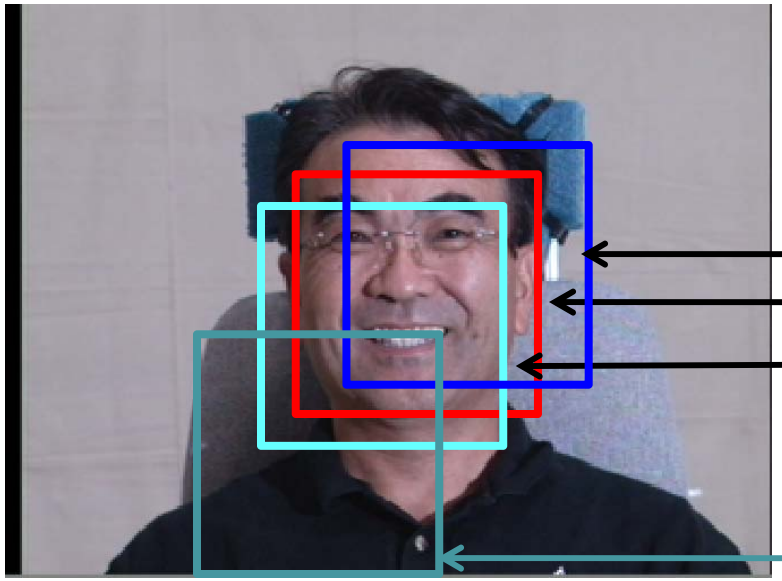


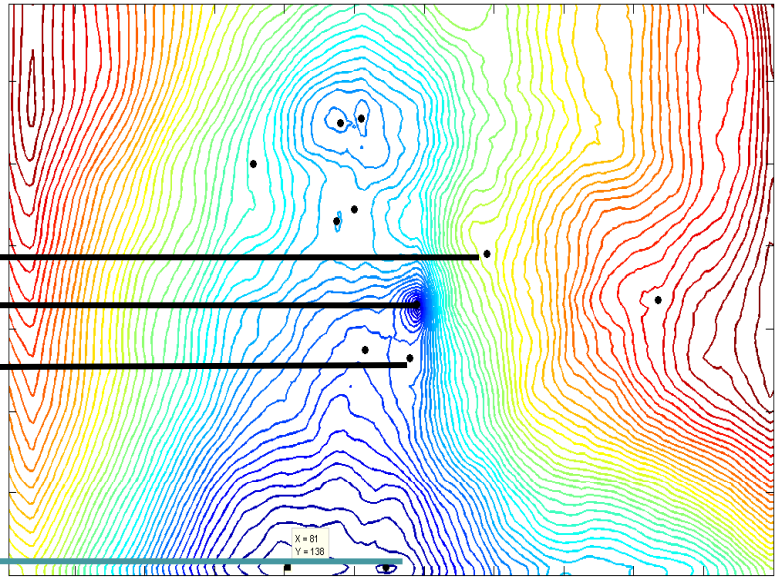
AMMs do not generalize that well

- 33% of convergence in unseen faces (th= 5 pixels)

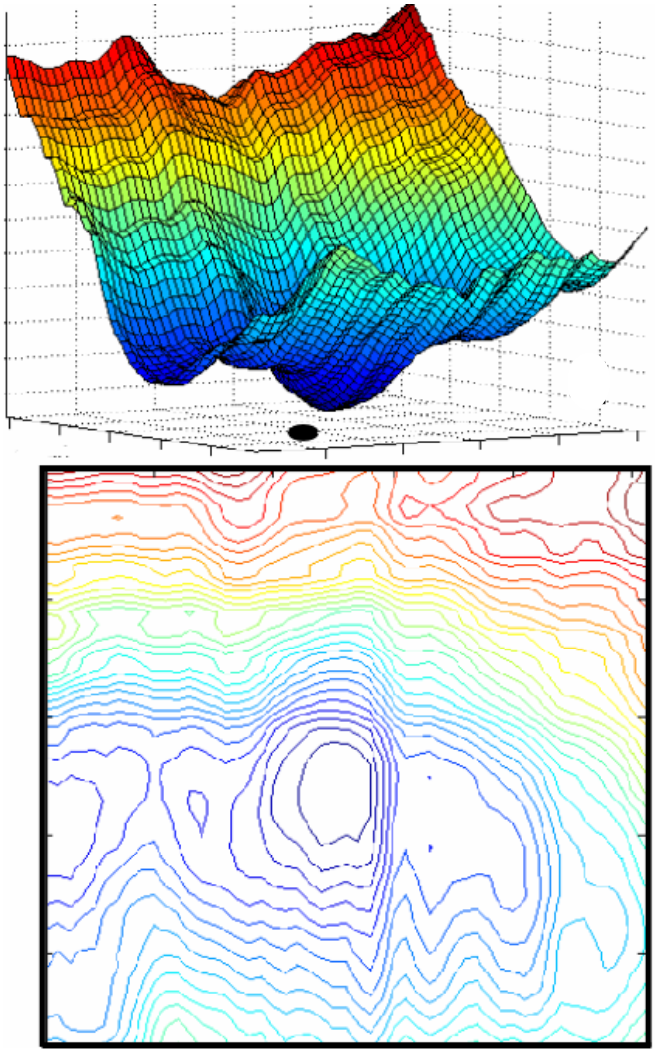
Problem





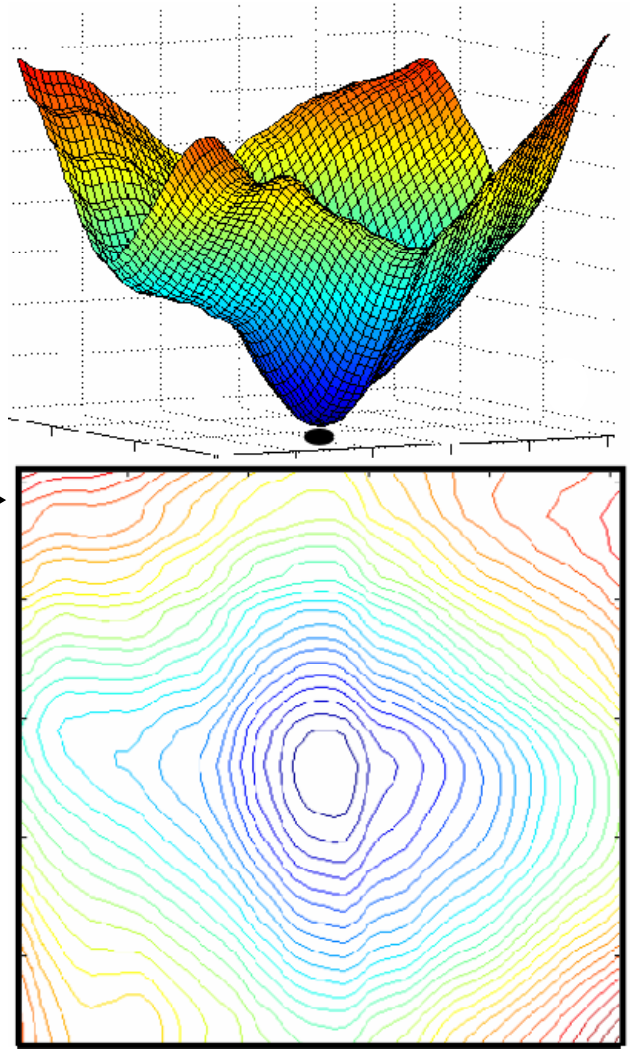


Problem II



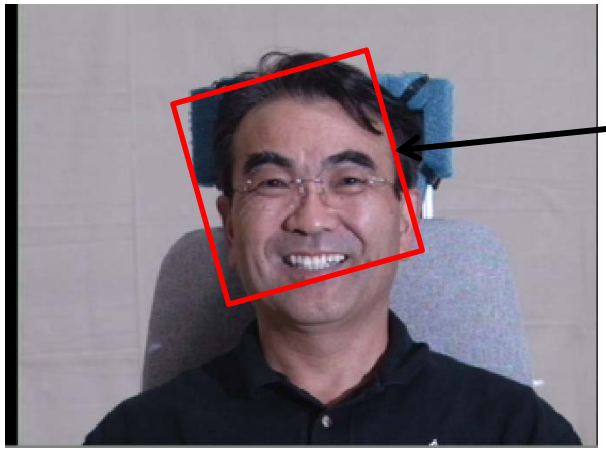
Learn the cost function

(Lecun et al. 03-09, Caetano et al 07, Lacoste-Julien et al. 06, Leordeanus and Hebert 05, Wimmer et al. 06, Lu 08)



One local minimum only at the right place

Alignment though optimization



pixels to align

Template

Image

$$\min_{\mathbf{p}} \left\| \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) - \mathbf{d}_{ref} \right\|_2^2$$

Geometric transformation

Motion parameters

1. Template Matching (Lucas-Kanade 81, E. H. Adelson 84, Kim 07,..),

2. Optical flow (Quam 84 , Anandan 89, Bergen et al 92, black & anandan 91, …),

3. AMs/MMs (Black & Jepson 98, Jones & Poggio 98, Cootes & Taylor 01, Blanz & Vetter 99, de la Torre et al 02, Baker et al 04 , …)

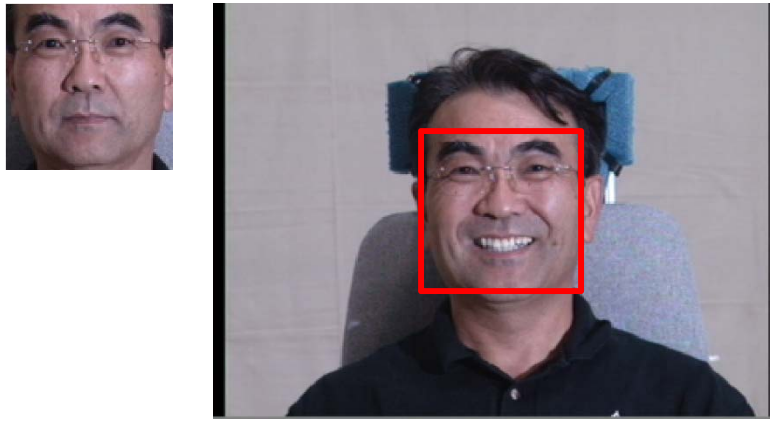
Quadratic cost function

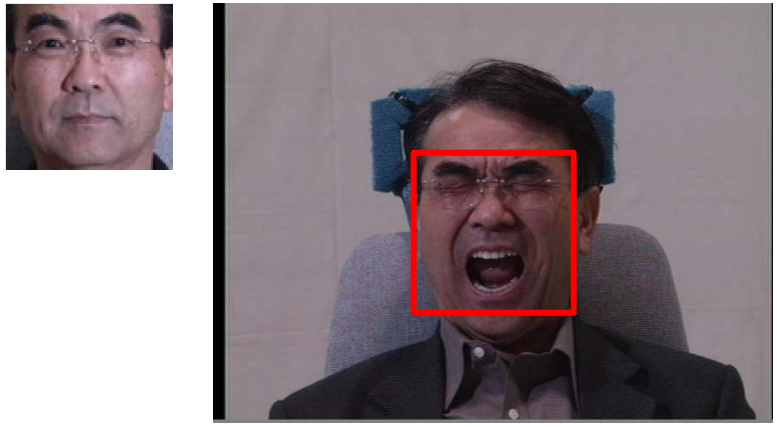
Learn **A** and **b**

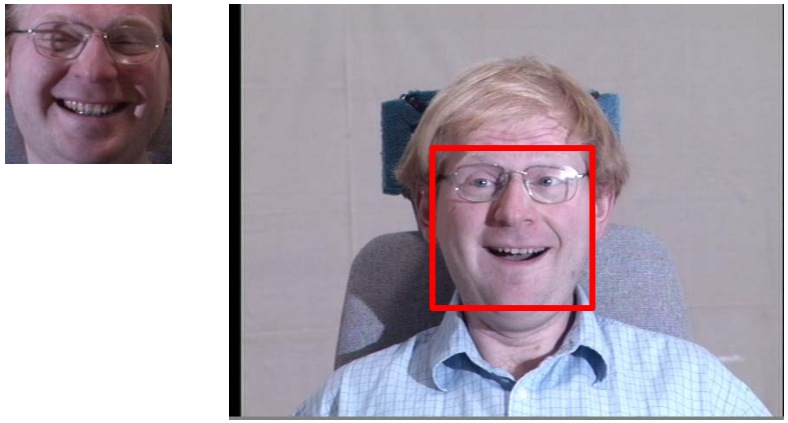
$$E(\mathbf{d},\mathbf{p}) = \mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))^T \; \mathbf{A} \; \mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) + 2\,\mathbf{b}^{\mathrm{T}}\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))$$

$$\mathbf{A} = \mathbf{I}_m \text{ and } \mathbf{b} = -\mathbf{d}_{ref}$$

Multiple training images







...

$$(\mathbf{d}_i, \mathbf{d}_i^{ref}, \mathbf{p}_i)$$

Image **Template** **Ground truth**

1st Desired property of the cost function

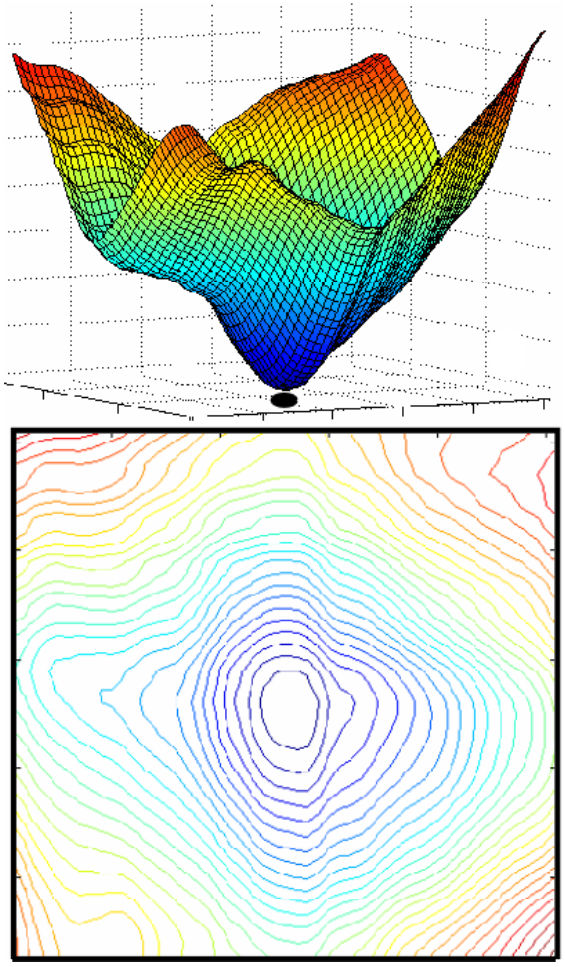
- Local minimum at the expected location

i th training image

$$\left.\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}}\right|_{\mathbf{p}_i} = \mathbf{0}$$

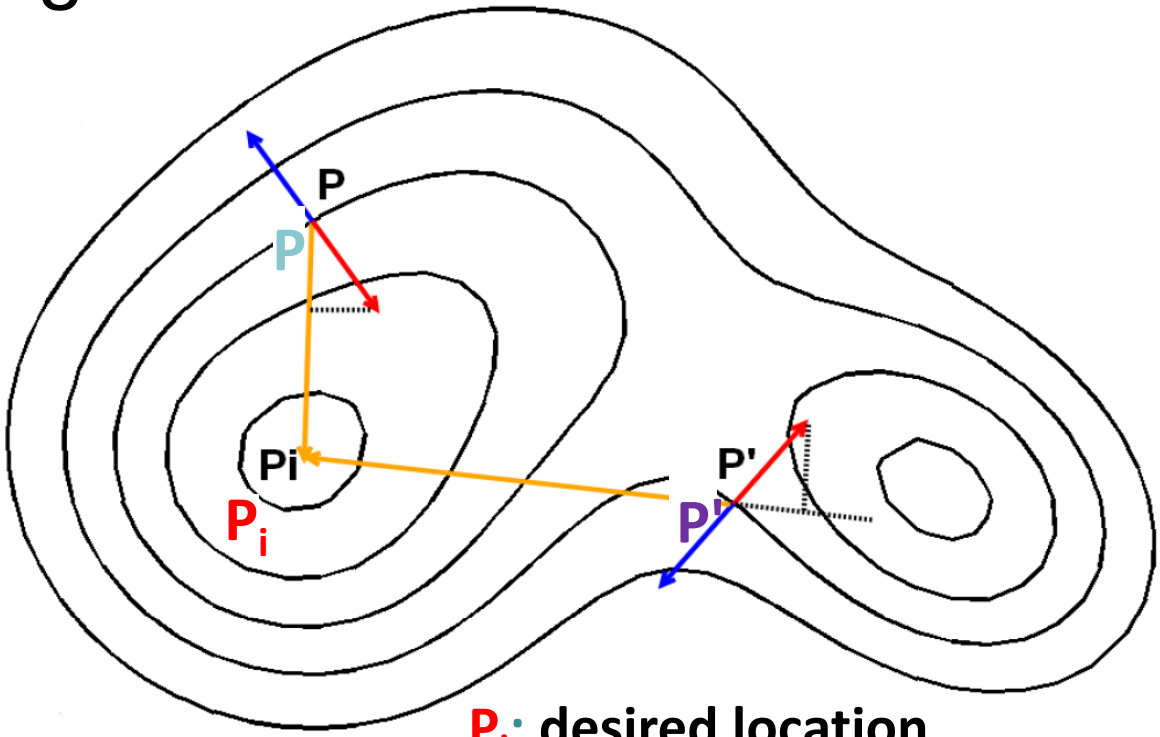
Correct alignment parameter

$$\left\| \left.\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}}\right|_{\mathbf{p}_i} \right\|_2^2 = \mathbf{0}$$



2on Desired property of the cost function

- Gradient descent agrees with the optimal walking direction



P_i: desired location

At P, gradient descent moves closer to P_i
At P', gradient descent moves away from P_i

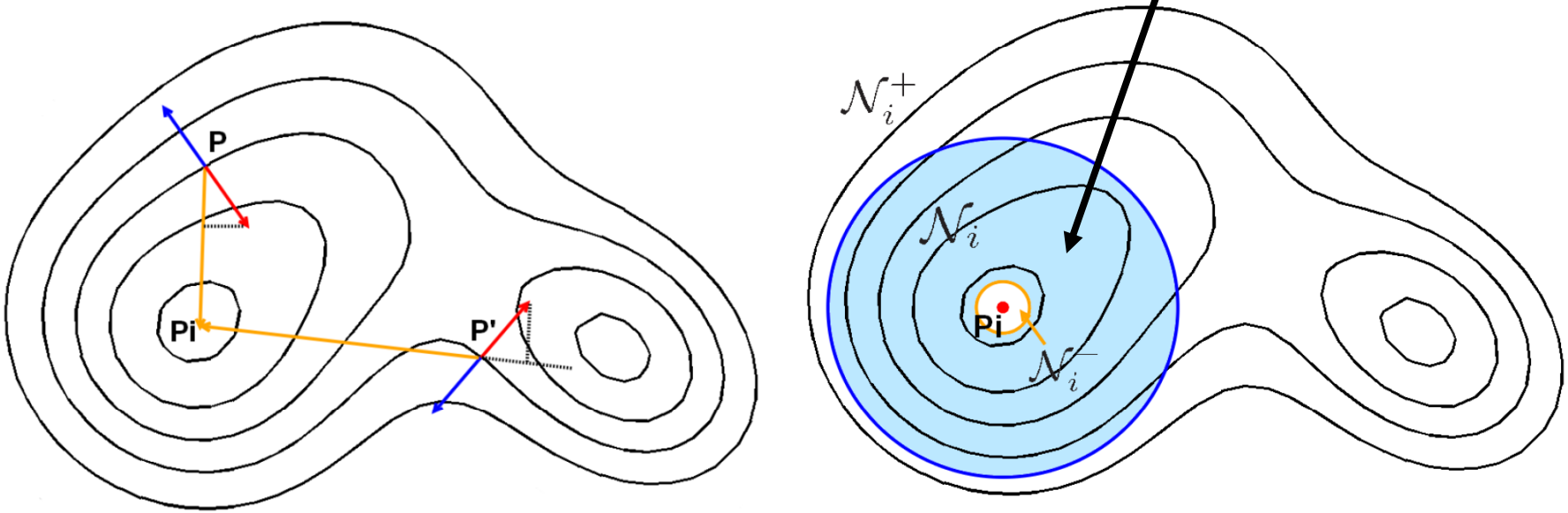
Enforcing the 2nd desired property:

$$\left\langle -\left(\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right)^T, \frac{\mathbf{p}_i - \mathbf{p}}{\|\mathbf{p}_i - \mathbf{p}\|^2} \right\rangle > 0 \quad \forall \mathbf{p} \in N_i$$

Gradient descent direction

Optimal direction

Appropriate neighbourhood



The learning problem

$$\left\| \left. \frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right|_{\mathbf{p}_i} \right\|_2^2 = 0 \quad \forall_i$$ 1st desired property

$$\left\langle -\left(\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right)^T, \frac{\mathbf{p}_i - \mathbf{p}}{\|\mathbf{p}_i - \mathbf{p}\|^2} \right\rangle > 0 \quad \forall \mathbf{p} \in N_i \ \forall_i$$ 2nd desired property

Convex optimization problem

$$\underset{\mathbf{A}, \mathbf{b}, \boldsymbol{\xi}}{\text{minimize}} \quad \sum_i \frac{1}{2} \left\| \left. \frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right|_{\mathbf{p}_i} \right\|_2^2 + C \sum_i \xi_i$$

$$s.t. \left\langle -\left(\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right)^T, \frac{\mathbf{p}_i - \mathbf{p}}{\|\mathbf{p}_i - \mathbf{p}\|^2} \right\rangle \geq M - \xi_i \quad \forall \mathbf{p} \in N_i \ \forall_i$$

$$\xi_i \geq 0 \ \forall_i$$

Weighted template alignment

$$\min_{\mathbf{p}} ||\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{d}_{ref}||_2^2$$

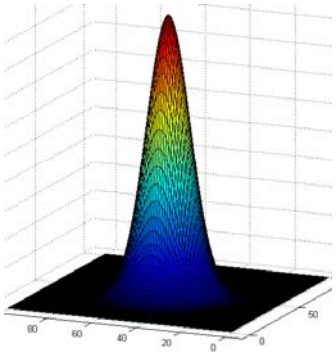
$$(\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{d}_{ref})^T diag(\mathbf{w})(\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{d}_{ref})$$

$$E(\mathbf{d},\mathbf{p}) = \mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))^T \mathbf{A}\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) + 2\mathbf{b}^T\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))$$

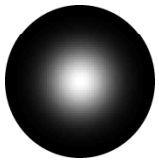
$$\mathbf{A} = diag(\mathbf{w})$$

$$\mathbf{b} = -diag(\mathbf{w})\mathbf{d}_{ref}$$

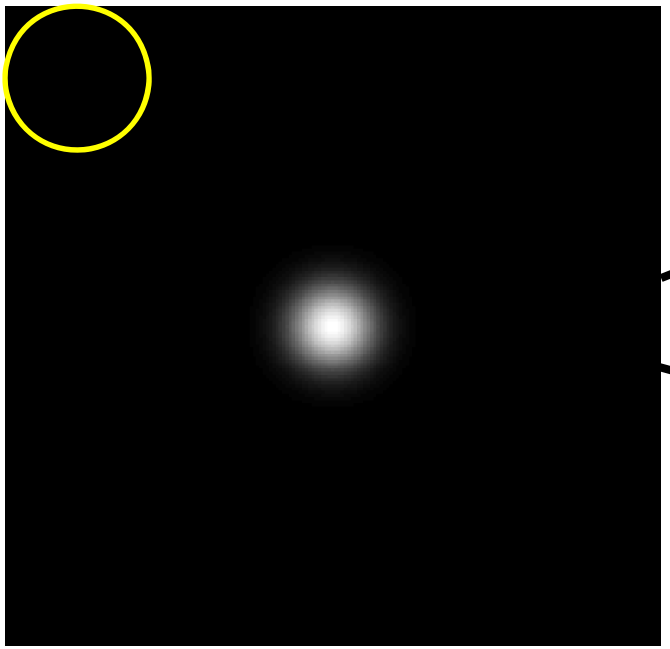
$$0 \leq w_i \leq 1$$



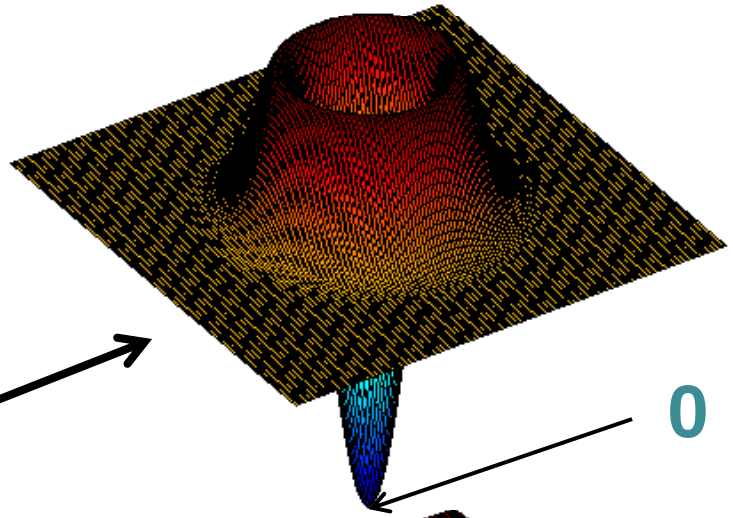
Error surfaces



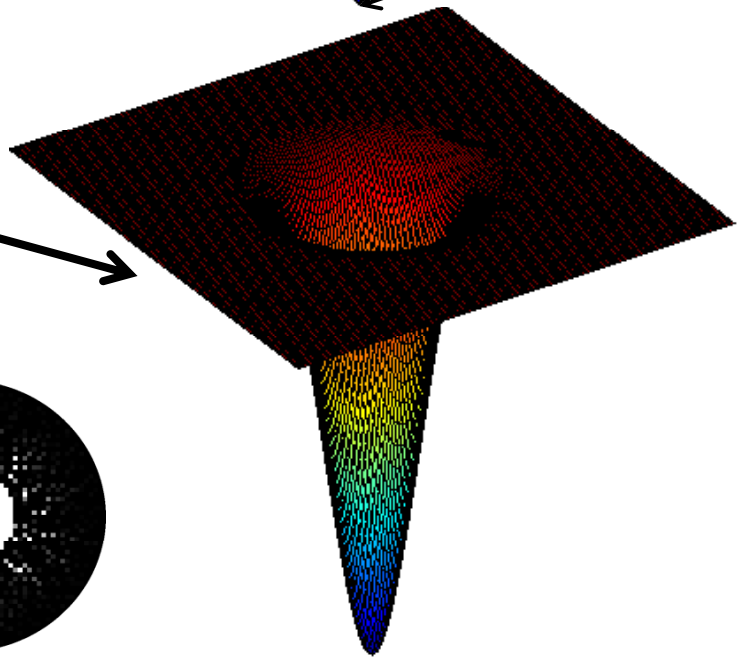
Template



Image



0



Weighted template

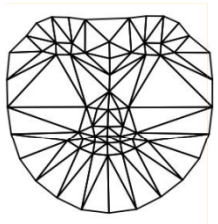
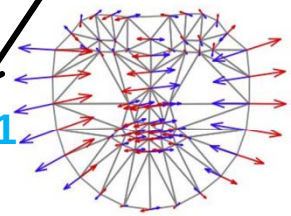
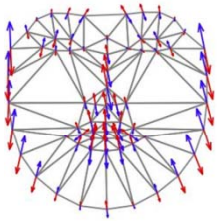
Weights

Active Appearance Models

(Cootes et al '96, De la Torre et al '00, Matthews and Baker '02)

$$\underset{\mathbf{p},\mathbf{c}}{\text{minimize}} \; ||\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{U}\mathbf{c}||_2^2$$

Affine parameters + Shape coefficients

Shape =  + s_1  + s_2  + …

Mean shape

PCA basis

U_0 $+c_1U_1$ $+c_2U_2$

Appearance Variation

Cost function for AAM

$$\underset{\mathbf{p},\mathbf{c}}{\text{minimize}} \left|\left|\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{U}\mathbf{c}\right|\right|_2^2$$

$$E(\mathbf{d},\mathbf{p}) = \mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))^T \; \mathbf{A} \; \mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) + 2\,\mathbf{b}^T\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))$$

$$\mathbf{A} = \mathbf{I} - \mathbf{U}\mathbf{U}^T = \mathbf{I} - \sum_{i=0}^{k} \mathbf{u}_i \mathbf{u}_i^T$$

$$\mathbf{b} = \mathbf{0}$$

weights to optimize

- Weighted basis alignment

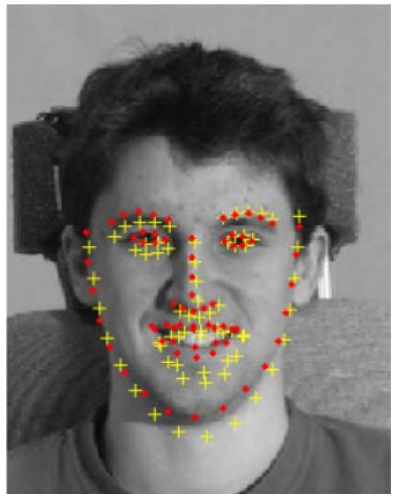
$$\mathbf{A} = \mathbf{I} - \mathbf{U}\mathbf{U}^T = \mathbf{I} - \sum_{i=0}^{k} \lambda_i \mathbf{u}_i \mathbf{u}_i^T \qquad 0 \le \lambda_i \le 1$$

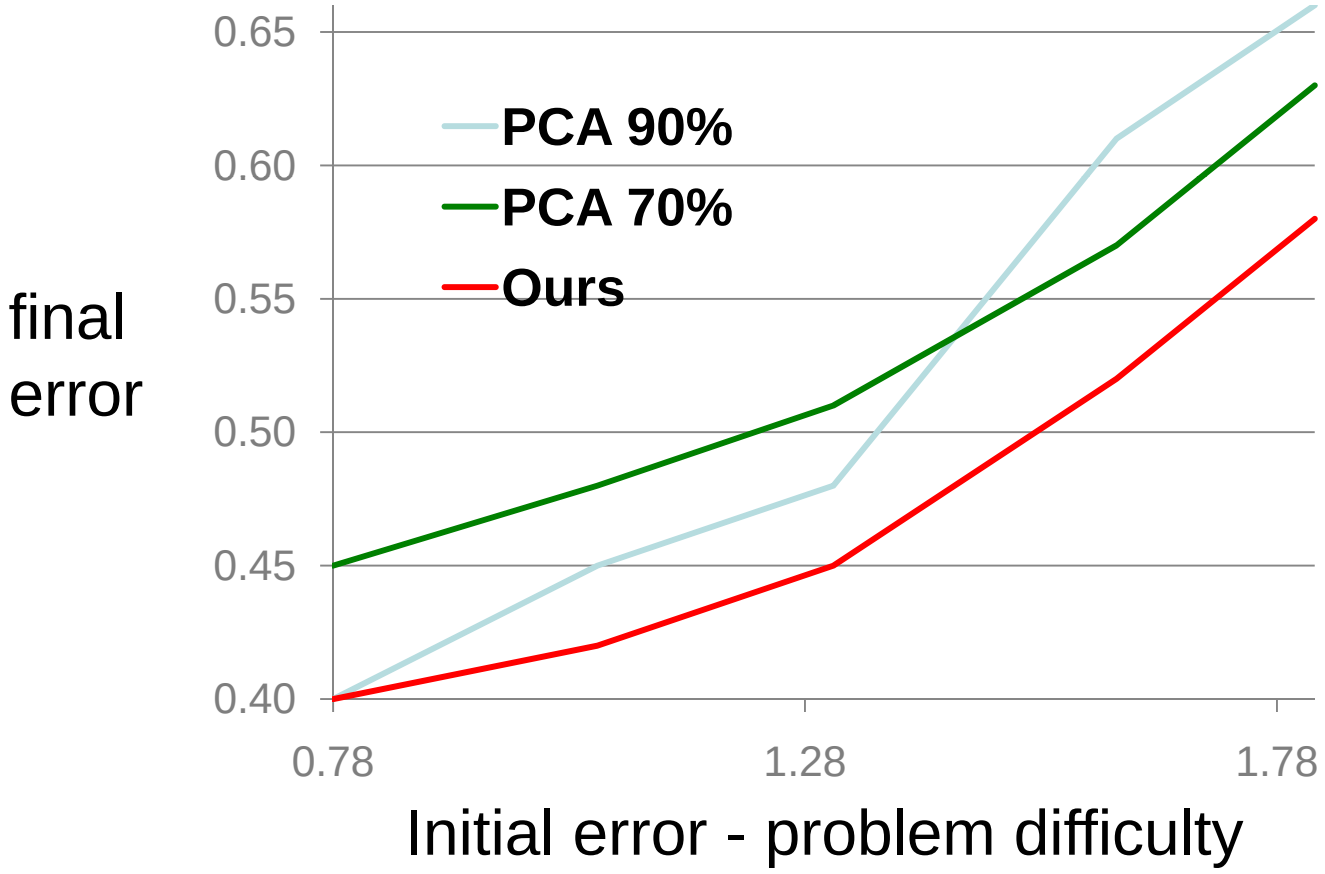
$$\mathbf{b} = \text{No constraint}$$

Results

Results – mean error curves

- MultiPIE database:
  - 400 training/500 testing
  - Expressions: smile, disgust, squint, surprise, scream



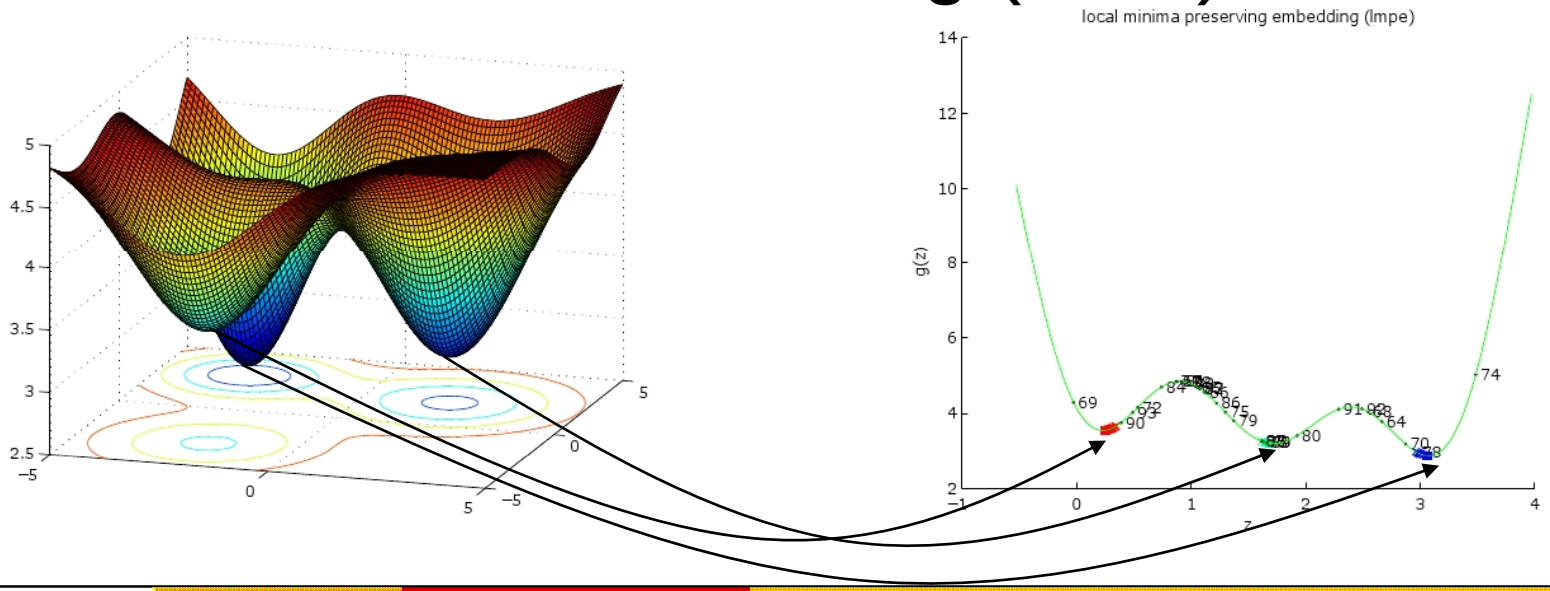


Conclusions

- Learning a metric for image alignment (**A**, **b** of the cost function) :
  - Template matching
  - Active Appearance Models (AAMs)
- Local minima only at the expected locations
- Convex quadratic formulation

Ongoing and future work

- Metric useful for:
  - Good regions/pixels to detect/track
  - More robust features (HOG, DAISY,…)
- Learn the functional form of the cost function
- Local Minima Embedding (LME)– ICML 2010



Outline

- Introduction
- Facial image analysis (FIA)
  - Applications
  - Facial feature detection and tracking
  - Supervised facial expression recognition
  - Unsupervised discovery of facial events
- Conclusions

Related work in facial expression analysis

- Emotion labels *(e.g. Padgett & Cottrell '96, Essa & Pentland '97, Cohen et al. '00, Yacoob & Black '97 , Chan et al. '04, Shan et al. '05, Wang & Yin '07, Yang et al. '08, Kumano et al. '09)*

- Action Units (Aus) labels *(e.g. Donato et al. '99, Tian et al. '01, Cohen et al. '03, Bazzo et al. '04, Bartlett et al. '06, Tong et al. '07, Zhang et al. '05, Pantic & Patras '06, Zhu et al '09, Simon et al. '10)*

Outer Brow Raiser **(AU2)**

Brow lowerer **(AU4)**

Upper lid raiser **(AU5)**

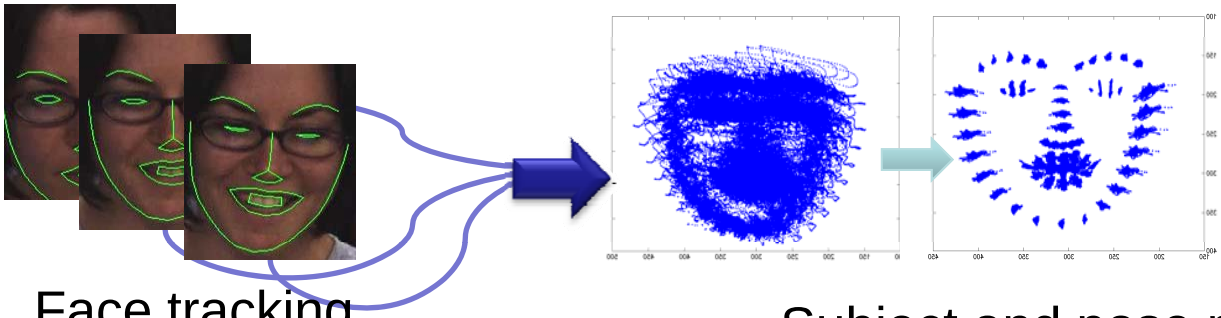
Lip corner depressor **(AU15)**

Chin raiser **(AU17)**

Lip stretcher **(AU20)**

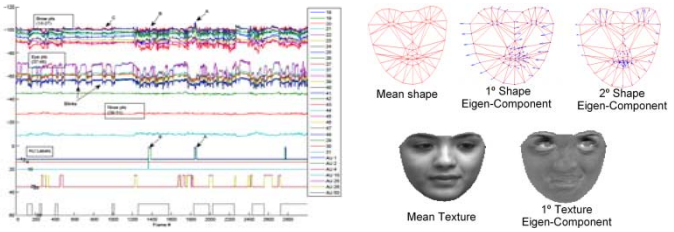
Facial expression recognition

- Most work differ in: registration, features and classifiers



Face tracking
(AAM)

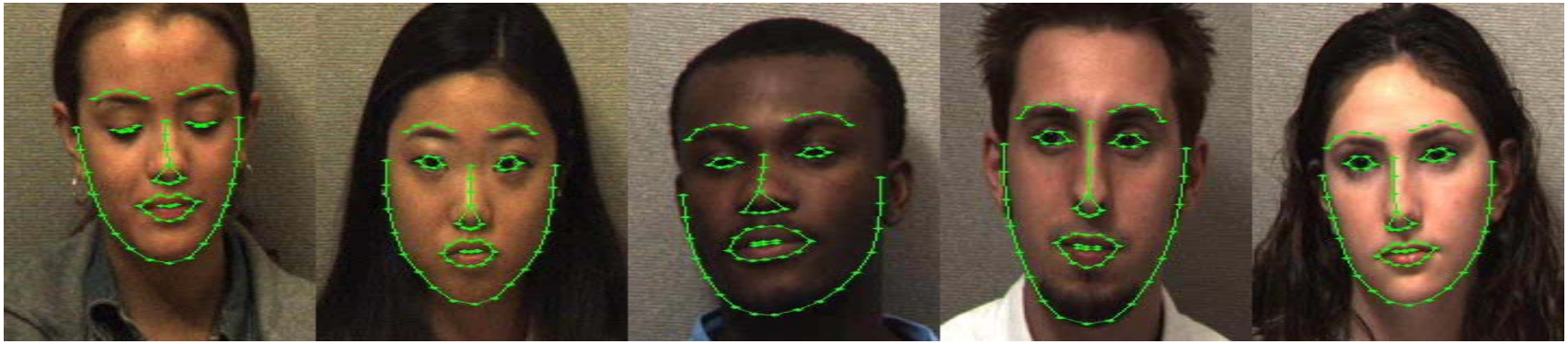
Subject and pose normalization
2D (affine, Euclidean)
3D (SFM)



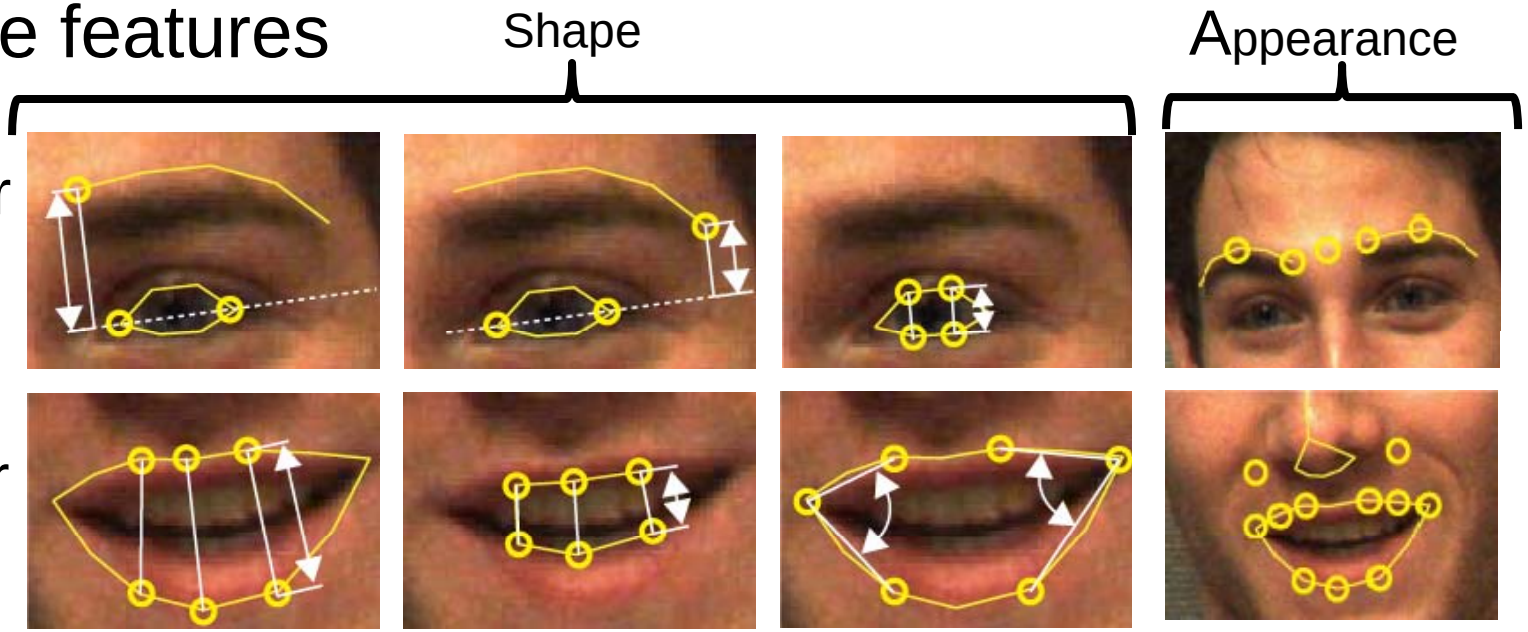
Feature extraction

Facial image features

- Active Appearance Models

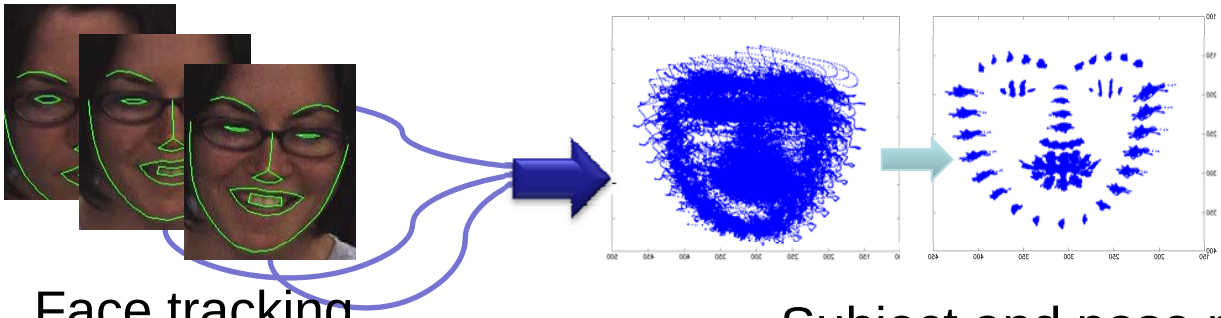


- Image features



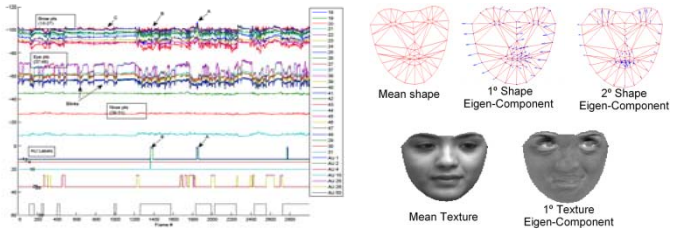
Facial expression recognition

- Most work differ in: registration, features and classifiers



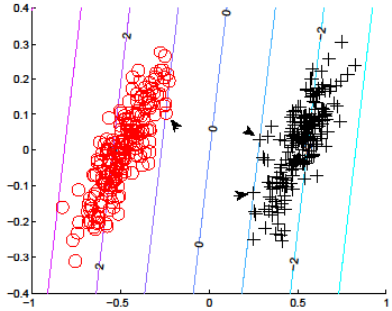
Face tracking (AAM)

Subject and pose normalization
2D (affine, Euclidean)
3D (SFM)



Feature extraction

AU present?



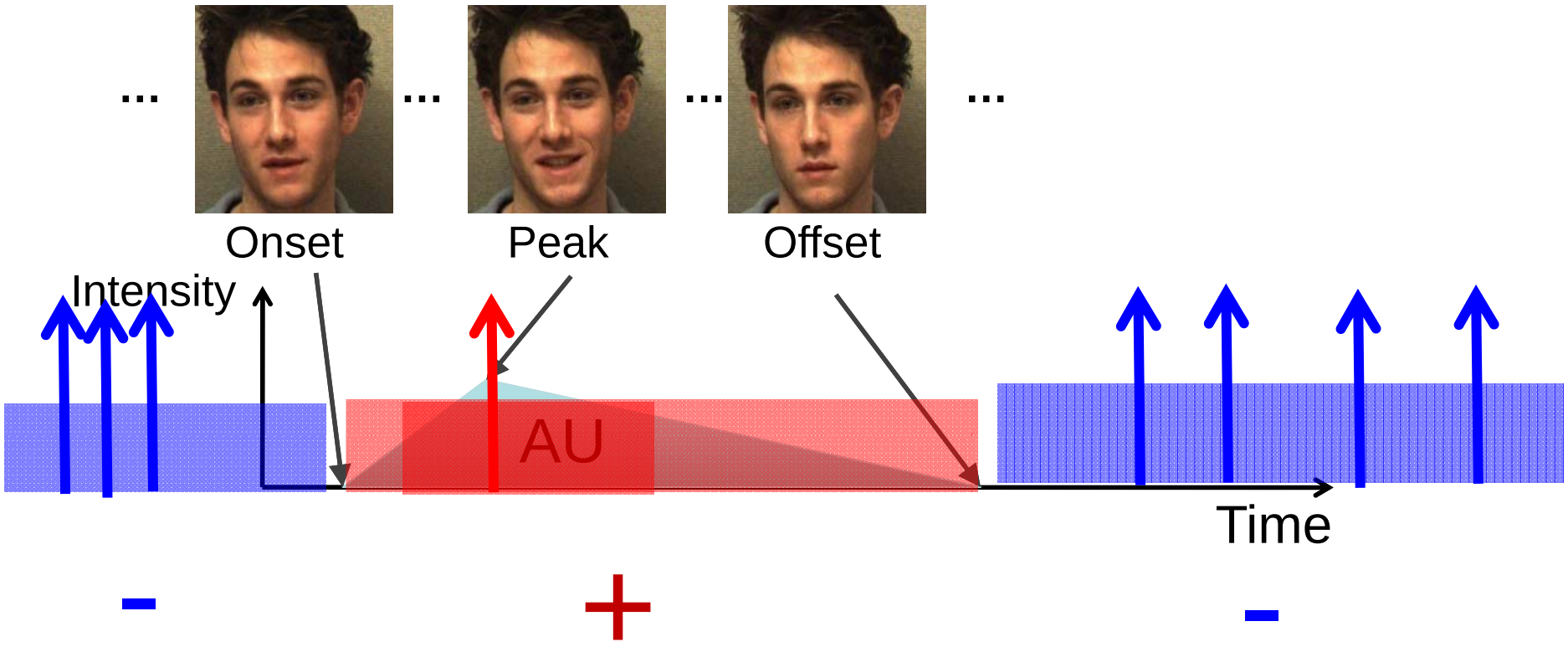
Classification

Classifiers

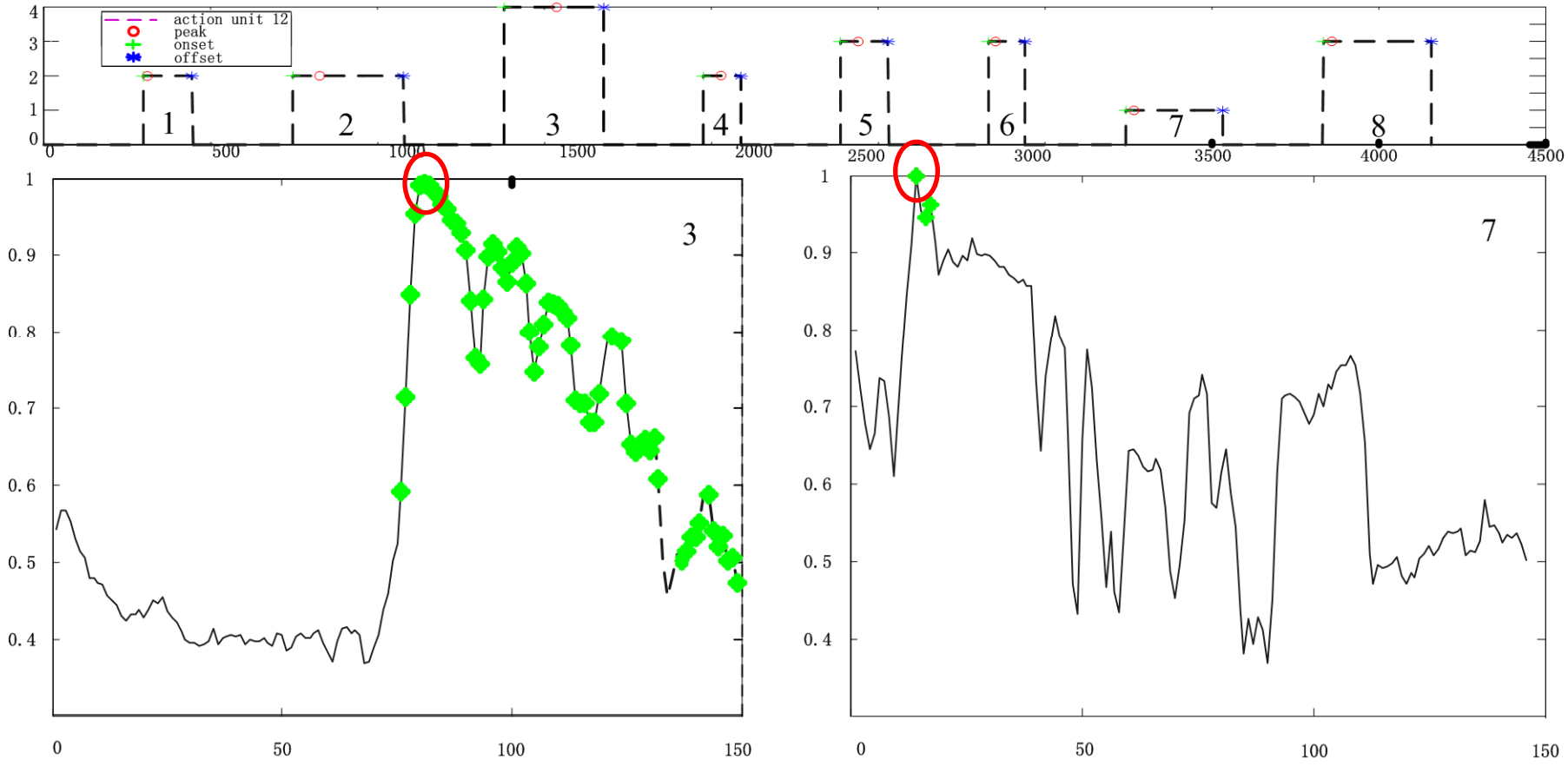
- Static (frame by frame)
  - Support Vector Machines (SVM)
  - Boosting
  - Gaussian Processes
  - Neural Networks
- Dynamic
  - Hidden Markov Models
  - Conditional Random Fields
  - Other extensions of dynamic Bayesian networks
  - Temporal bags of words

Facial expression recognition

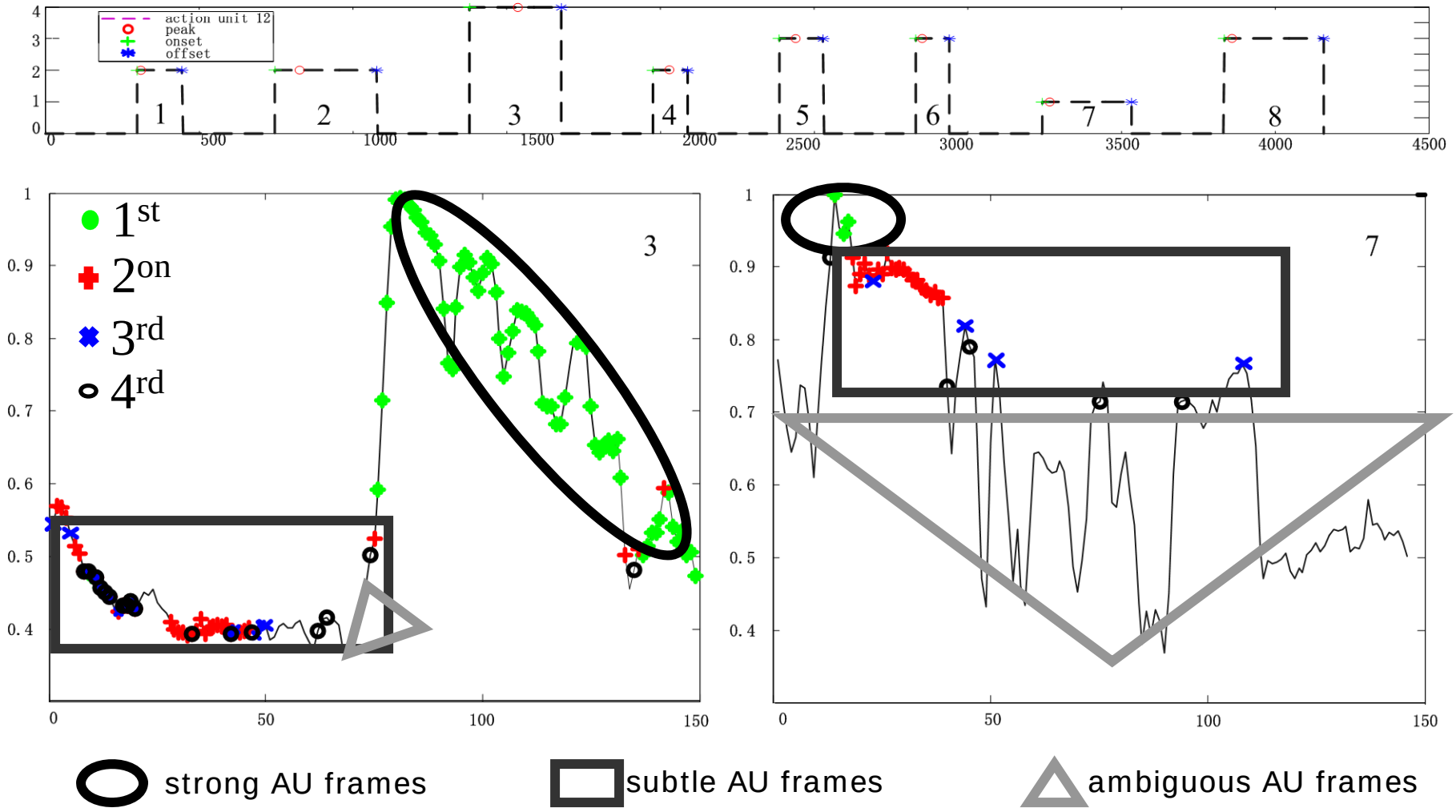
- Beyond registration, features and classifiers also make good use of the data!



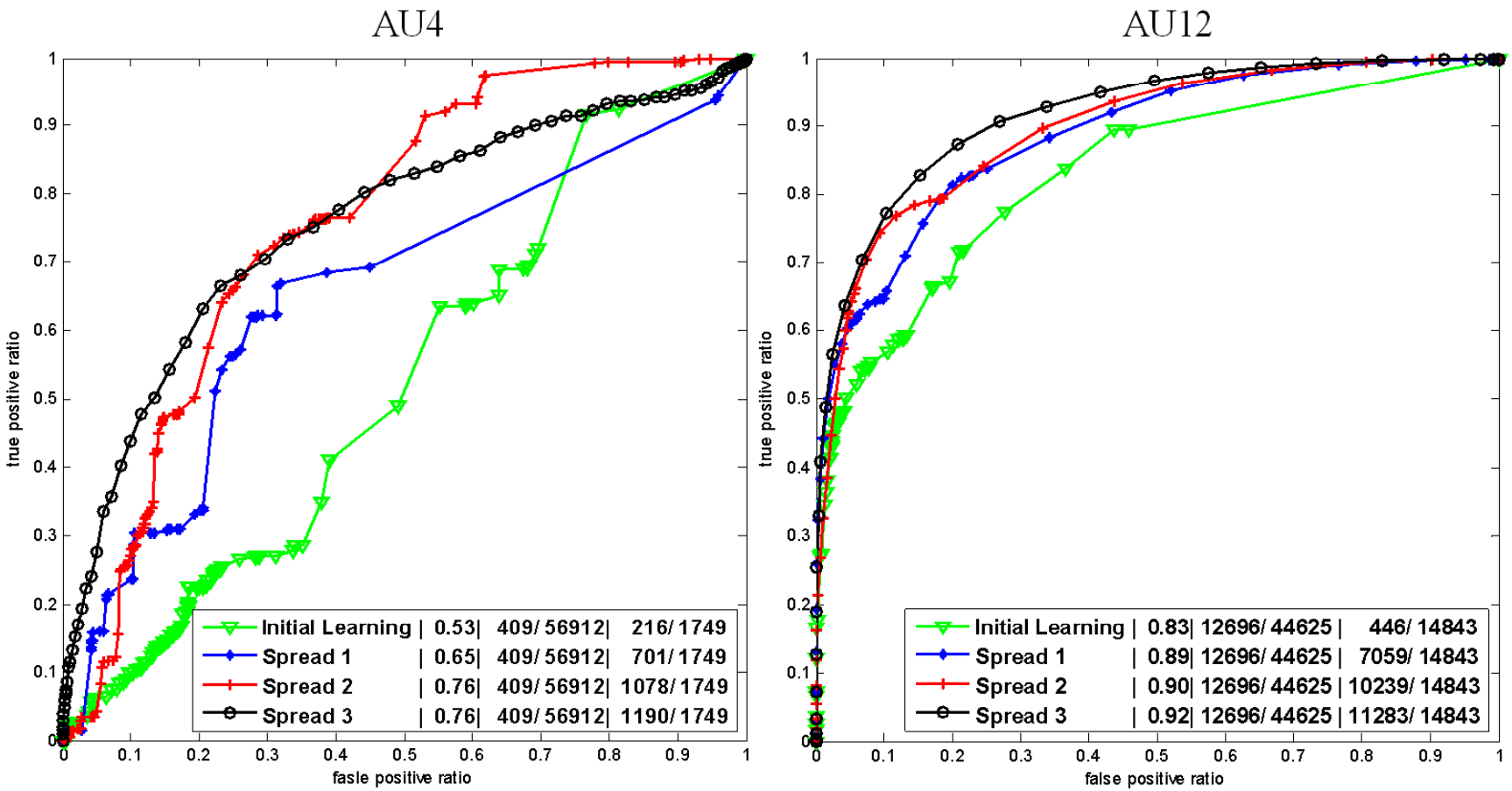
Dynamic Cascade with Bidirectional Bootstrapping (DCBB)



Dynamic Cascade with Bidirectional Bootstrapping (DCBB)



Results for AU4 and AU12

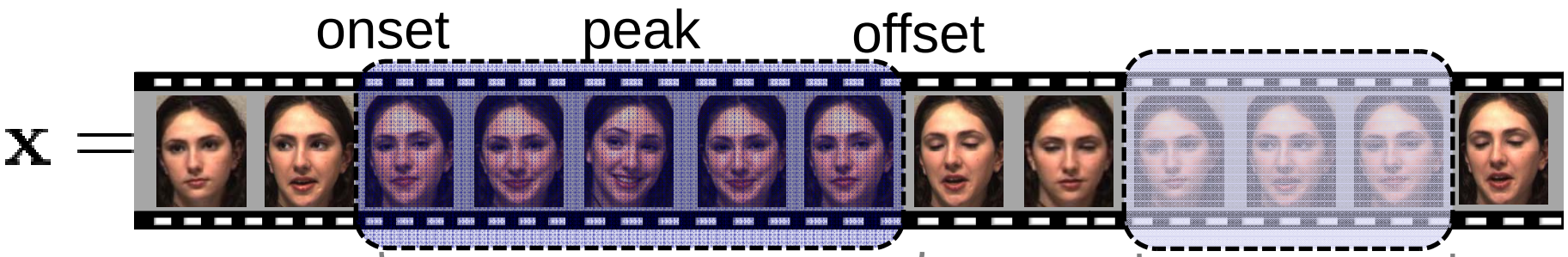


The first number between lines | denotes the area under the ROC, the second number is the size of positive samples in the testing dataset and separated by / is the size of negative samples in the testing dataset. The third number denotes the size of positive samples in training working sets and separated by / the total frames of target AU in training data sets.

Summary of results

	AU1	AU2	AU4	AU6	AU7	AU10	AU12	AU14	AU15	AU17
Peak+ Boost	0.75	0.71	0.53	0.93	0.56	0.52	0.83	0.50	0.52	0.59
All frames+ Boost	0.69	0.71	0.58	0.95	0.66	0.64	0.90	0.69	0.83	0.80
DCBB	**0.76**	**0.75**	**0.76**	**0.97**	**0.69**	**0.72**	**0.92**	**0.72**	**0.86**	**0.81**

Dynamic approach with bag of words



$$\varphi(\mathbf{x}, \mathbf{y}) =$$

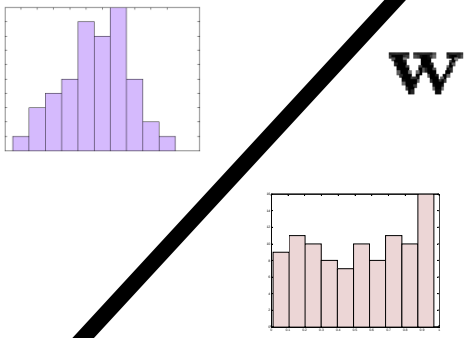
$$= \varphi(\mathbf{x}, \hat{\mathbf{y}})$$

$$\varphi(\mathbf{x}, [startframe, endframe])$$

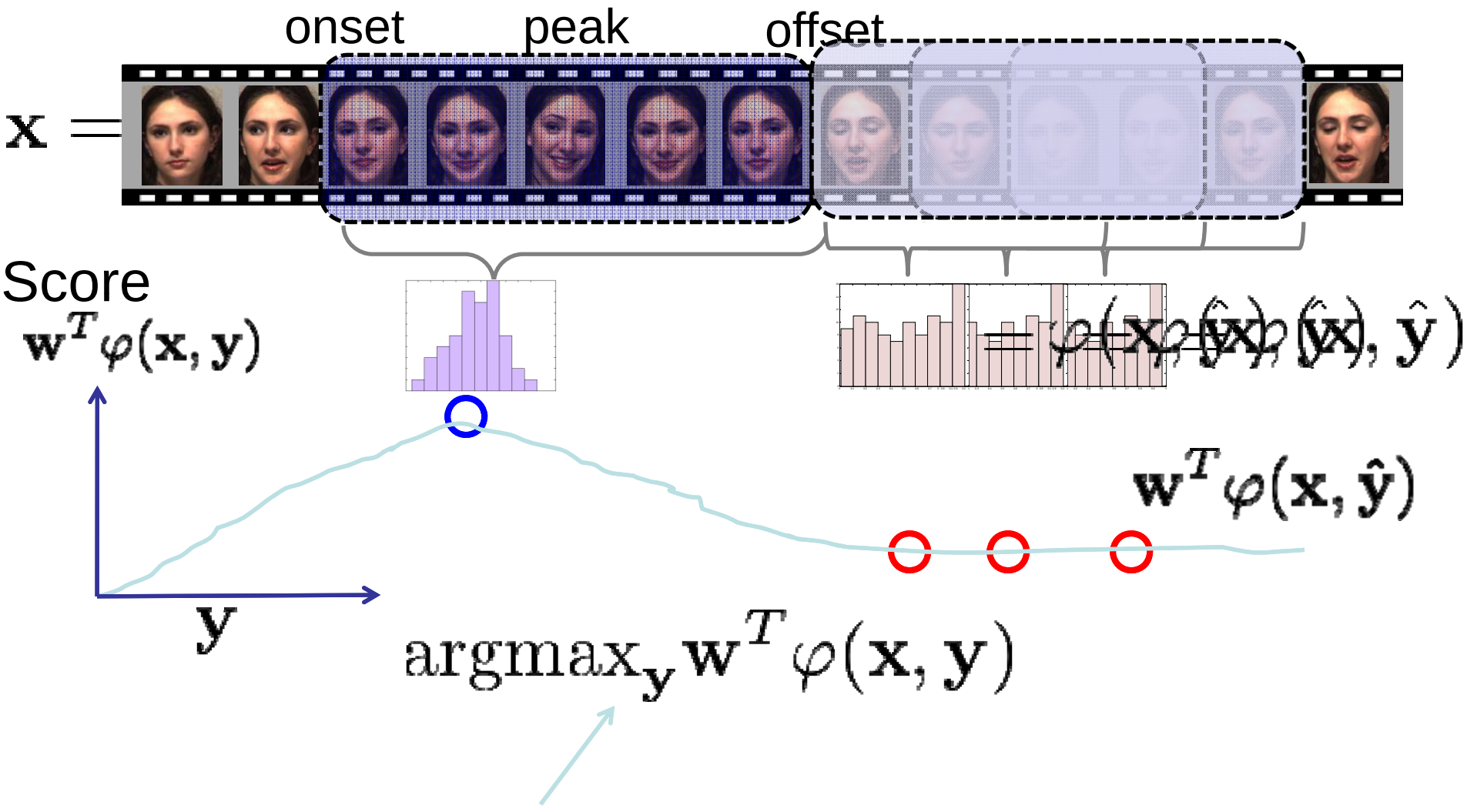
Bag of **temporal** words

Score function

$$\mathbf{w}^T \varphi(\mathbf{x}, \mathbf{y})$$

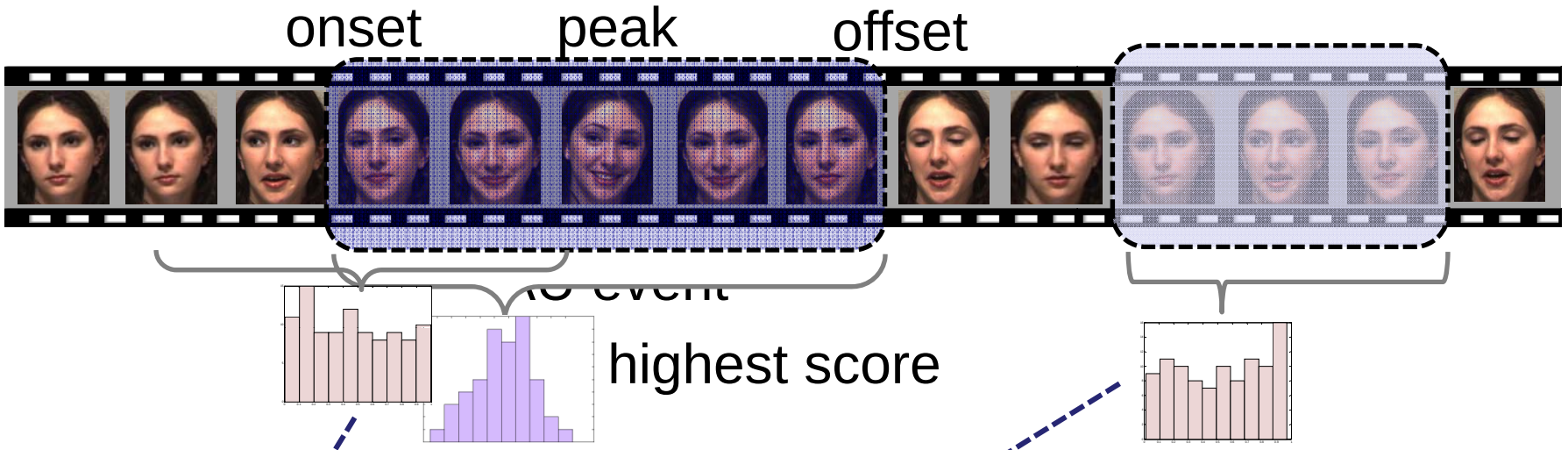


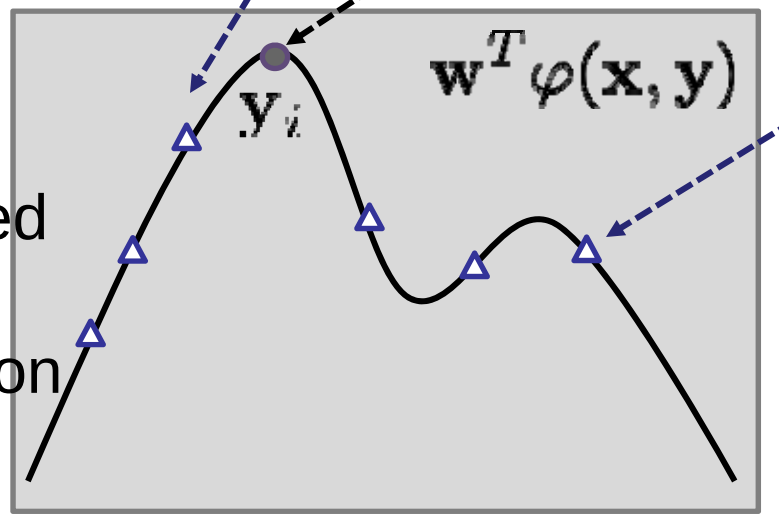
Testing



Efficient linear time algorithm

Training



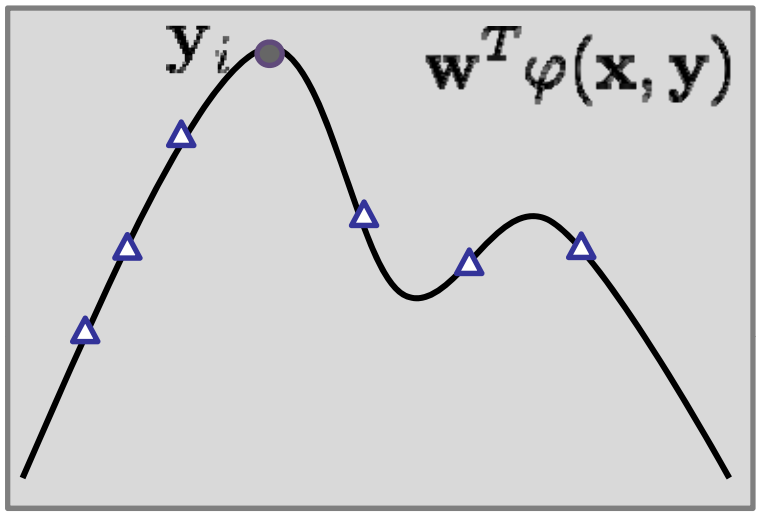


desired score function

$$\mathbf{w}^T\varphi(\mathbf{x},\mathbf{y}_i) > \mathbf{w}^T\varphi(\mathbf{x},\mathbf{y}) + \Delta(\mathbf{y}_i,\mathbf{y})$$

Structured output SVM

desired score function



$\mathbf{argmax}_{\mathbf{y}}\mathbf{w}^T\varphi(\mathbf{x},\mathbf{y})$

(Tsochantaridis '05)

All possible segments!

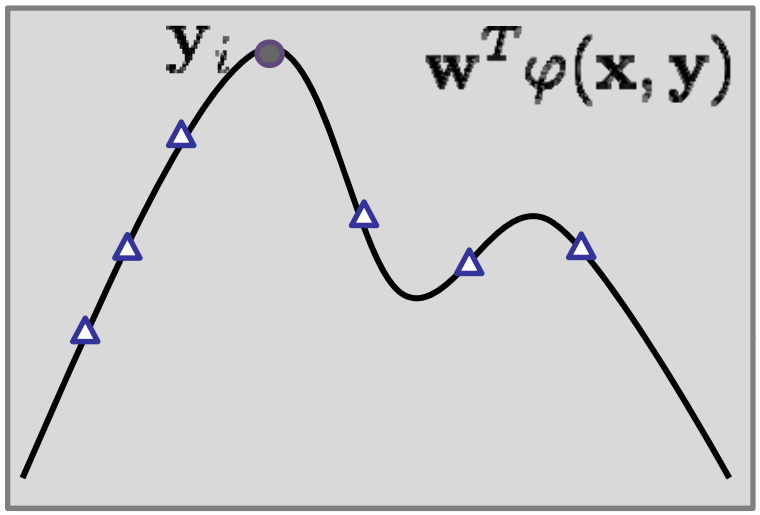
$$\min_{\mathbf{w},\boldsymbol{\xi}} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i,$$

$$\text{s.t. } \mathbf{w}^T\varphi(\mathbf{x}_i,\mathbf{y}_i) \geq \mathbf{w}^T\varphi(\mathbf{x}_i,\mathbf{y}) + \Delta(\mathbf{y}_i,\mathbf{y}) - \xi_i \;\forall \mathbf{y}$$

$$\xi_i \geq 0 \;\forall i.$$

Structured output SVM

desired score function



$\mathbf{argmax}_{\mathbf{y}}\mathbf{w}^T\varphi(\mathbf{x},\mathbf{y})$

(Tsochantaridis '05)

Constraint generation

$$\min_{\mathbf{w},\boldsymbol{\xi}} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i,$$

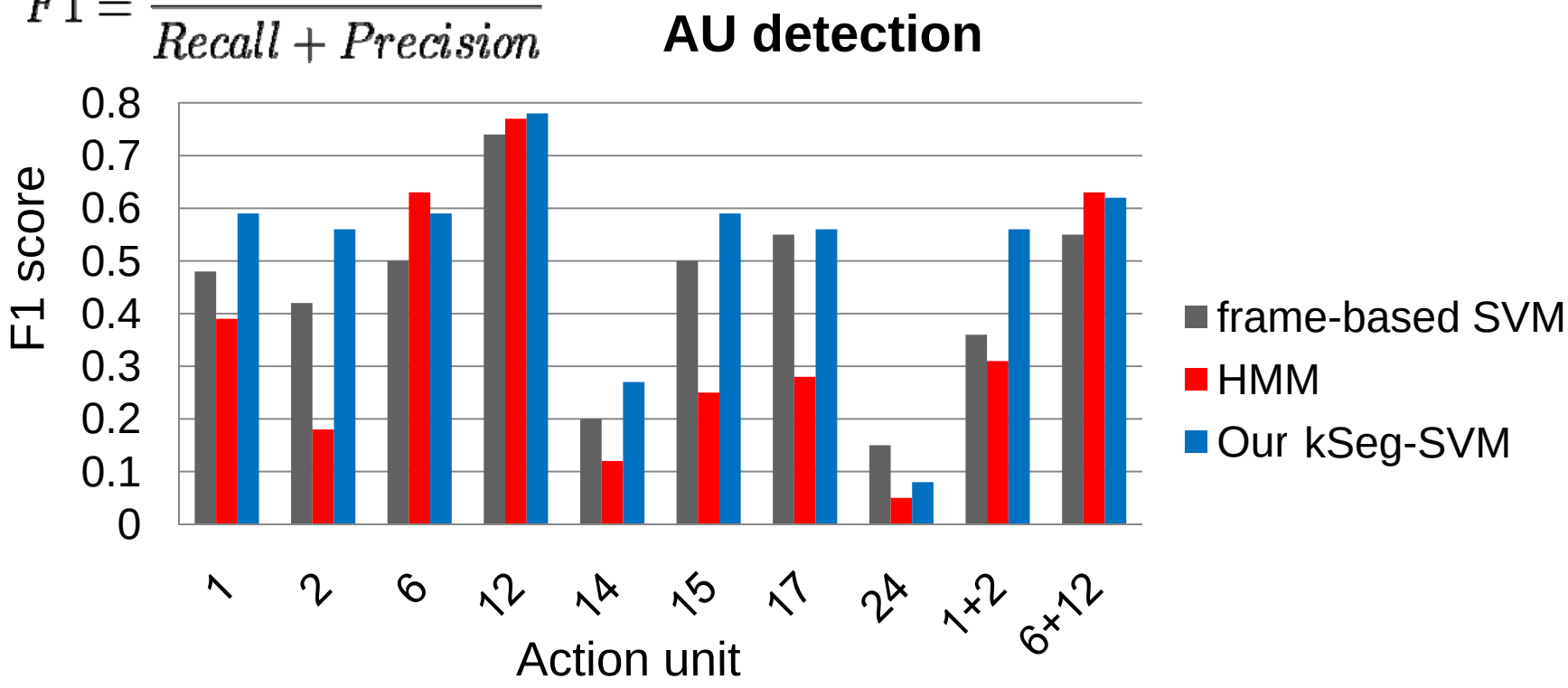
$$\text{s.t. } \max_{\mathbf{y}}\{\Delta(\mathbf{y}_i,\mathbf{y}) + \mathbf{w}^T\varphi(\mathbf{x}_i,\mathbf{y})\} \leq \mathbf{w}^T\varphi(\mathbf{x}_i,\mathbf{y}_i) + \xi_i$$

$$\xi_i \geq 0 \; \forall i.$$

Experiment – RU-FACS-1 dataset

- Interview videos of 28 subjects
- Video length: 2.5 minutes

$$F1 = \frac{2.Recall.Precision}{Recall + Precision}$$

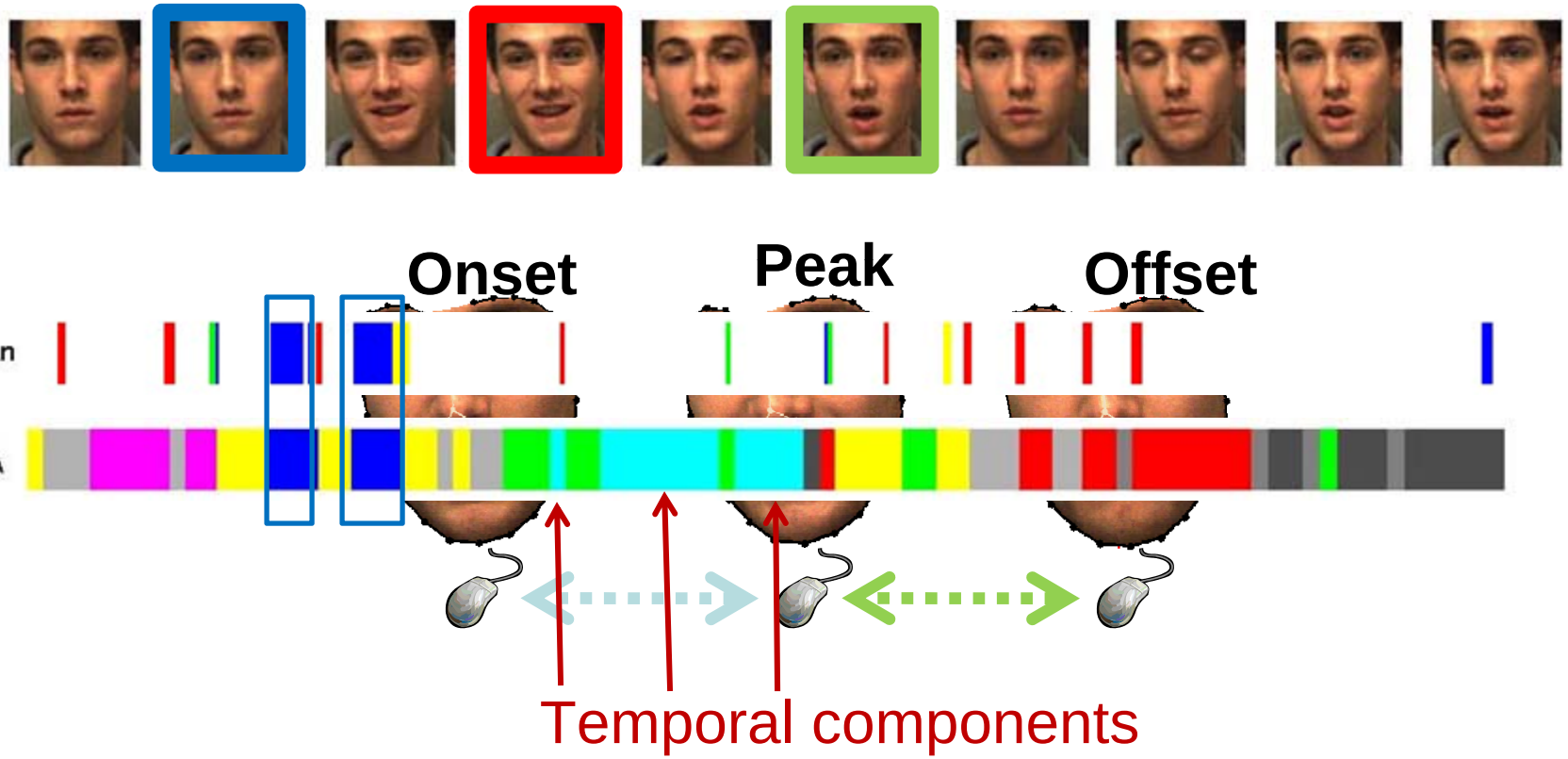


Outline

- Introduction
- Facial image analysis (FIA)
  - Applications
  - Facial feature detection and tracking
  - Supervised facial expression recognition
  - Unsupervised discovery of facial events
- Conclusions

Motivation I

- Improving speed and reliability of FACS coding

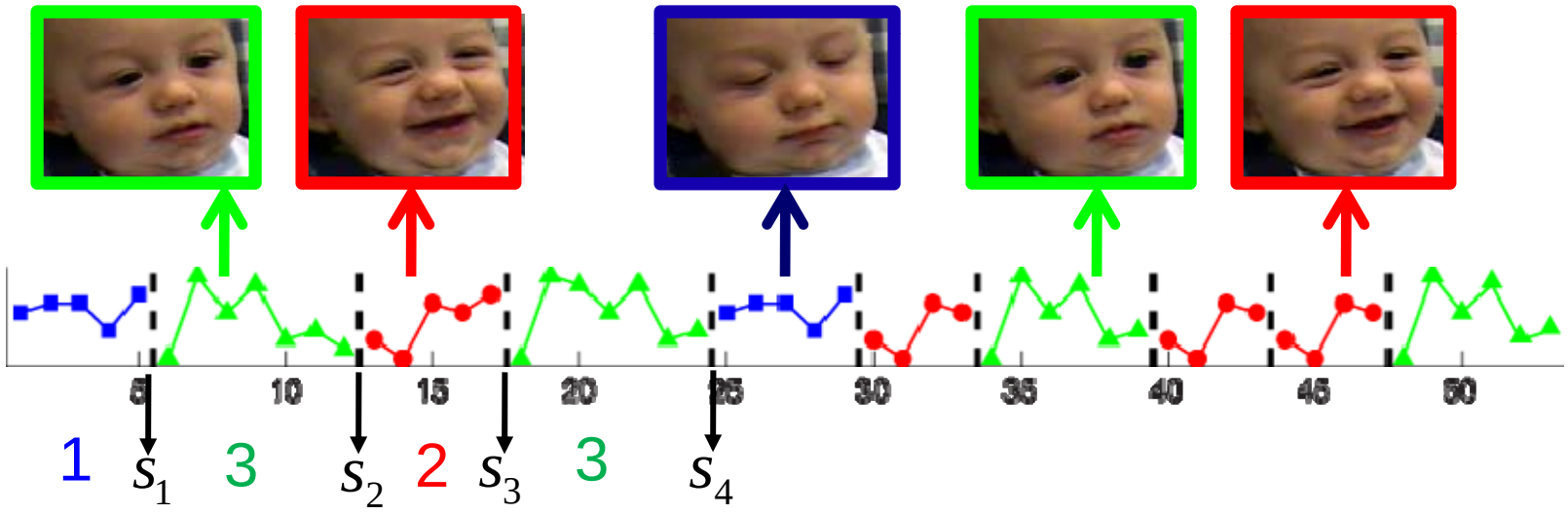


Motivation II

- Learning a taxonomy of facial expressions

Discovery facial events

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

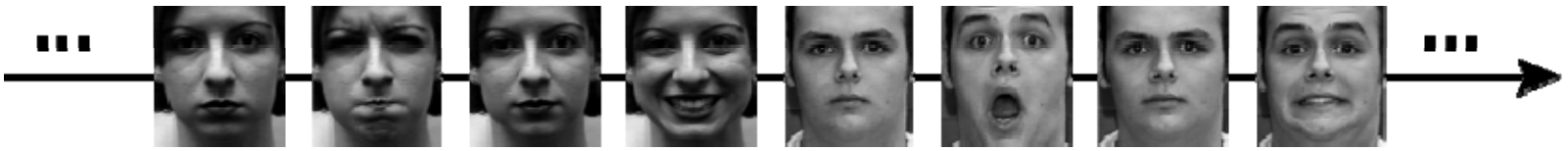


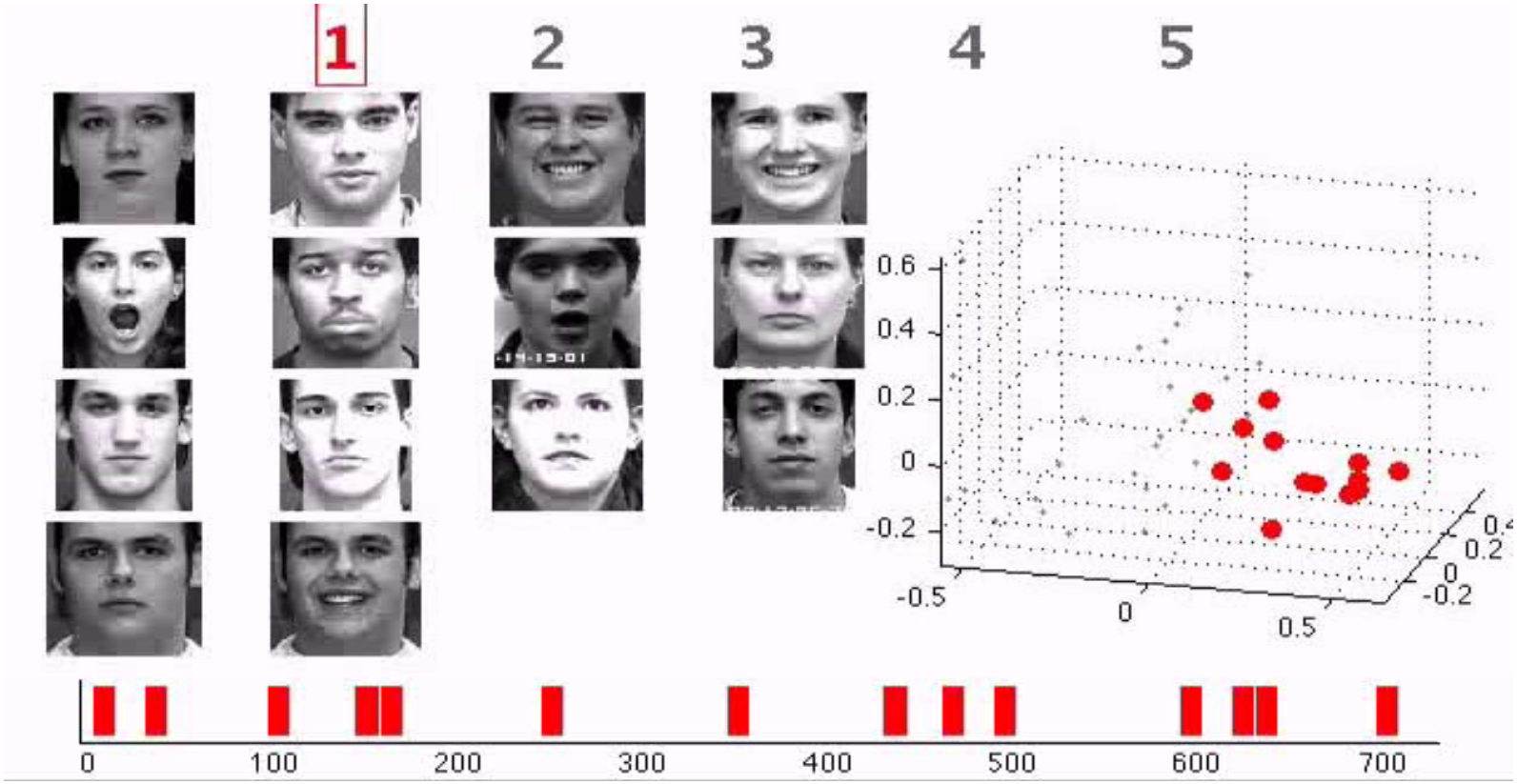
Related work in time series

- Change point detection *(e.g. Page '54, Stephens 94', Lai '95, Ge and Smyth '00, Steyvers & Brown '05, Murphy et al. '07, Harchaoui et al. '08)*
- Segmental HMMs *(e.g. Ge and Smith '00, Kohlmoren et al. '01, Ding & Fan '07)*
- Mixtures of HMMs *(e.g. Fine et al. '98, Murphy & Paskin '01, Oliver et al. '02, Alon et al. '03)*
- Switching LDS (*e.g. Pavolvic et al. '00*, *Oh et al. '08, Turaga et al. '09)*
- Hierarchical Dirichelet Process (*e.g. Beal et al. '02, Fox et al. '08)*
- **Aligned Cluster Analysis (ACA)**

Facial event discovery across subjects

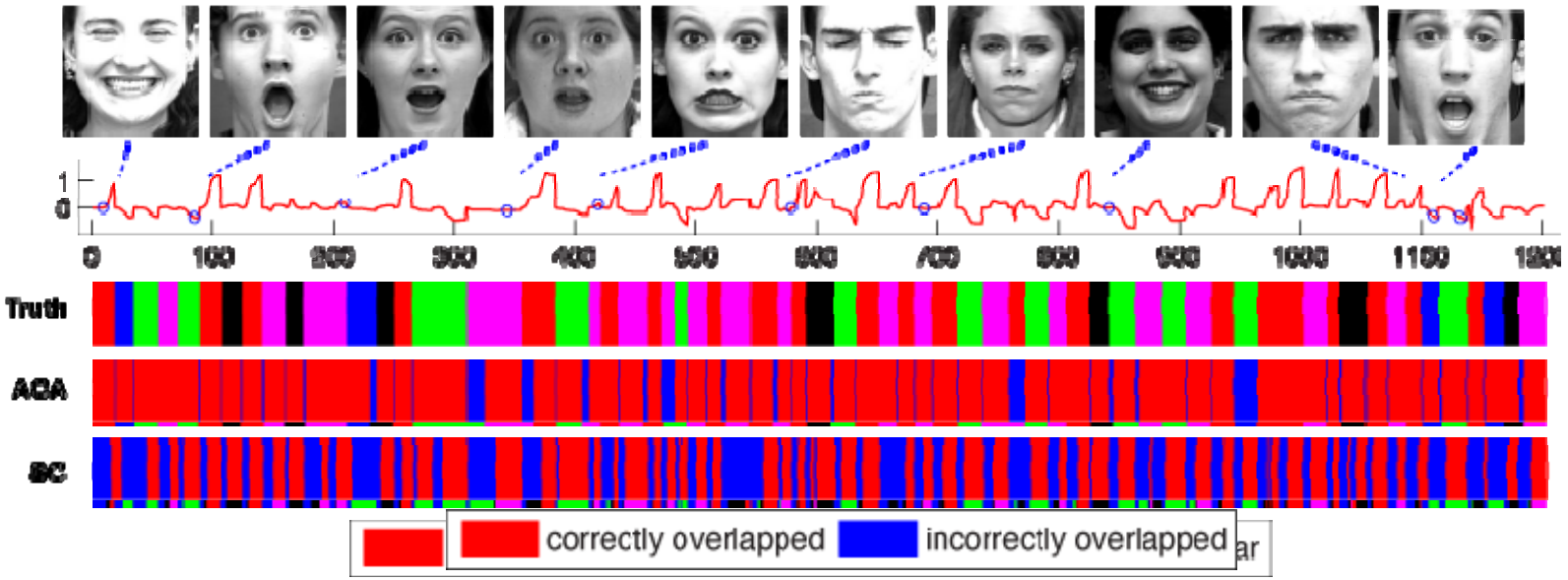
- Cohn-Kanade: 30 people and five different expressions (surprise, joy, sadness, fear, anger)





Facial event discovery across subjects

- Cohn-Kanade: 30 people and five different expressions (surprise, joy, sadness, fear, anger)



- 10 sets of 30 people

ACA	Spectral Clustering (SC)
0.87(.05)	0.56(.04)

Unsupervised facial event discovery

Conclusions and future work

- Applications for FIA
- Algorithms for:
  - Facial feature detection/tracking
  - Facial expression recognition
  - Unsupervised facial expression
- Future work:
  - Diverse applications in psychology, marketing, surveillance or personal use
  - Generic models, strong pose changes and subtle facial behavior

Takeo Kanade

Jeff Cohn

Howard Wactlar

Rebecca Thurston

Thanks

Questions?

Minh Nguyen

Feng Zhou

Human Sensing Lab

http://humansensing.cs.cmu.edu

Component Analysis Lab

http://ca.cs.cmu.edu