Surveillance and aggression detection using multimodal cameras

Delft University of Technology
Netherlands Defence University

Prof.drs.dr. Léon Rothkrantz
Ir. Dragos Datcu
Ir Zhenke Yang
Ir. Ramon Dor

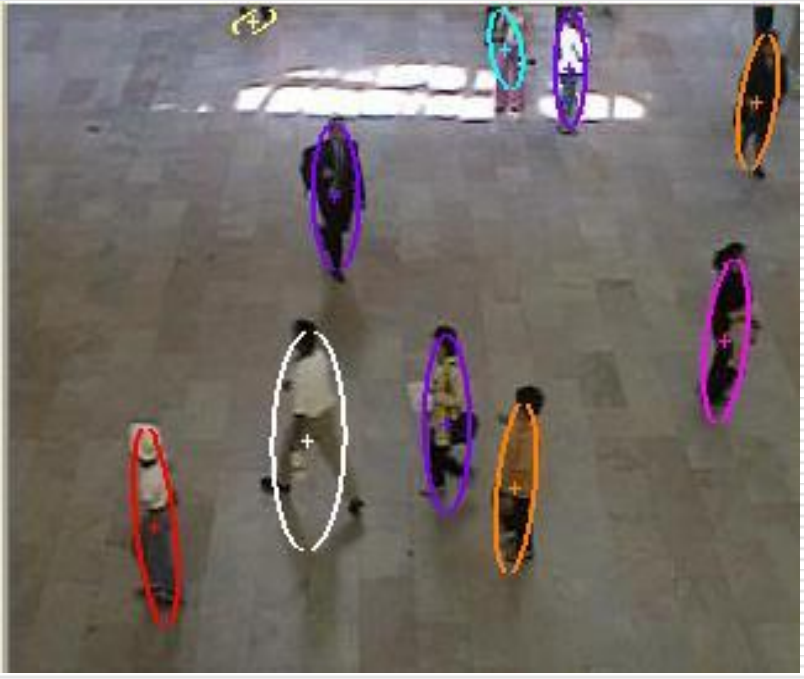
December 2009 San Sebastiaan

TU Delft

Outline presentation

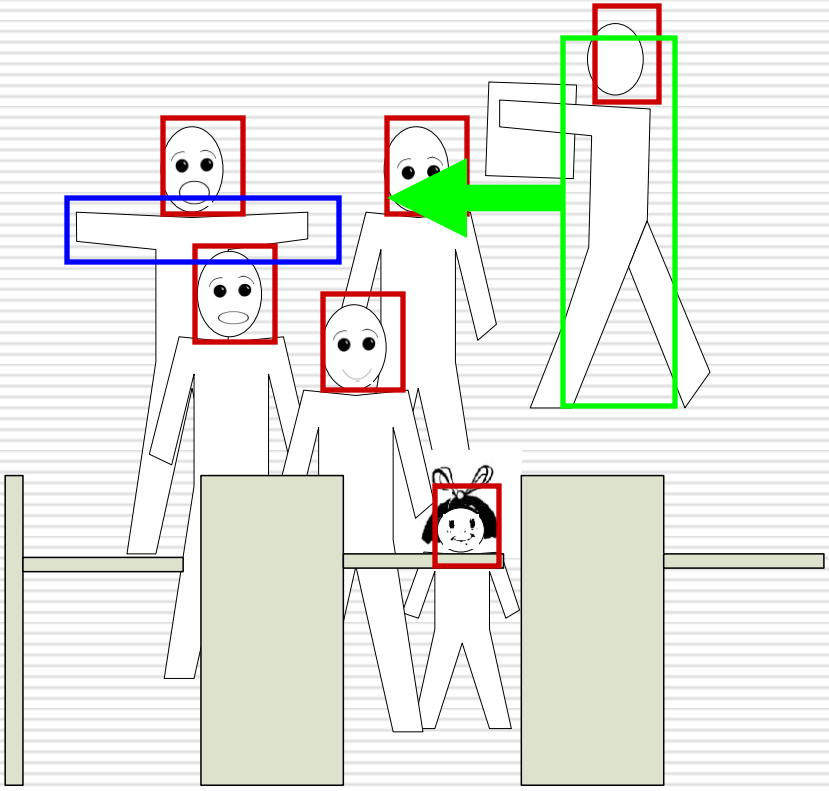
- Problem definition, visual speech, nonverbal communication, aggression
- Model of visual speech
- Experiments, in train environment
- Conclusion

Sensing, localizing, tracking, classification, identification



Surveillance multimodal cameras

Video



Face/facial expression recognition

Emotion assessment

Gesture recognition

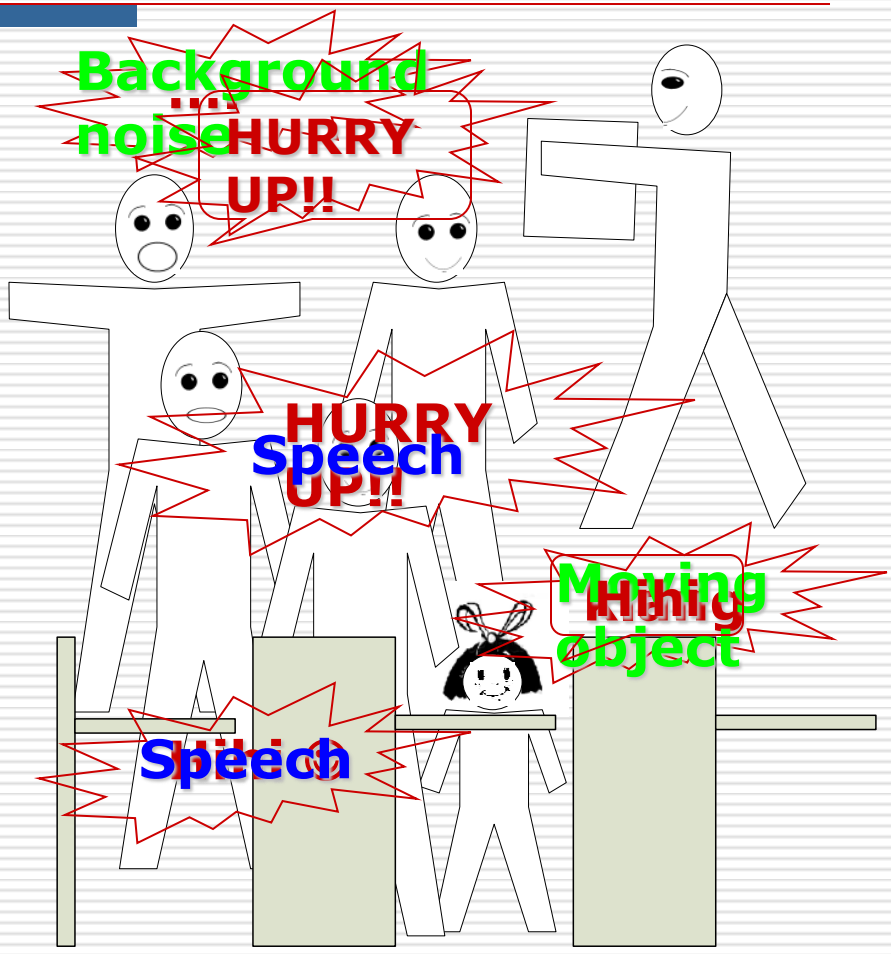
Motion tracking

Audio

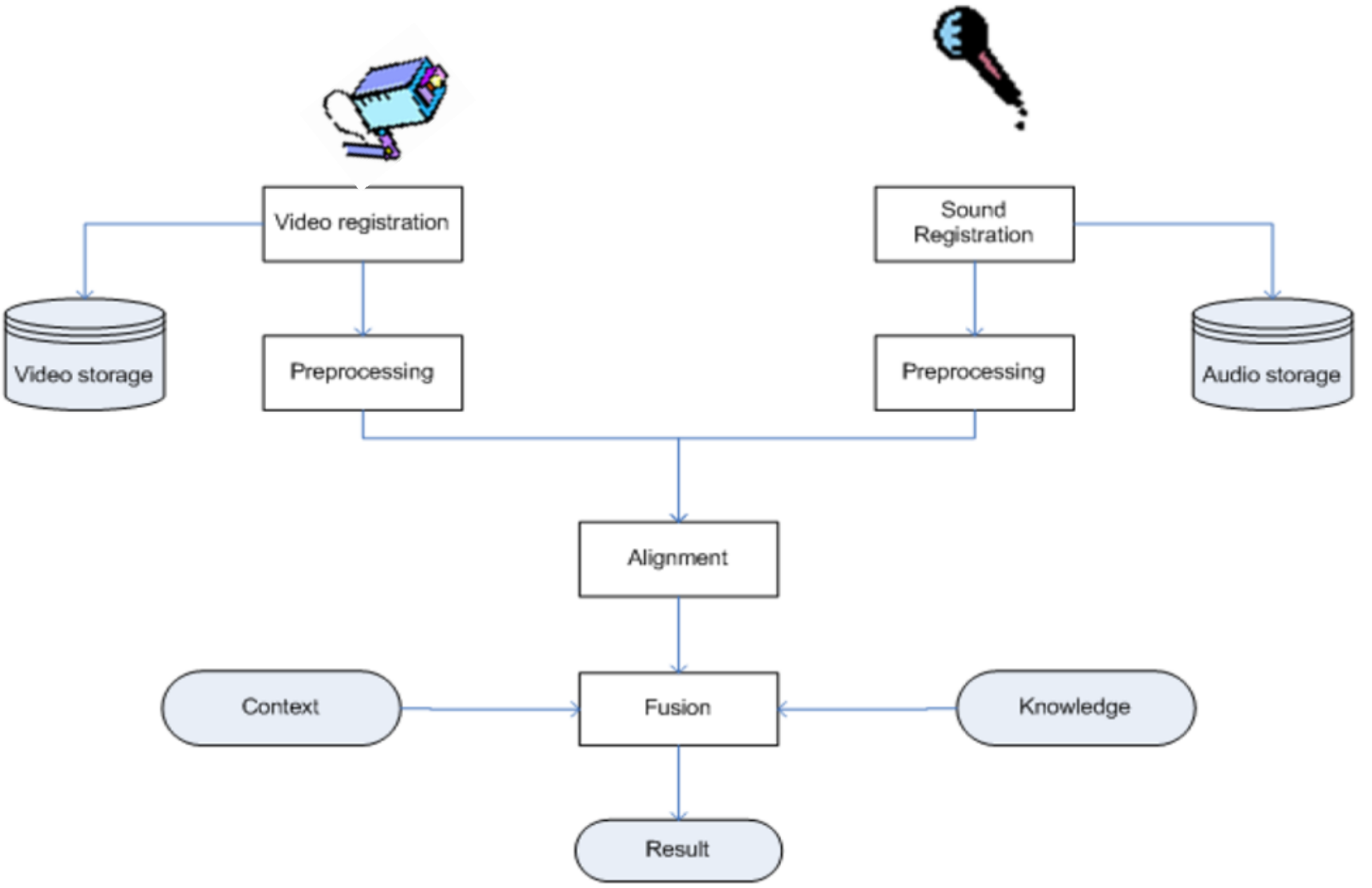
Sound classification

Source localization

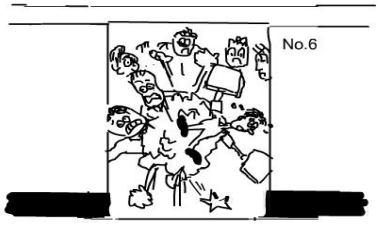
Stress/aggression detection

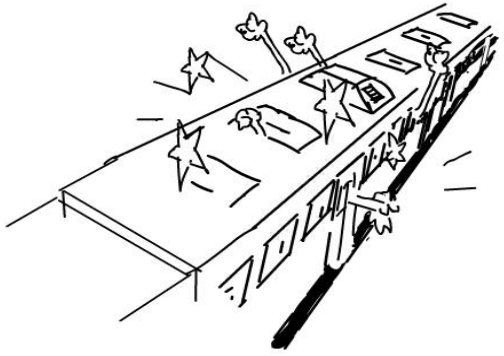


MULTIMODAL SURVEILLANCE SYSTEM



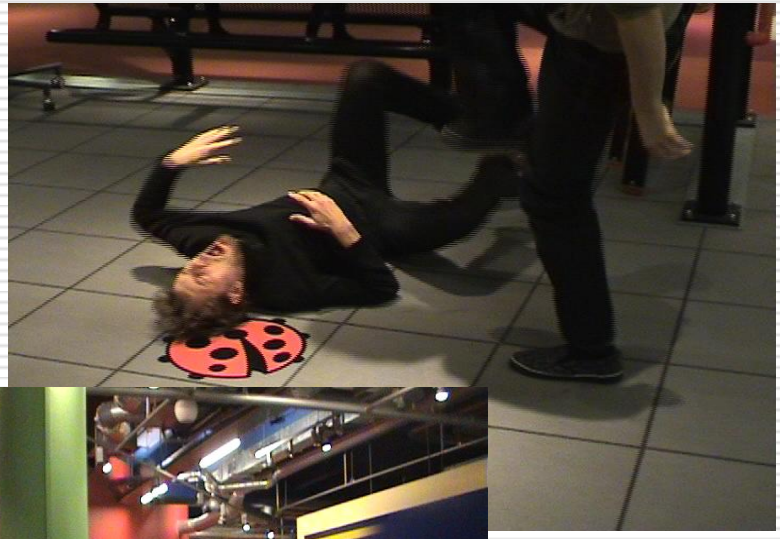
Finding or fighting a place



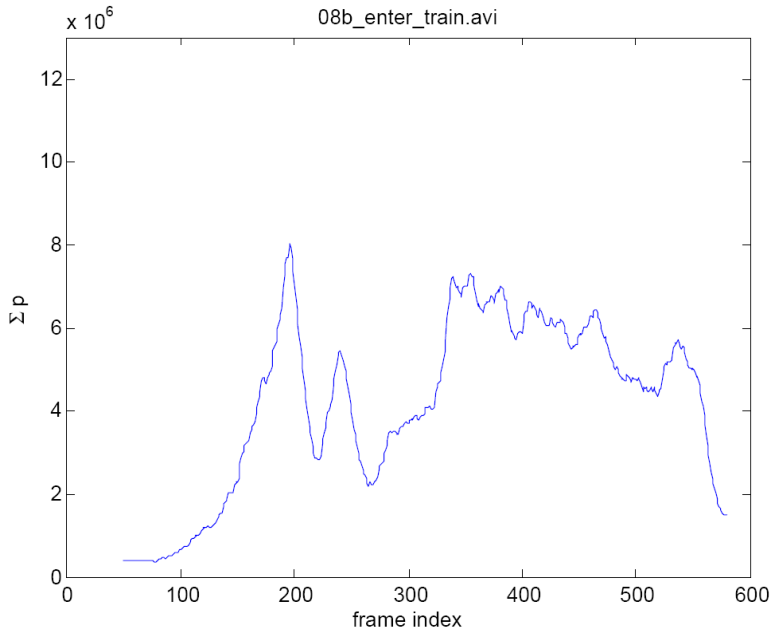


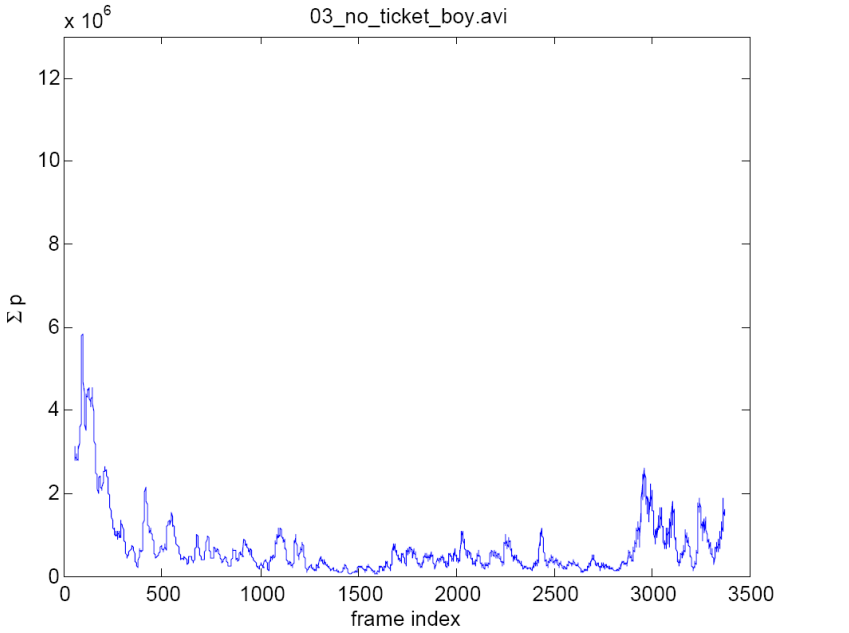
Intrusion at the last moment

Aggression at a railwaystation



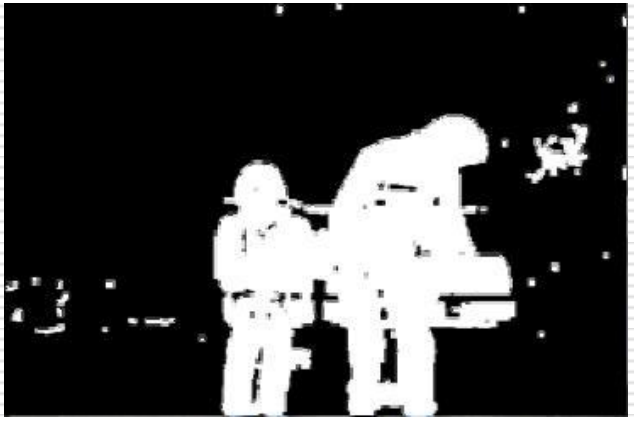
Energy graphs fighting people-laying on the ground



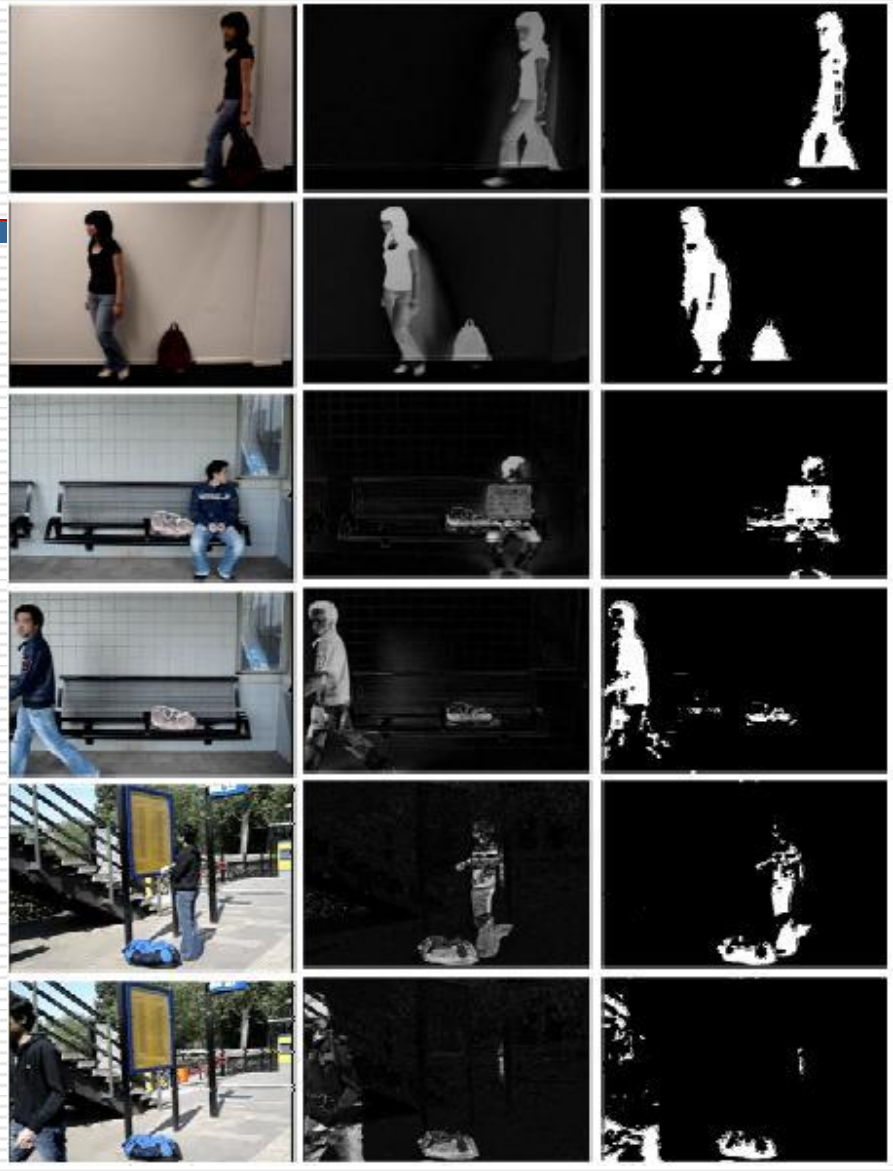


TU Delft

Background subtraction



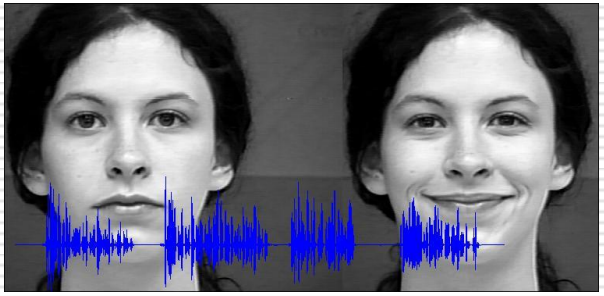
Segmentation of human bodies and luggage



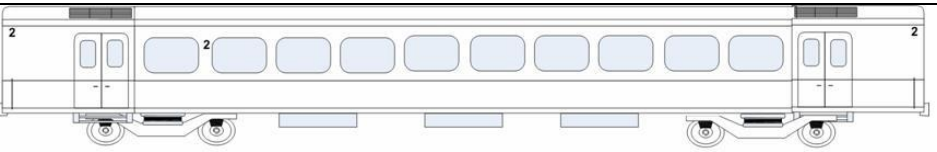
What makes visual speech recognition so hard?

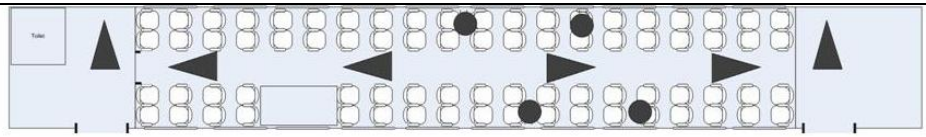
- Visemes ➔ Smaller word separability
- Speech info in audio > Speech info in video

Application : Surveillance in trains



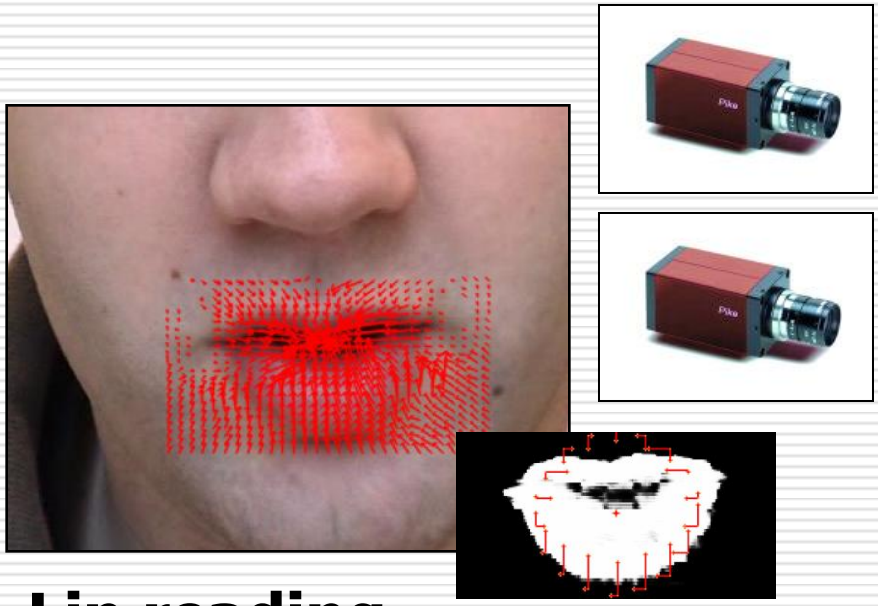
Video, Audio processing
Sound localization, pattern rec.



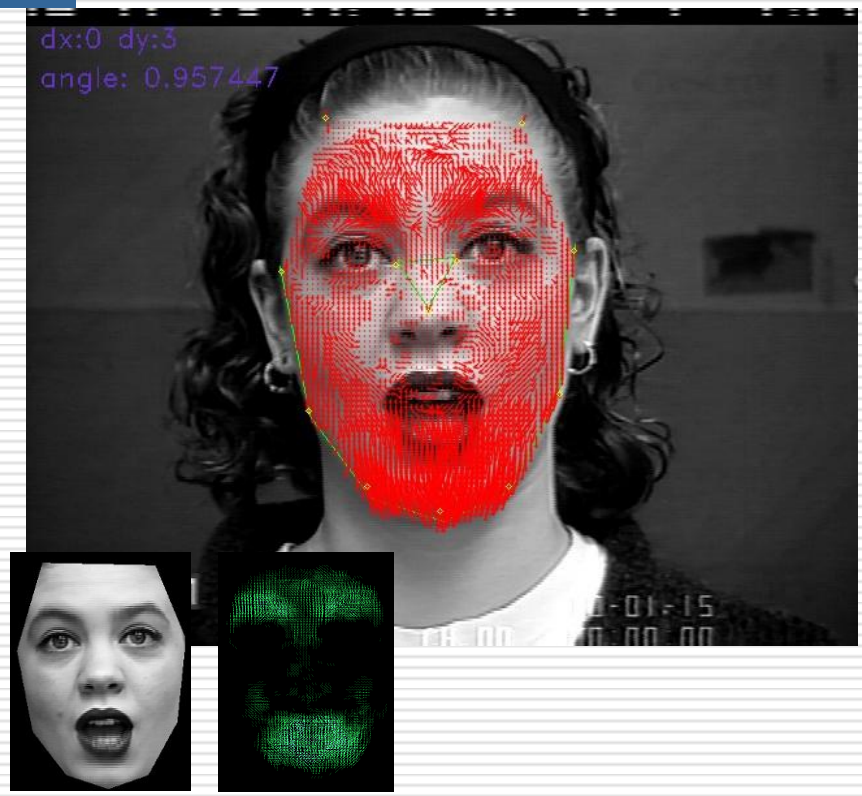


TU Delft

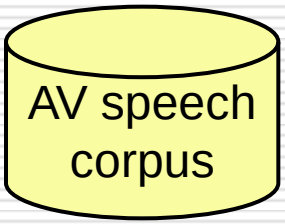
Automatic recognition of facial expressions and lipreading using vector flow



Lip reading



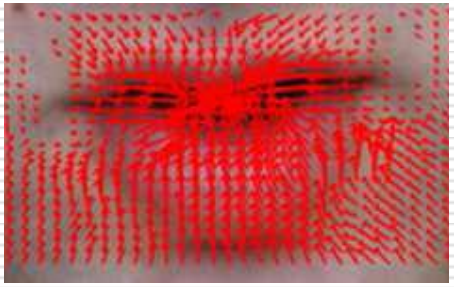
Facial expression recognition

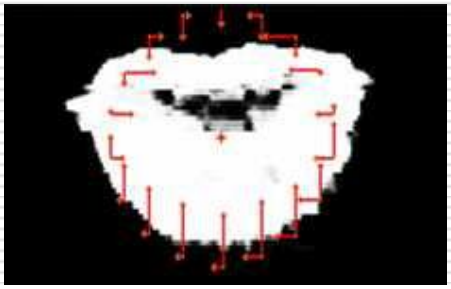


New speech corpus

Tracking the face – Optical flow

- Capturing apparent motion of subsequent images in a grid of motion vectors

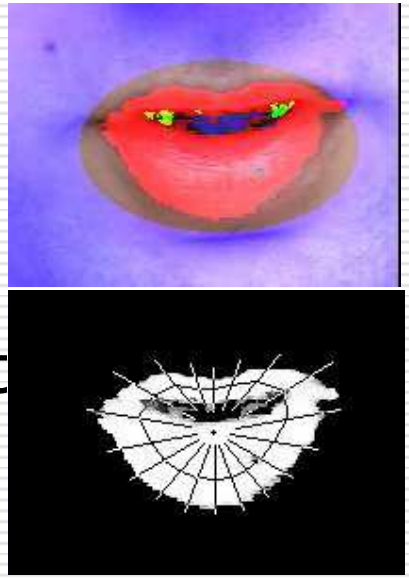




- Advantages
  - No lip model required
  - Good at capturing motion
- Disadvantage
  - Slow

Tracking the face – Lip Geometry Estimation

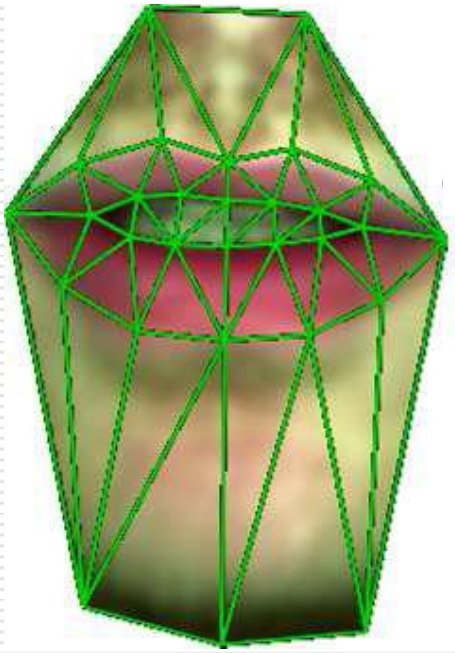
- Applying some color filters and capturing the lip contours in polar coordinates



- Advantages
  - No lip model required
  - More or less person-independent
- Disadvantage
  - Not robust to external factors

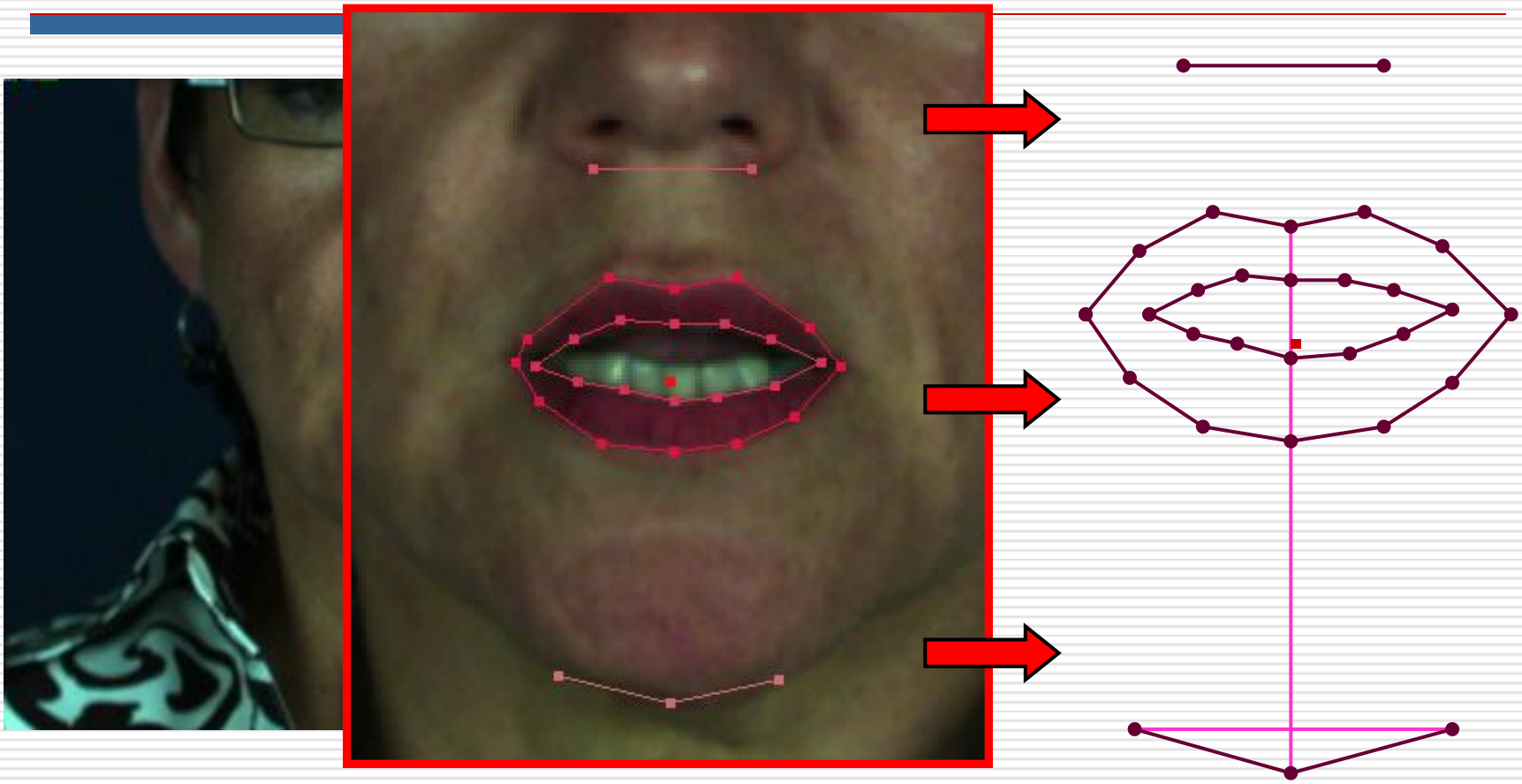
Tracking the face – Active Appearance Models

- Point tracking according to a statistical lip model



- Disadvantage
  - Requires annotated training
  - images
- Advantages
  - Robust against external
  - factors
  - Fast!

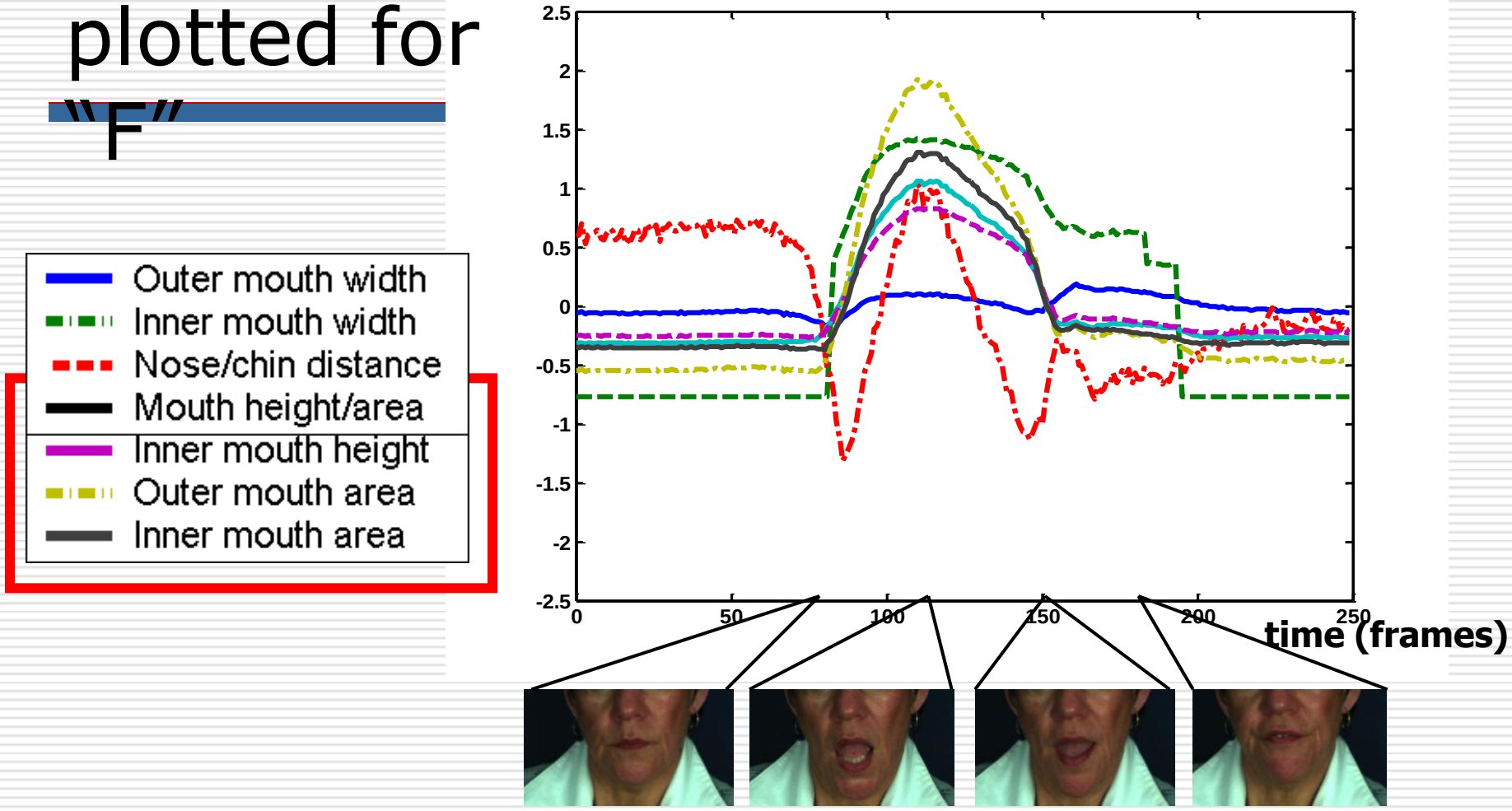
Active Appearance Models – Defining the lip model



AAM model point coordinates

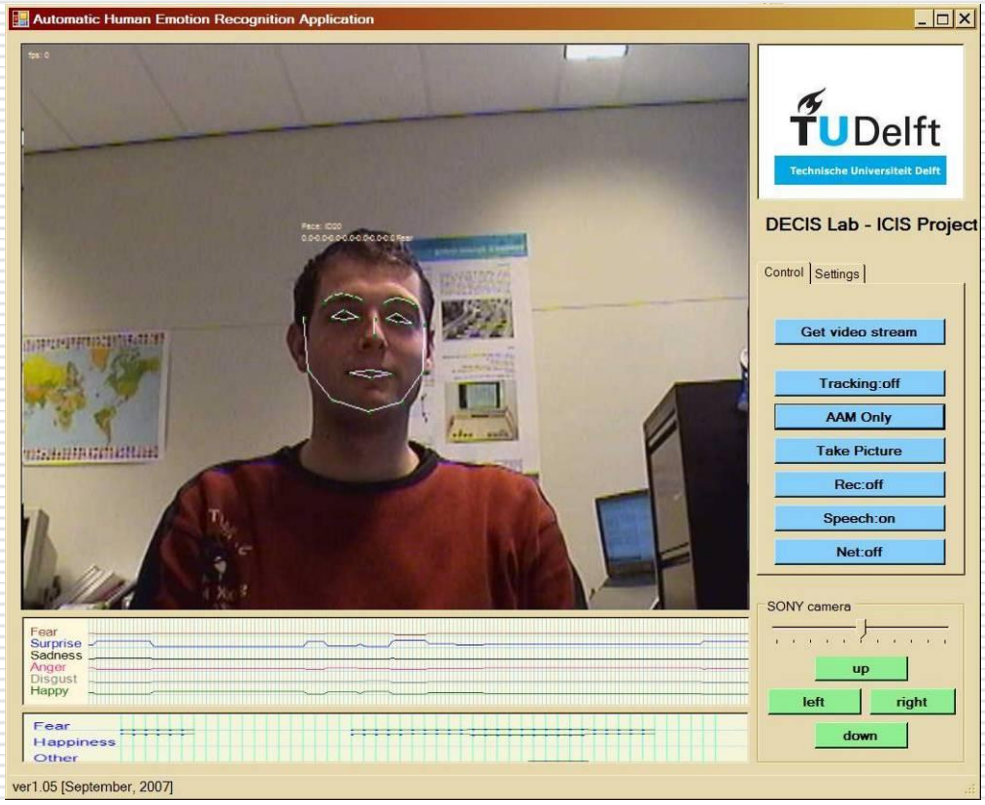
Features plotted for "F"

Feature extraction



TU Delft

Automatic recognition of facial expressions using active Appearance model



Automatic bi-modal human emotion recognition


9962 20-416 3:Camera 3 10-mei-2006 11:41:58 AM CEST
9962 20-416 4:Camera 4 10-mei-2006 11:41:58 AM CEST
9962 20-416 5:Camera 5 10-mei-2006 11:41:58 AM CEST
9962 20-416 6:Camera 6 10-mei-2006 11:41:58 AM CEST

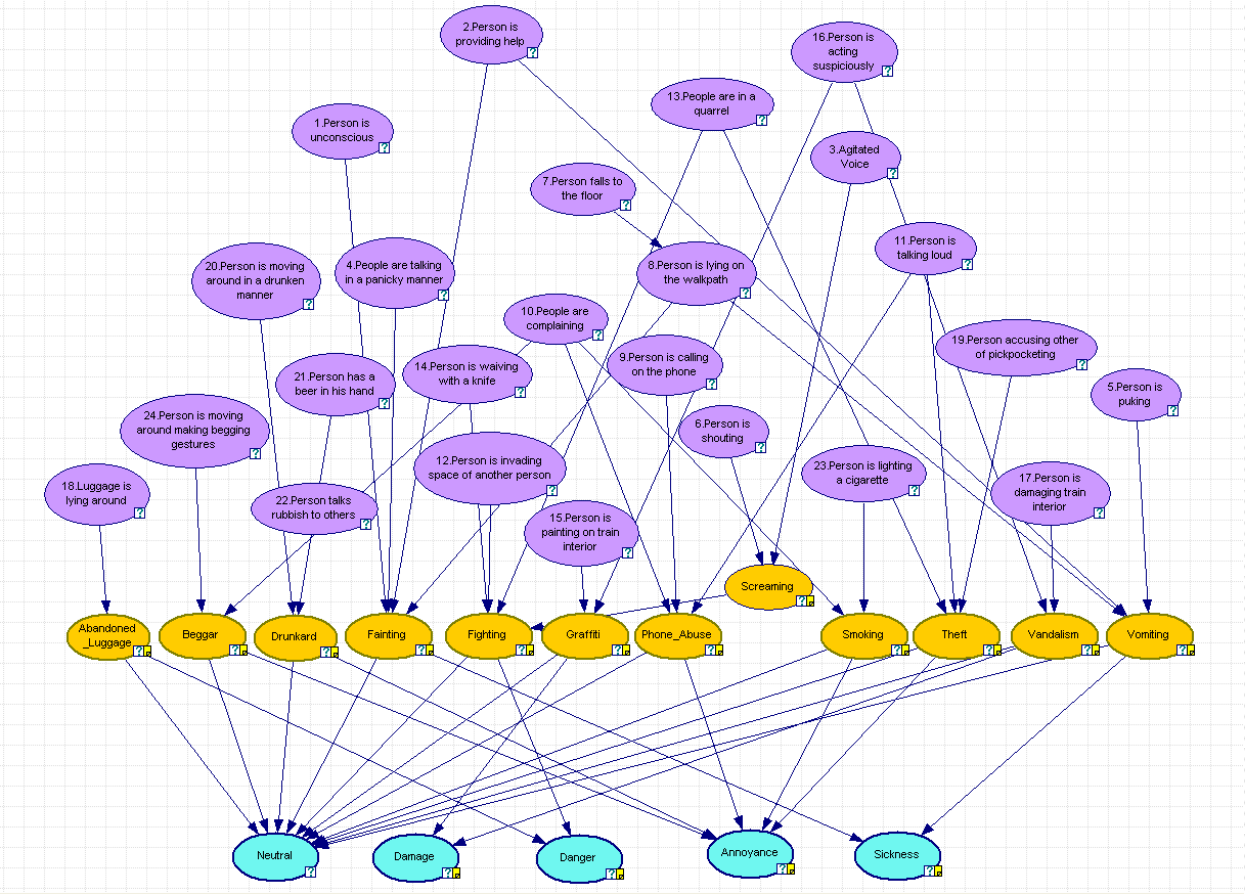
TU Delft

Face localisation

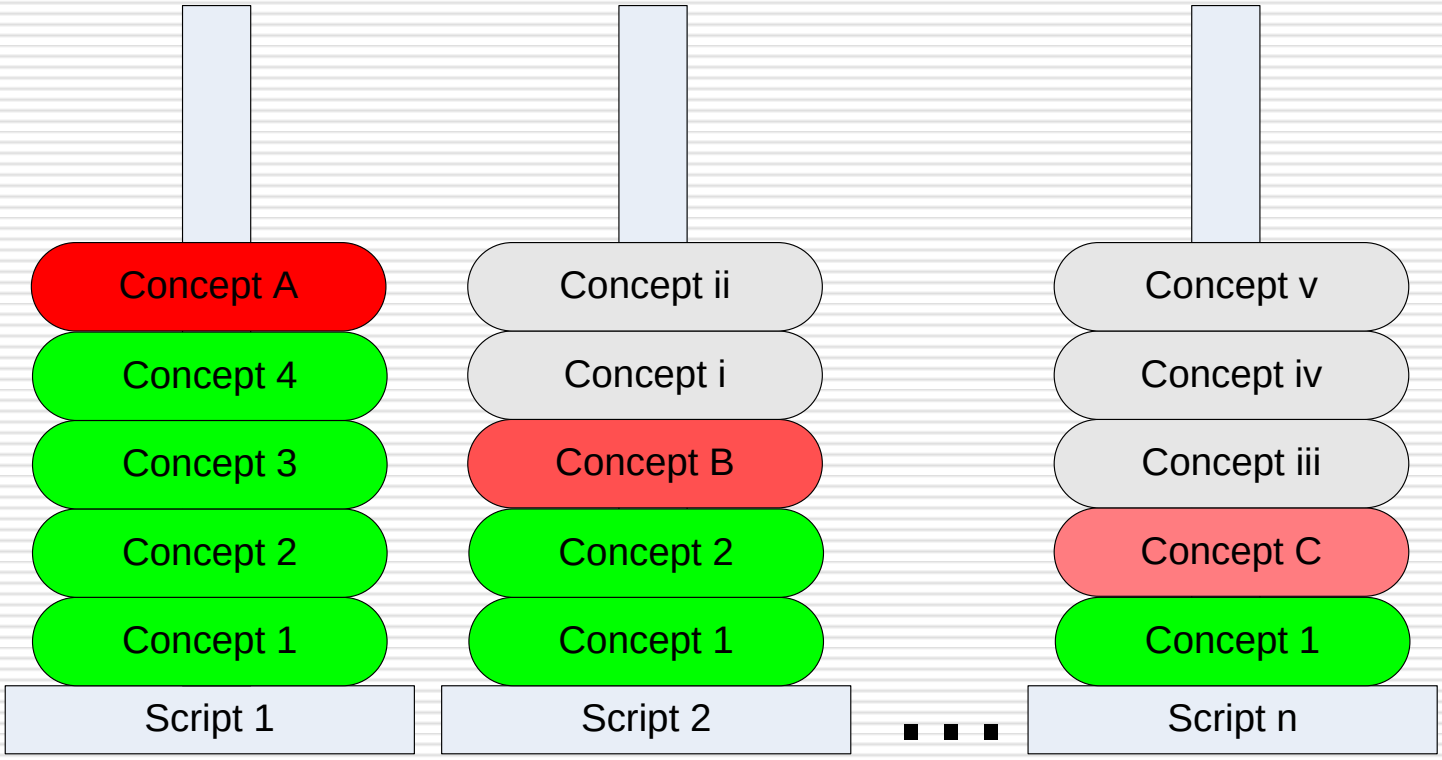
Features extraction

- **F1-Location in the image**
- **F2 –Direction**
- **F3-Speed**
- **F4 Brownian motion**
- **F4- Running**
- **F5- Crawling**
- **F6- Fighting**
- **F7-(Un-) Friendly persons**
- **F8 Tracked path goal directed**
- **F9 Tracked path without goal**
- **F10 Exceptional path**

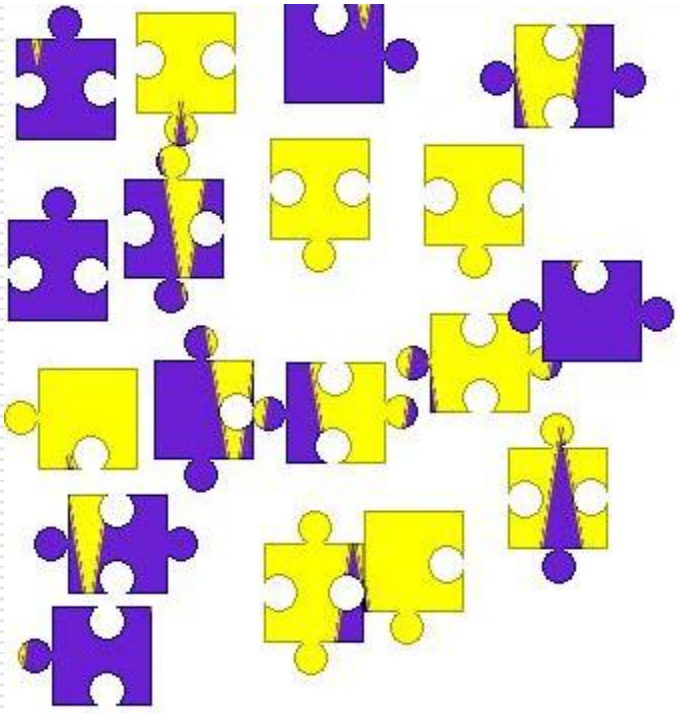
Bayesian reasoning



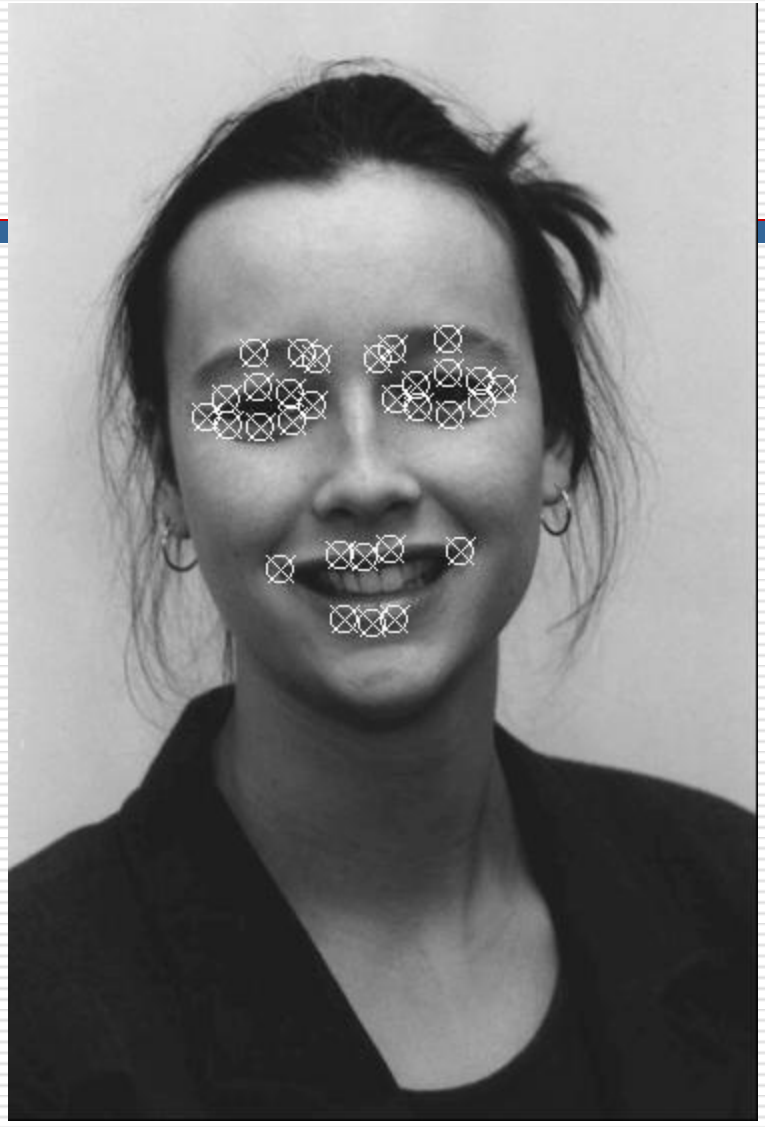
Building Hypothesis

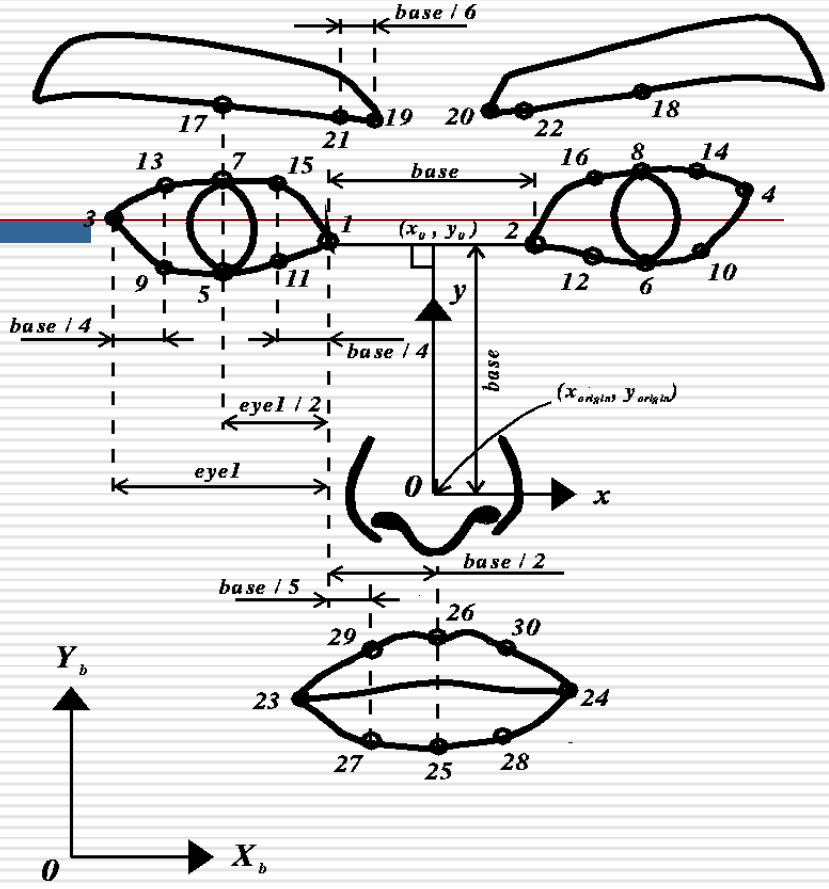


Jig Saw feature puzzling



Would you like to join me for a dinner ?




base / 6
17
19
21
20
22
18
13
7
15
base
16
8
14
4
3
1
(x_0, y_0)
2
9
5
11
12
6
10
y
base / 4
base / 4
base
(x_{origin}, y_{origin})
eye1 / 2
eye1
0
x
base / 2
base / 5
26
29
30
Y_b
23
24
27
25
28
0
X_b

TU Delft

Cohn Kanade db

Six basic emotions

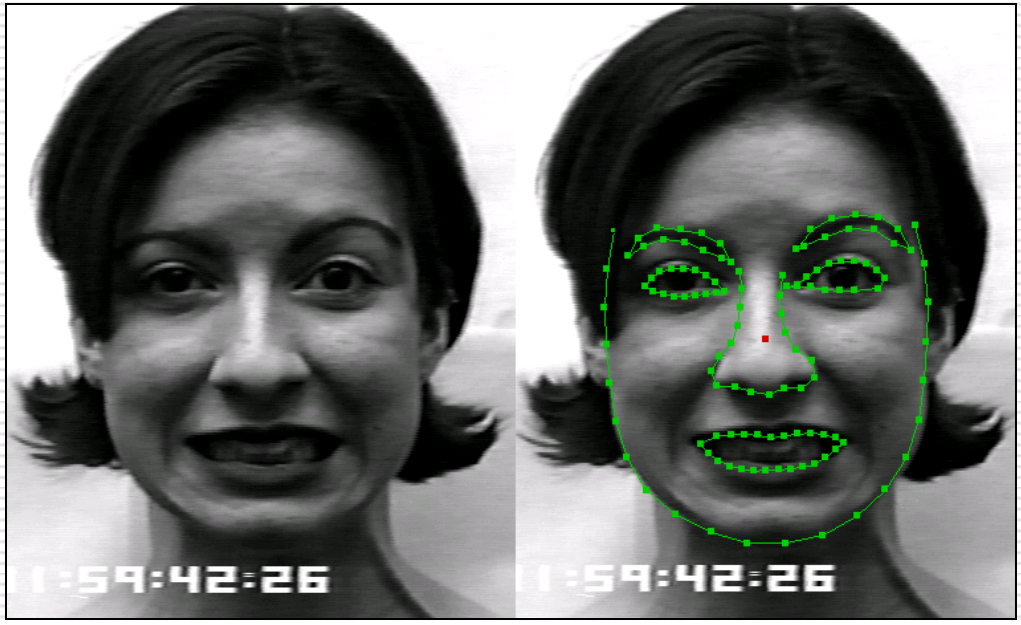
Cohn Kanade db

95 subjects

AAM

Cootes, T.F., Edwards, G.J., Taylor, C.J., "Active Appearance Models", 1998.

Face (Shape and Texture) modeling



AAM - Fitting – step1

AAM - Fitting – step2

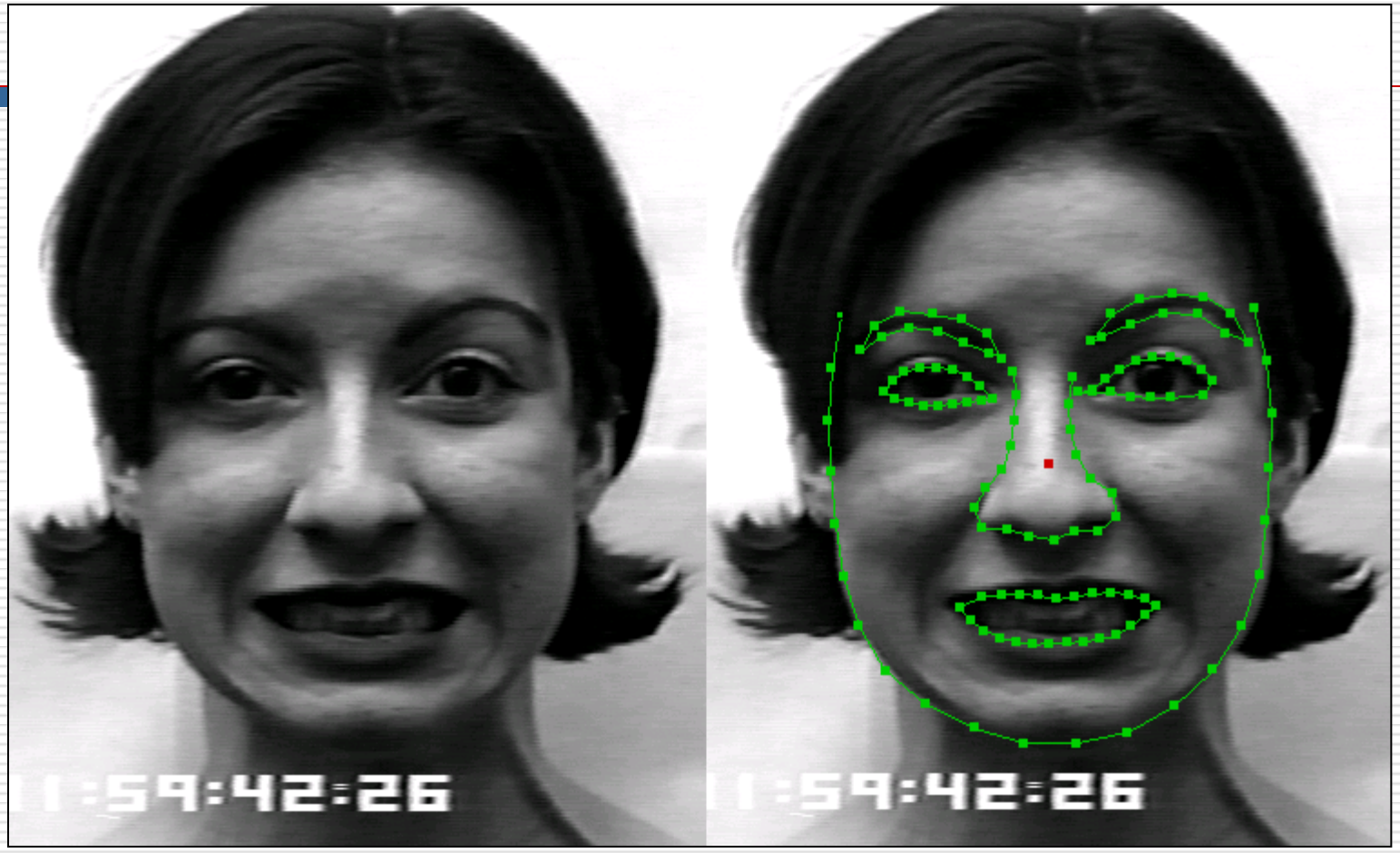
AAM - Fitting – step3

AAM - Fitting – step4

AAM - Fitting – step5

AAM - Fitting – step6

AAM in a static frame



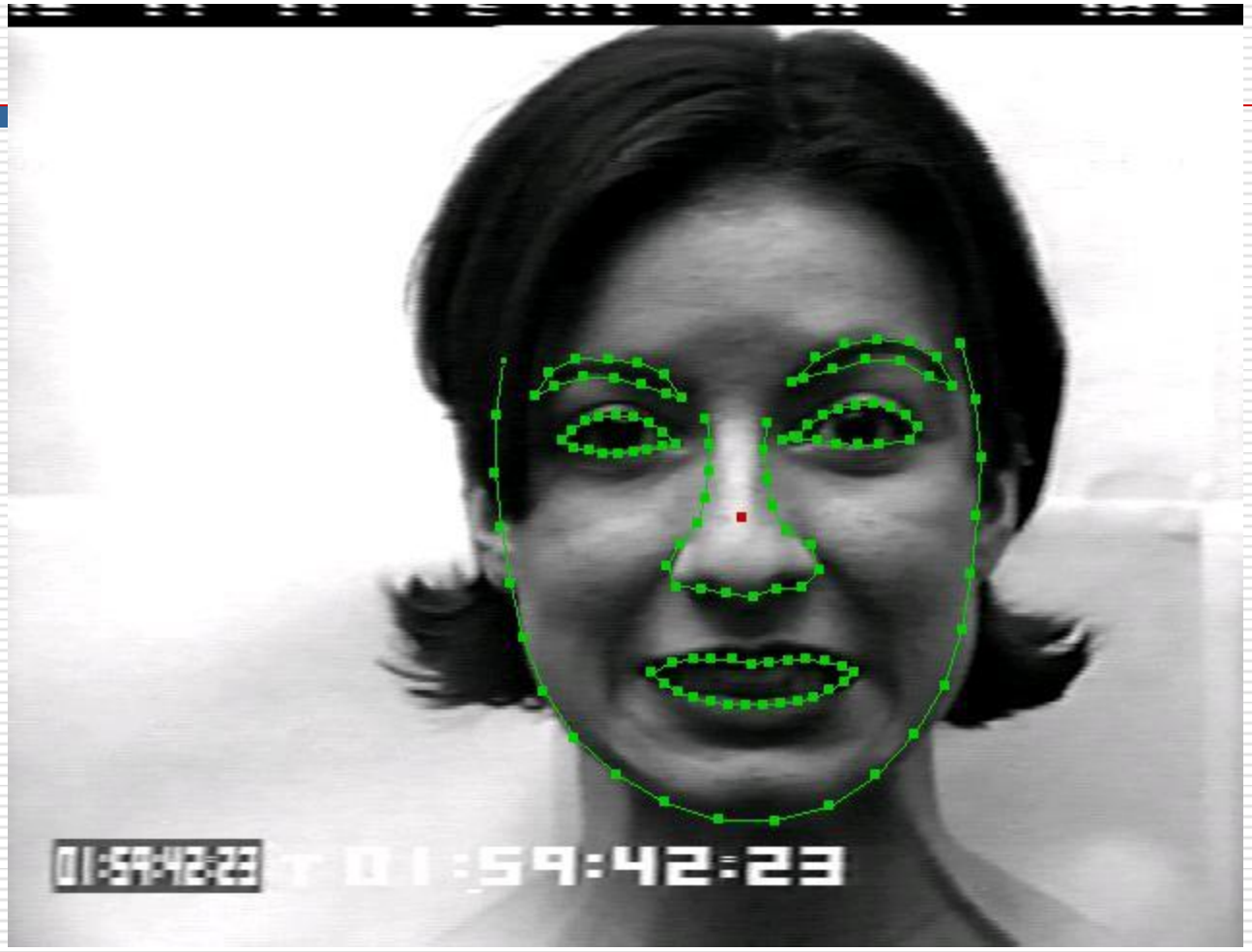
AAM in video sequences

AAM in video sequences

AAM in video sequences

AAM in video sequences

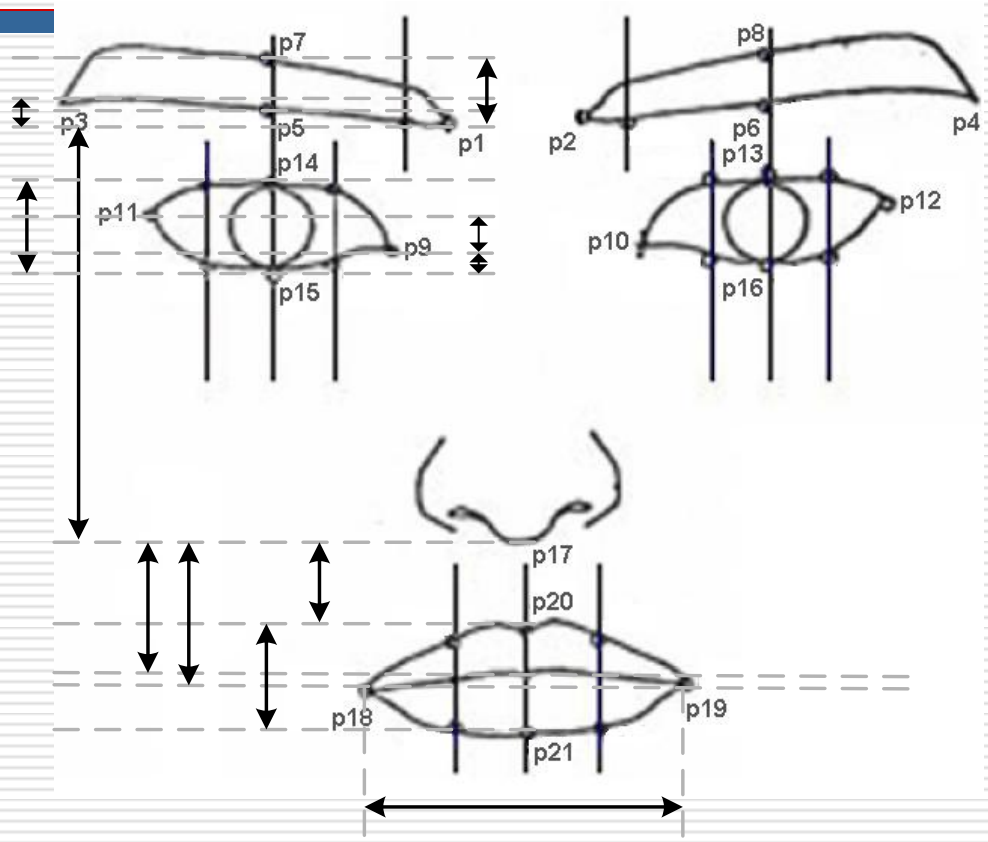
AAM in video sequences



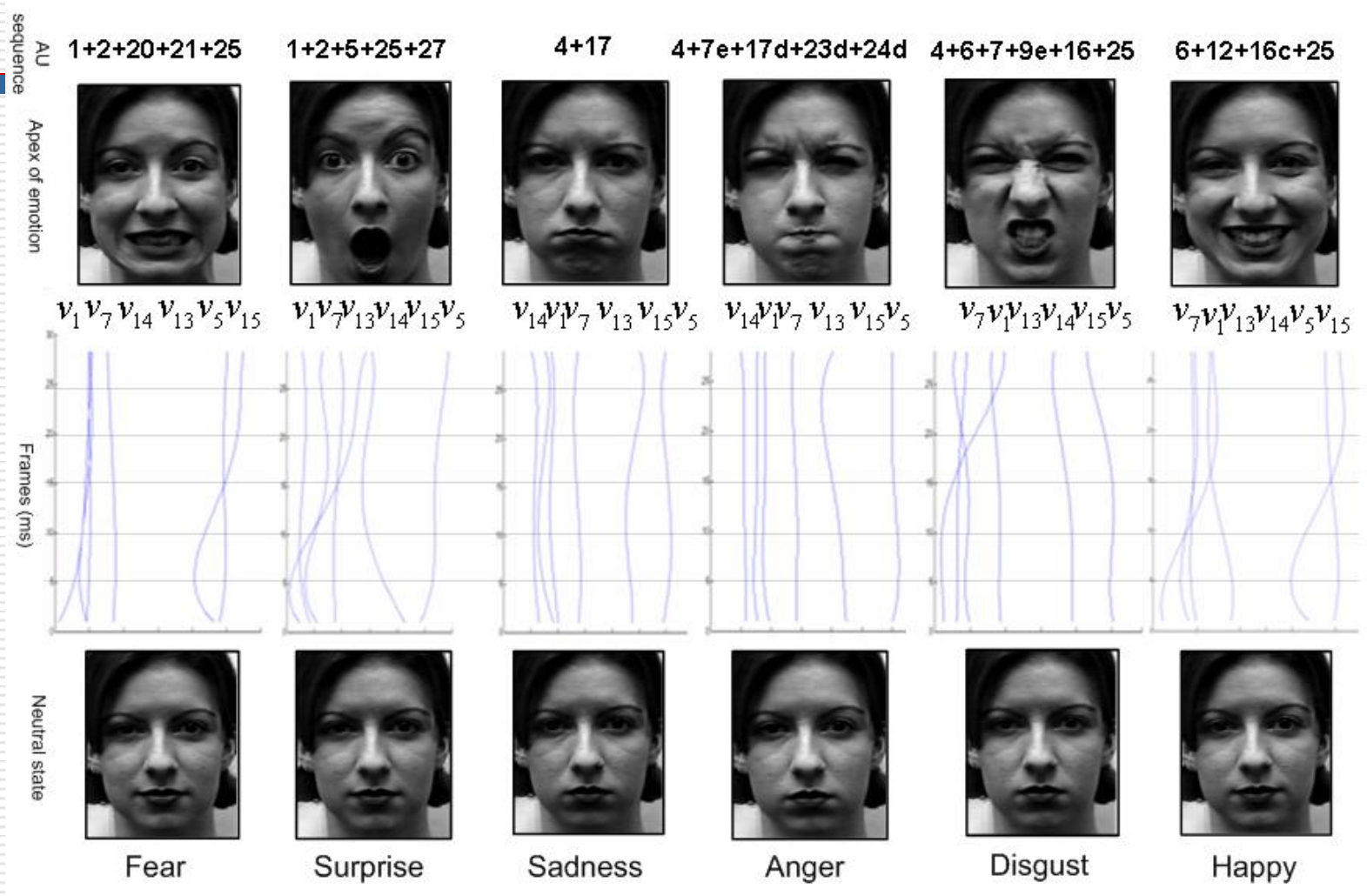
AAM in video sequences

FCP model (Kobayashi &Hara extension)

		Visual feature			Visual feature
v_1	$(P_1, P_7)_y$	Left eyebrow	v_{10}	$(P_{13}, P_{16})_y$	Right eye
v_2	$(P_1, P_3)_y$	Left eyebrow	v_{11}	$(P_{10}, P_{12})_y$	Right eye
v_3	$(P_2, P_8)_y$	Right eyebrow	v_{12}	$(P_{10}, P_{16})_y$	Right eye
v_4	$(P_2, P_4)_y$	Right Eyebrow	v_{13}	$(P_{17}, P_{20})_y$	Mouth
v_5	$(P_1, P_{17})_y$	Left Eyebrow	v_{14}	$(P_{20}, P_{21})_y$	Mouth
v_6	$(P_2, P_{17})_y$	Right eyebrow	v_{15}	$(P_{18}, P_{19})_y$	Mouth
v_7	$(P_{14}, P_{15})_y$	Left eye	v_{16}	$(P_{17}, P_{18})_y$	Mouth
v_8	$(P_9, P_{11})_y$	Left eye	v_{17}	$(P_{17}, P_{19})_x$	Mouth
v_9	$(P_9, P_{15})_y$	Left eye			



Temporal patterns of six basic emotions



FACS

1970s

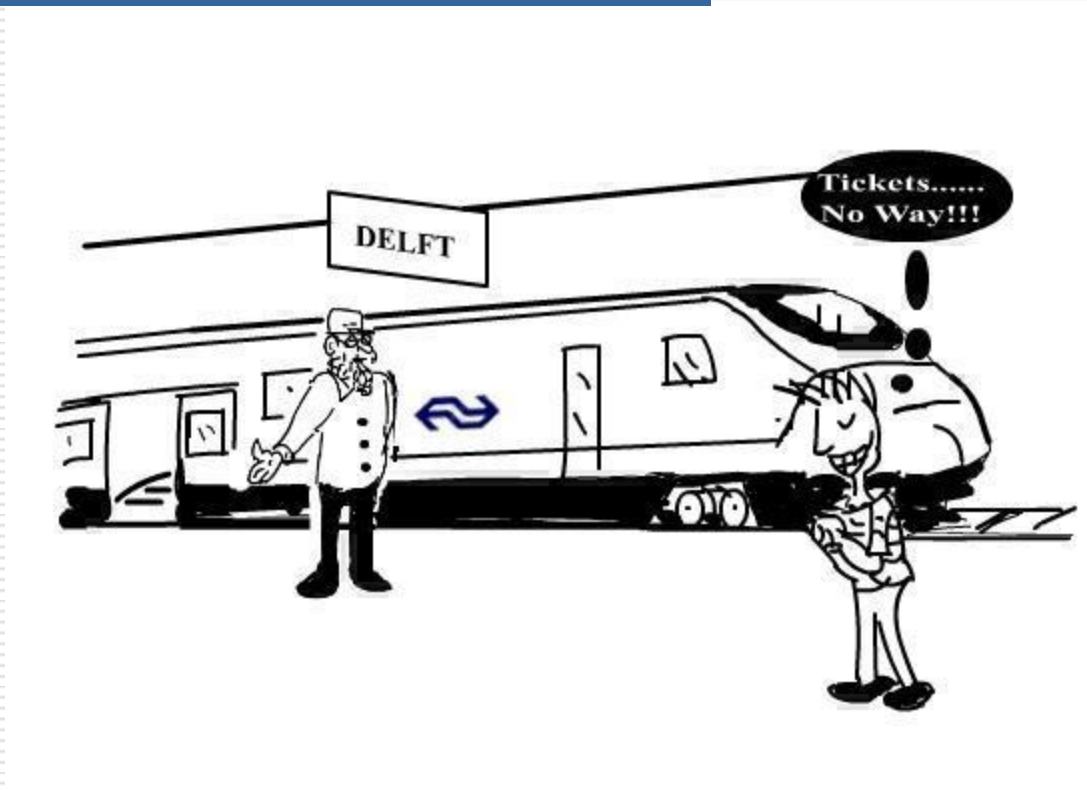
Paul Ekman

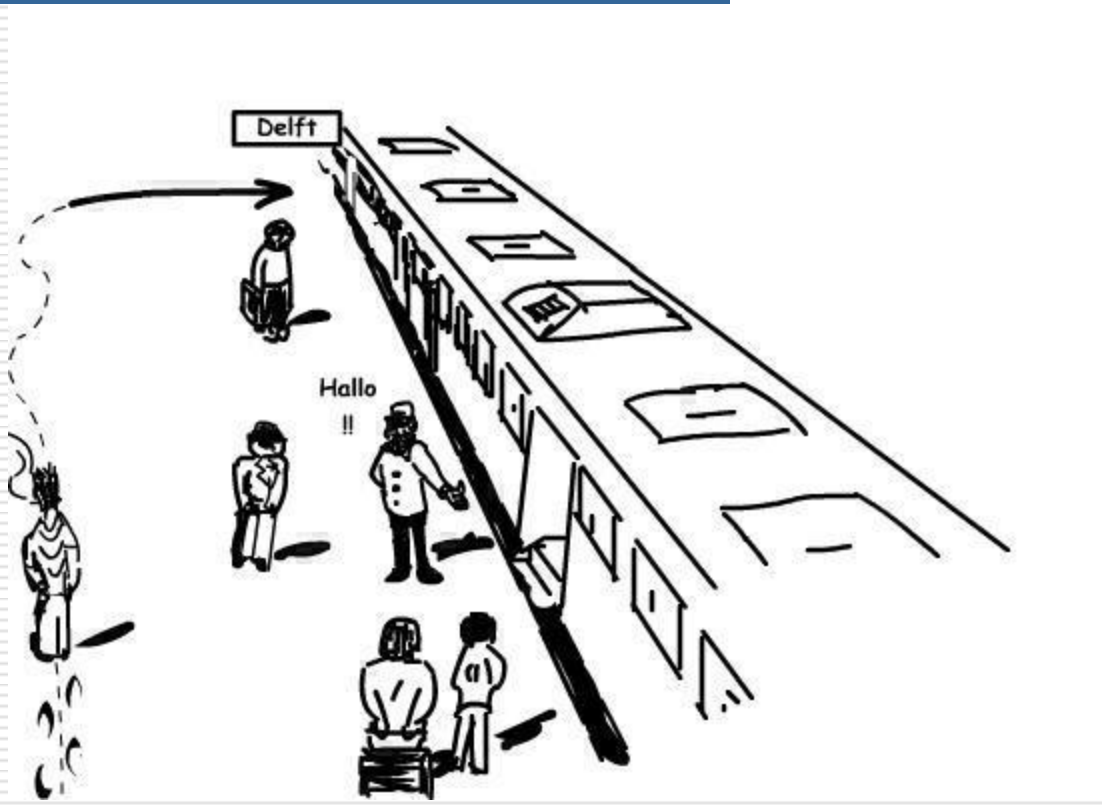
W.V. Friesen

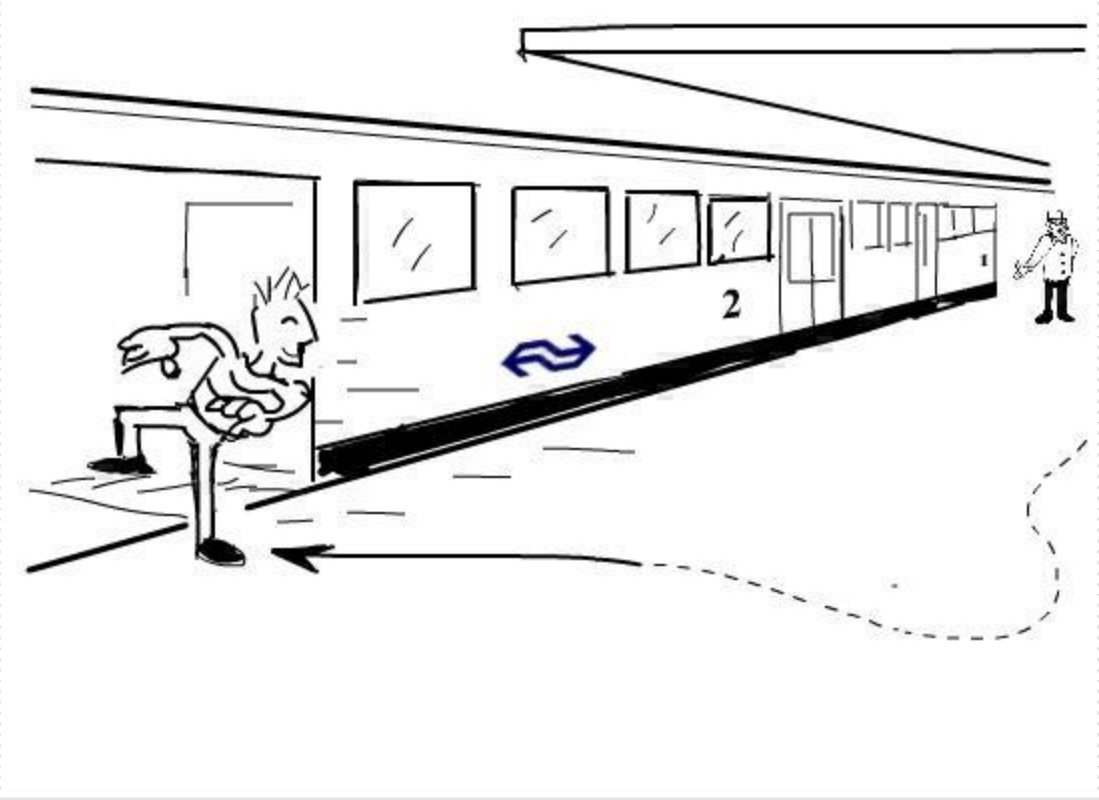
Facial Action Coding System

AU	Description	Facial muscle	Example
1	Inner Brow Raiser	Frontalis, pars medialis	
2	Outer Brow Raiser	Frontalis, pars lateralis	
4	Brow Lowerer	Corrugator supercilii, Depressor supercilii	
5	Upper Lid Raiser	Levator palpebrae superioris	
6	Cheek Raiser	Orbicularis oculi, pars orbitalis	
7	Lid Tightener	Orbicularis oculi, pars palpebralis	
9	Nose Wrinkler	Levator labii superioris alaquae nasi	
10	Upper Lip Raiser	Levator labii superioris	
11	Nasolabial Deepener	Zygomaticus minor	
12	Lip Corner Puller	Zygomaticus major	
13	Cheek Puffer	Levator anguli oris (a.k.a. Caninus)	
14	Dimpler	Buccinator	
15	Lip Corner Depressor	Depressor anguli oris (a.k.a. Triangularis)	
17	Chin Raiser	Mentalis	
18	Lip Puckerer	Incisivii labii superioris and Incisivii labii inferioris	
20	Lip stretcher	Risorius w/ platysma	
22	Lip Funneler	Orbicularis oris	
23	Lip Tightener	Orbicularis oris	
24	Lip Pressor	Orbicularis oris	
25	Lips part**	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris	
26	Jaw Drop	Masseter, relaxed Temporalis and internal Pterygoid	
27	Mouth Stretch	Pterygoids, Digastric	
28	Lip Suck	Orbicularis oris	

Example

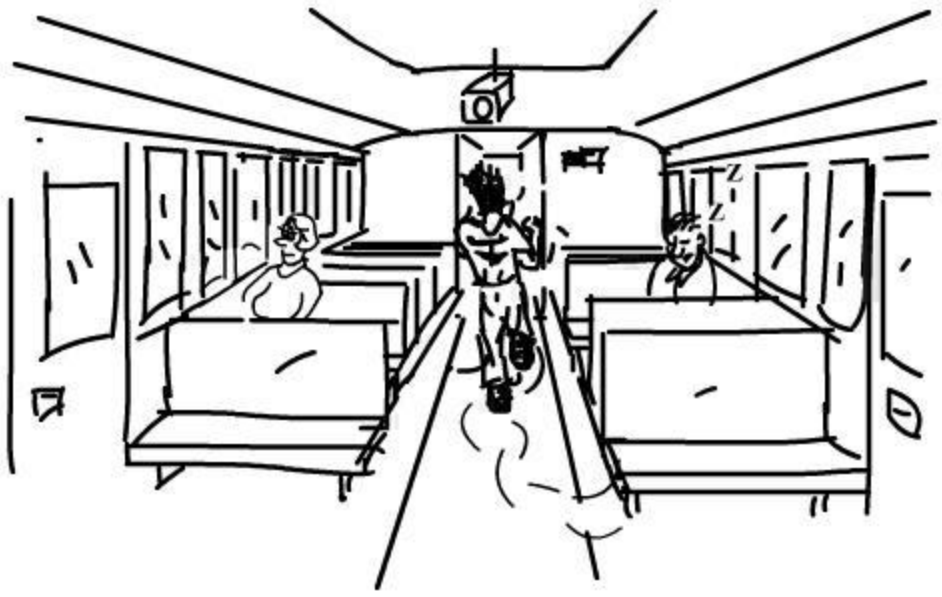



Delft
Hallo
!!


2


NRC


TOILET
Plan A
Plan B
Oops!!!!
Hello!
NRC




Toilet
open


Toil
closed


Toilet


Toilet
open


Toilet

Conclusion

- Model and implementation of automated surveillance system, based on visual speech/nonverbal communication and probabilistic reasoning
- To be developed more advanced models of speech recognition in noisy environments (noise canceling)
- To be developed more advanced models of body gestures

QUESTIONS????