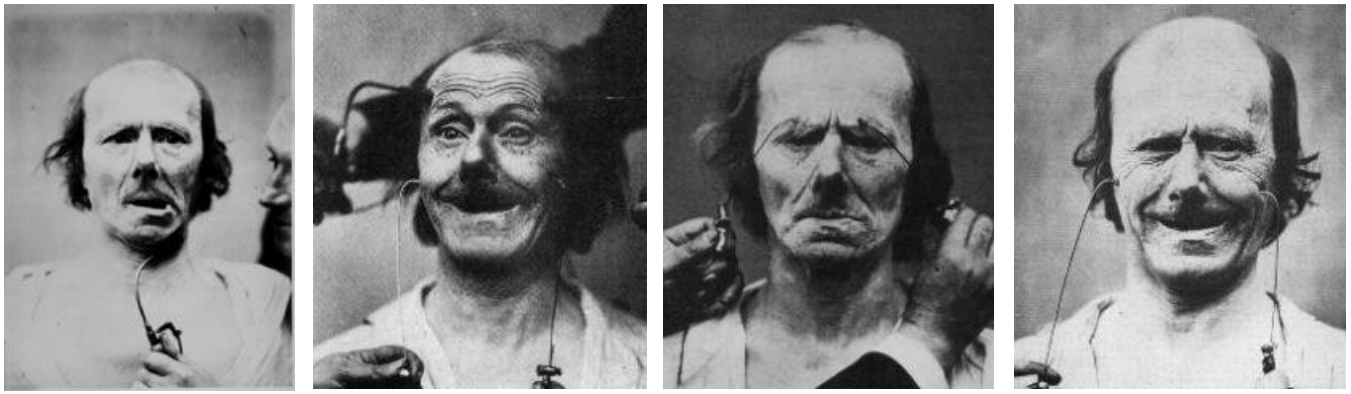
Karlsruher Institut für Technologie

Automatic Facial Expression Analysis

Vorlesung “Computer Vision für Mensch-Maschine-Interaktion”, WS 2013 / 2014

Tobias Gehrig (tobias.gehrig@kit.edu)
Prof. Dr.-Ing. Rainer Stiefelhagen (rainer.stiefelhagen@kit.edu)

FAKULTÄT FÜR INFORMATIK, INSTITUT FÜR ANTHROPOMATIK, MASCHINENSEHEN FÜR MMI



What is Facial Expression?

- *Facial expressions are the facial changes in response to a person‘s internal emotional states, intentions, or social communications.*

Role of Facial Expressions

- Almost the most powerful, natural, and immediate way (for human beings) to communicate emotions and intentions
- Face can express emotion sooner than people verbalize or realize feelings
- Faces and facial expressions are an important aspect in interpersonal communication and man-machine interfaces

Facial Expressions

- Facial expression(s):
  - nonverbal communication
  - voluntary / involuntary
  - results from one or more motions or positions of the muscles of the face
  - closely associated with our emotions

- The fact:
  - Most people's success rate at reading emotions from facial expression is only a little over 50 percent.

Facial Expression Analysis vs. Emotion Analysis

- Emotion analysis requires higher level knowledge, such as context information.
  - e.g. smile could be caused by happiness or frustration
    [Hoque and Picard, *„Acted vs. Natural frustration and delight: Many people smile in natural frustration“*, FG 2011]

- Besides emotions, facial expressions can also express intention, cognitive processes, physical effort, etc.

Emotions conveyed by Facial Expressions

- Six basic emotions (according to Ekman)
  - assumed to be innate


Happiness


Surprise

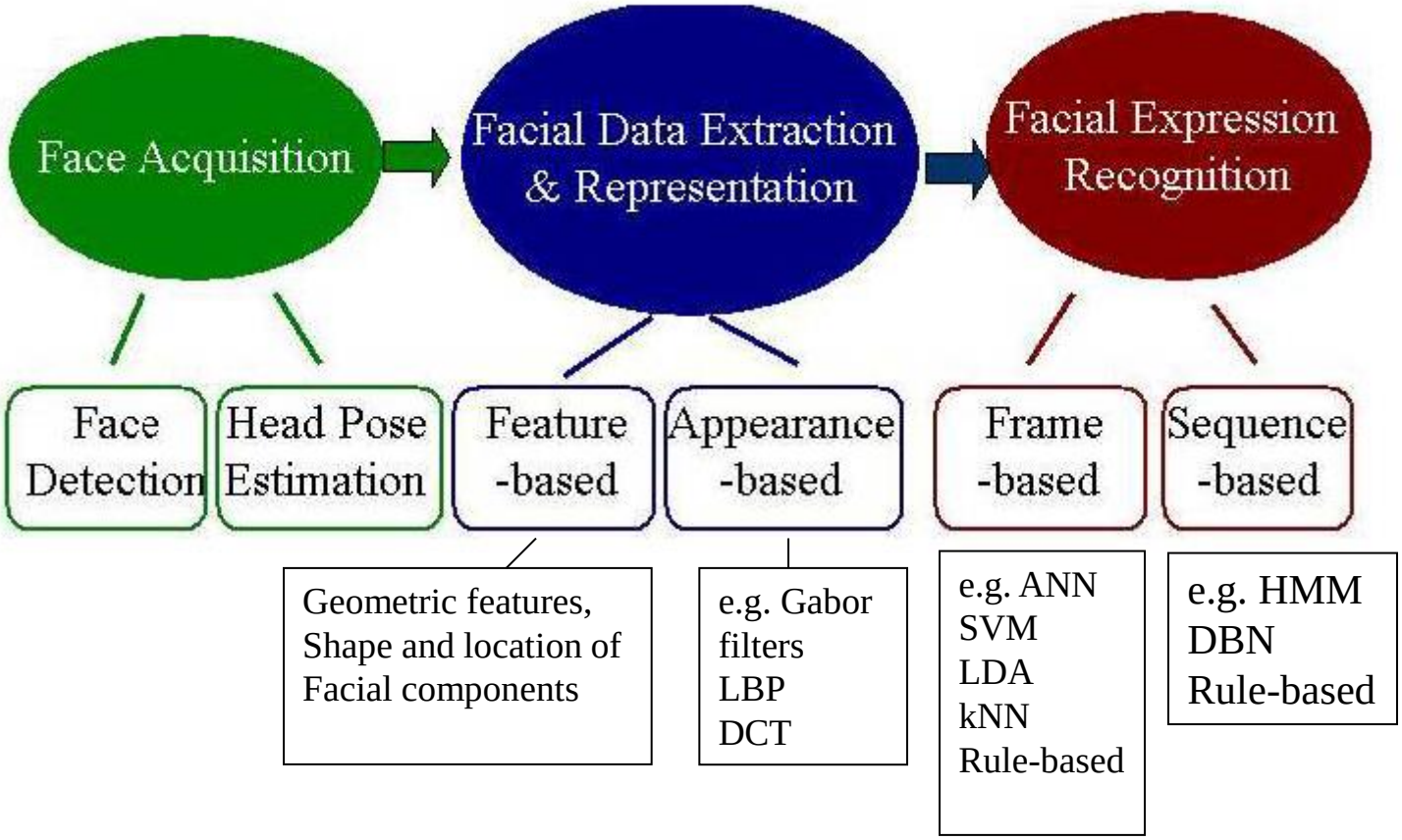

Sadness


Fear


Disgust


Anger

Basic Structure of Facial Expression Analysis Systems



Problem Space for Facial Expression Analysis

- Level of Description (emotions vs. facial actions)
- Individual Differences in Subjects
- Transitions Among Expressions
- Intensity of Facial Expression
- Deliberate vs. Spontaneous Expression
- Head Orientation and Scene Complexity
- Image Acquisition and Resolution
- Reliability of Ground Truth
- Databases
- Relation to Other Facial Behavior or Nonfacial Behavior

Level of Description

- Emotions
  - Discrete classes:
    - Six basic emotions

      Emotion-specified facial expression

      1. Disgust; 2. Fear; 3. Joy; 4. Surprise; 5. Sadness; 6. Anger

    - Positive, neutral, negative
  - Continuous valued dimensions: Arousal, Valence, ...
- Action Units,
  - ...

Facial Action Coding System (FACS)

- Developed by Ekman & Friesen (1978)
- A human-observer based system designed to detect subtle changes in facial features.
- Viewing videotaped facial behavior in slow motion, trained observer can manually FACS code all possible facial displays
- These facial displays are referred to as action units (AU) and may occur individually or in combinations.

Interpretation of Facial Actions

- Facial expressions can be linked with psychological interpretations, e.g. emotions
  - → e.g. Facial Action Coding System Affect Interpretation Dictionary (FACSAID) project (Ekman & Friesen)

- Facial expressions are also assumed to reveal something about sincereness or lying
  - Can be used to reveal deception (Ekman)
    - e.g. through analysis of micro-expressions (very quick involuntary expression that last less then a quarter-second)
  - Detection of voluntary vs. involuntary smiles
    - insincere and voluntary smile: contraction of zygomatic major alone
    - sincere and involuntary (Duchenne) smile: contraction of zygomatic major and inferior part of orbicularis oculi
    - temporal dynamics also very important

- Has also been used for analysis of depression and pain

Action Units (AUs)

- There are 44 **AU**s
- 30 **AU**s related to **contractions** of special **facial muscles**:
  - 12 AUs for upper face
  - 18 AUs for lower face

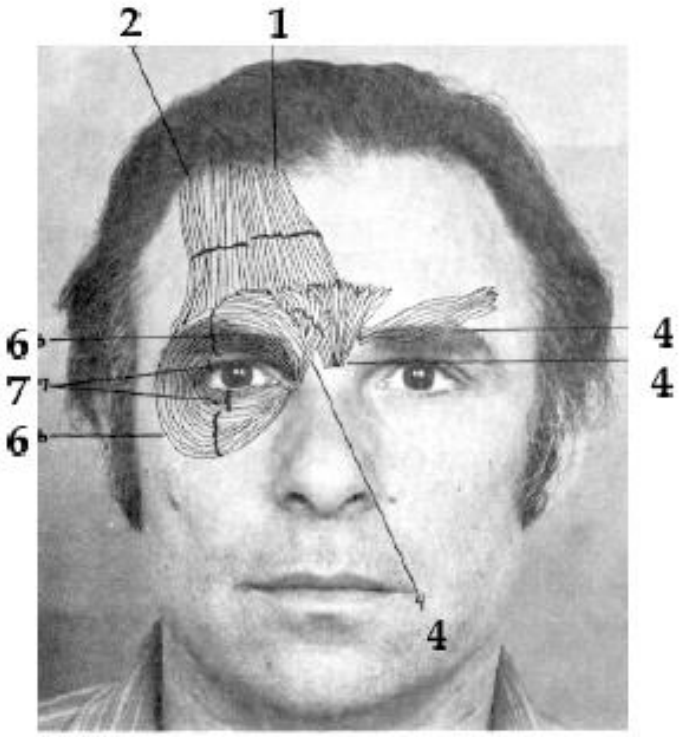
- Anatomic basis of the remaining 14 is unspecified
  => referred to in FACS as miscellaneous actions

- For action units that vary in intensity, a 5-point ordinal scale is used to measure the degree of muscle contraction

Upper Face Action Units

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink

Upper Facial Action Units (cont.)



(Ekman & Friesen 1978)

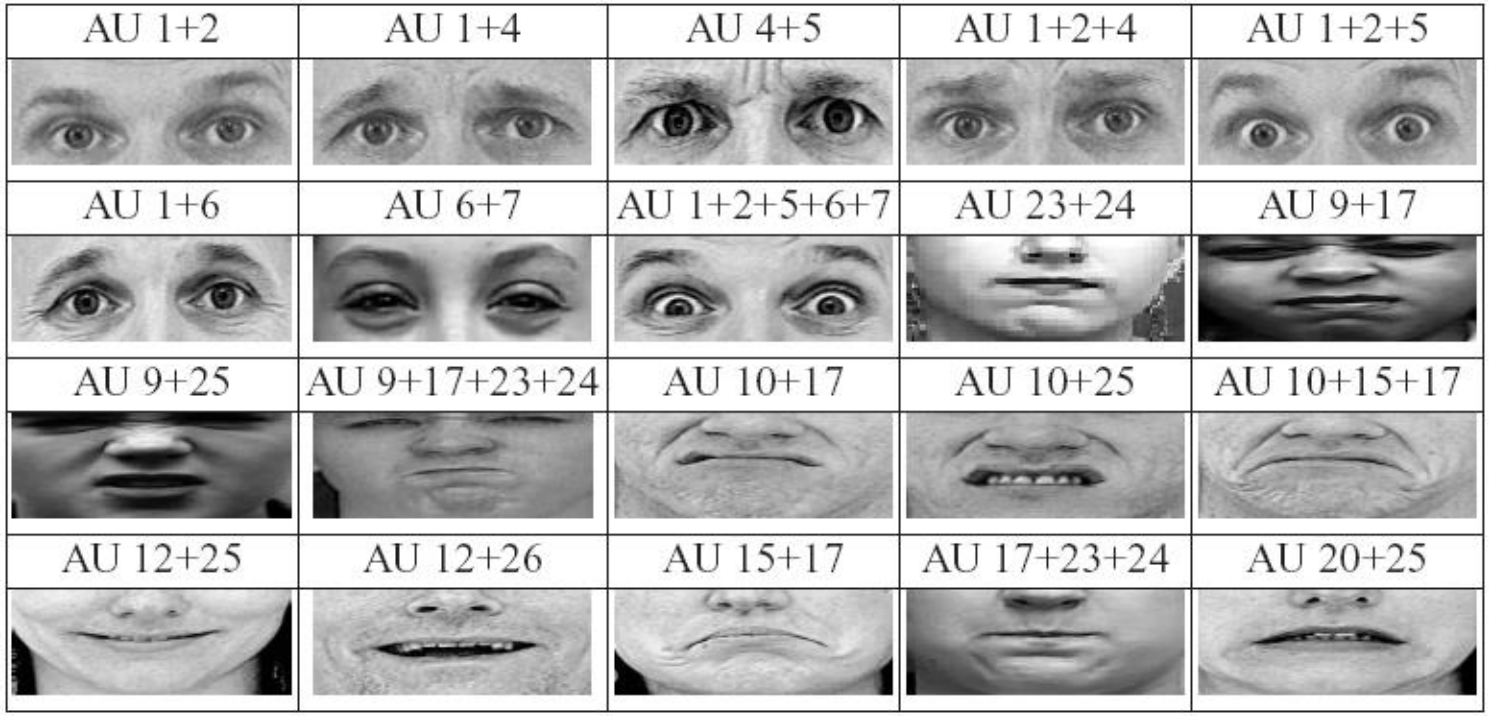
Lower Face Action Units

Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Miscellaneous Actions

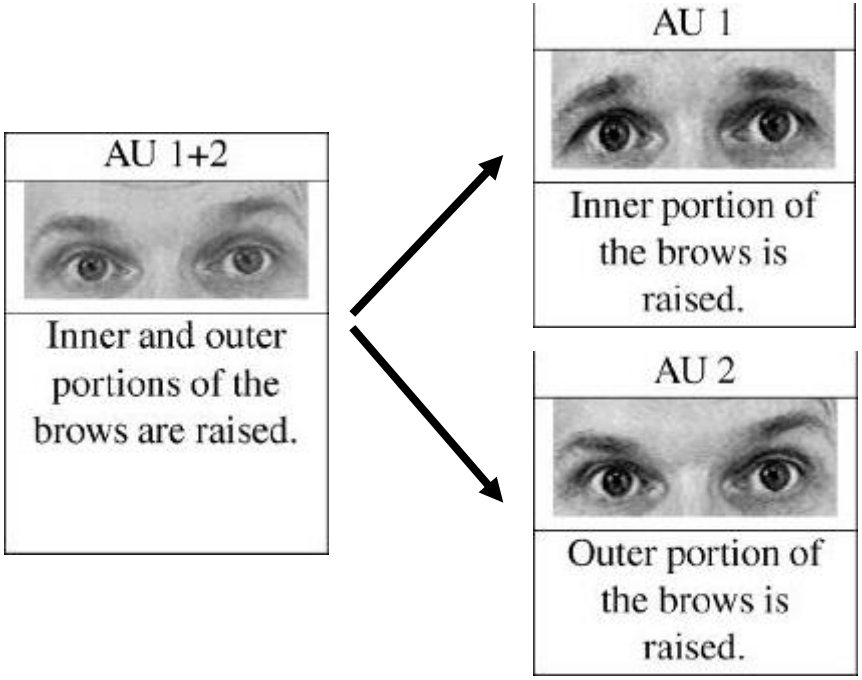
AU	Description
8	Lips toward
19	Tongue show
21	Neck tighten
29	Jaw thrust
30	Jaw sideways
31	Jaw clench
32	Bite lip
33	Blow
34	Puff
35	Cheek suck
36	Tongue bulge
37	Lip wipe
38	Nostril dilate
39	Nostril compress

Some Examples of Combination of FACS Action Units

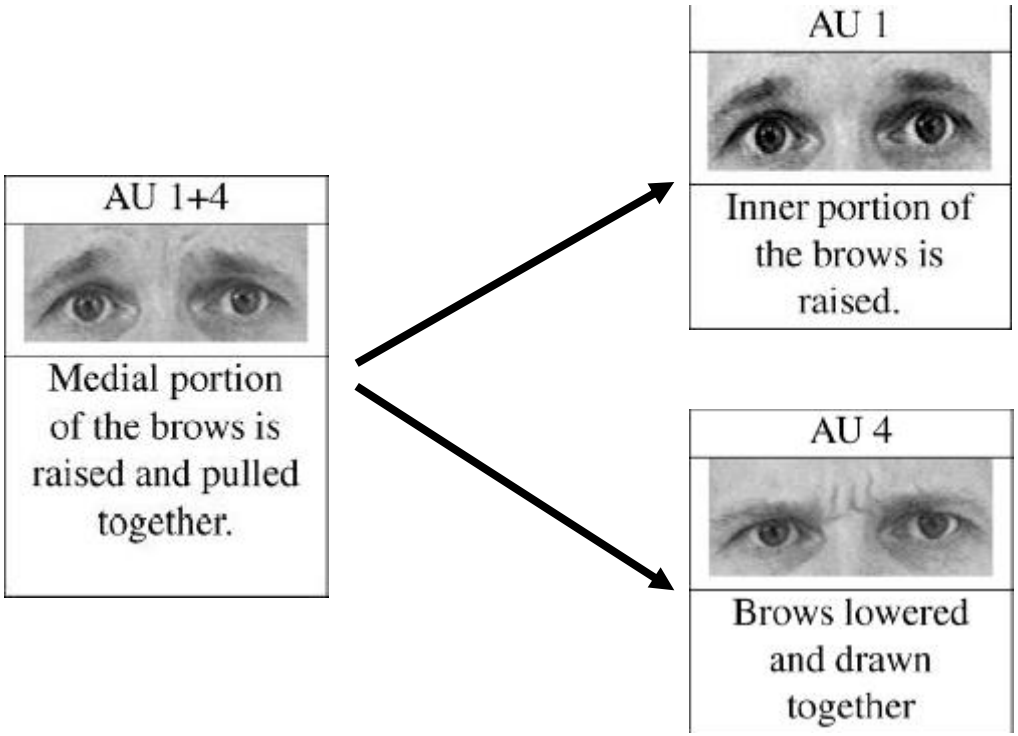


Combinations of AUs

- More than 7000 different AU combinations have been observed.
- ***Additive*** *– appearance of single* ***AU****s does* ***not change***
- ***Nonadditive*** *– appearance of single* ***AU****s does* ***change***



Nonadditive AU Combination Sample



Individual Differences in Subjects

- Variations in appearance:
  - Face shape,
  - texture,
  - color,
  - facial and scalp hair
  - due to sex, ethnic background, and age differences.

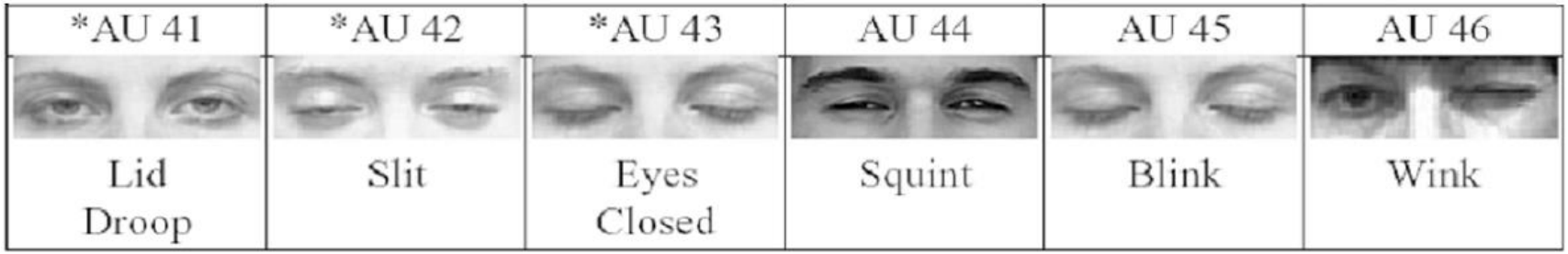
- Variations in expressiveness

Transitions Among Expressions

- A simplifying assumption in facial expression analysis is that expressions are singular and begin and end with a neutral position.
- Transitions from action units or combination of actions to another may involve no intervening neutral state.
- Parsing the stream of behavior is an essential requirement of a robust facial analysis system, and training data are needed that include dynamic combinations of action units, which may be either additive or nonadditive.

Intensity of Facial Expression

- Facial actions can vary in intensity
- FACS coding uses 5-point intensity scale to describe intensity variation of action units
- Some related action units function as sets to represent intensity variation. In the eye region, action units 41, 42, and 43 or 45 can represent intensity variation from slightly drooped to closed eyes.



Deliberate vs. Spontaneous Expression

- Most face expression data have been collected by asking subjects to perform a series of expressions.
- These facial actions may differ in appearance and timing from spontaneously occuring ones.
- Deliberate and spontaneous facial behavior are mediated by separate motor pathways.
- Differences in the timing of spontaneous and deliberate facial actions are particularly important in many pattern recognition approaches, such as hidden Markov modelling, are highly dependent on the timing of the appearance change.

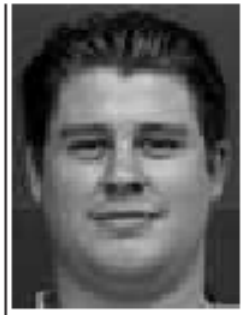
Head Orientation and Scene Complexity

- Face orientation relative to the camera, the presence and actions of other people, and background conditions influence face analysis.
- Large out-of-plane rotation in head position is common and often accompanies change in expression.
- To develop pose invariant methods of face expression analysis, image data are needed in which facial expression changes in combination with significant nonplanar change in pose.

Image Acquisition and Resolution

- The image acquisition factors that may influence facial expression analysis:
  - the number of video cameras,
  - the capture quality of the cameras
  - the frame rate,
  - the size of the face image relative to total image dimensions,
  - the ambient lighting.

A Face at Different Resolutions

Face				
Process	96 x 128	69 x 93	48 x 64	24 x 32
Detect?	Yes	Yes	Yes	Yes
Pose?	Yes	Yes	Yes	Yes
Recognize?	Yes	Yes	Yes	Maybe
Features?	Yes	Yes	Maybe	No
Expressions?	Yes	Yes	Maybe	No

Reliability of Ground Truth

- To be able to assess the performance of facial expression analysis systems, training and test data should be accurately labeled.
- To ensure internal validity, expression data must be manually coded, and the reliability of the coding verified.
- Interobserver reliability can be improved by providing rigorous training to observers and monitoring their performance.
- Monitoring is best achieved by having observers independently code a portion of the same data. As a general rule, 15% to 20% of data should be comparison-coded.

Databases

- A facial expression database should include:
  - Various facial expressions,
  - Large number of subjects,
  - Subjects from both genders,
  - Subjects from different age groups,
  - Subjects from different ethnic background,
- The recording conditions should imitate the real-life conditions

Cohn-Kanade AU-Coded Facial Expression Database

- 100 subjects from varying ethnic backgrounds.
- 23 different facial expressions (single action units and combinations of action units)
- Frontal faces, small head motion
- Variations in lighting: ambient lighting, single-high-intensity lamp, dual high-intensity lamps with reflective umbrellas
- Coded with FACS and assigned emotion-specified labels (happy, surprise, anger, disgust, fear, sadness)

Relation to other Facial Behavior or Nonfacial Behavior

- One of several channels of nonverbal communication.
- The message values of various modes may differ depending on context.
- For robustness, should be integrated with:
  - Gesture
  - Prosody
  - Speech
- i.e. Combining facial features with acoustic features would help to separate the effects of facial actions due to facial expression and those due to speech related movements.

Properties of an Ideal Facial Expression Analysis System

Robustness	
Rb1	Deal with subjects of different age, gender, ethnicity
Rb2	Handle lighting changes
Rb3	Handle large head motion
Rb4	Handle occlusion
Rb5	Handle different image resolution
Rb6	Recognize all possible expressions
Rb7	Recognize expressions with different intensity
Rb8	Recognize asymmetrical expressions
Rb9	Recognize spontaneous expressions
Automatic process	
Am1	Automatic face acquisition
Am2	Automatic facial feature extraction
Am3	Automatic expression recognition
Real-time process	
Rt1	Real-time face acquisition
Rt2	Real-time facial feature extraction
Rt3	Real-time expression recognition
Autonomic Process	
An1	Output recognition with confidence
An2	Adaptive to different level outputs based on input images

The Research Directions of the Facial Expression Analysis

- Build more robust systems for face acquisiton, facial data extraction and representation, and facial expression recognition to handle head motion (in-plane and out-of-plane), occlusion, lighting changes, and lower intensity of expressions.
- Employ more facial features to recognize more expressions and to achieve a higher recognition rate.
- Recognize facial action units and their combinations rather than emotion-specified expressions.
- Recognize action units as they occur spontaneously.
- Develop fully automatic and real-time AFEA systems.

Recent Advances

Example 1 (Tian et al.)

Example 2 Bartlett et al.

Table 11.6. Recent Advances of Facial Expression Analysis

	Method	Property		
		Real time	Fully automatic	Others
CMU S1 [81, 82, 83]	Detect face in 1st frame	Yes	Yes	Handle limited head motion;
	Initialize features in 1st frame	No	No	faces with glasses and hair;
	Track geometric features	Yes	Yes	recognize more than 30
	Extract appearance features	Yes	Yes	AUs and combinations;
	Neural network classifier	Yes	Yes	1st frame is frontal and expressionless
CMU S2 [14, 57, 90]	Detect face in 1st frame	Yes	Yes	Handle large head motion;
	Map face model in 1st frame	Yes	Yes	handle lighting changes;
	Track 3D head	Yes	Yes	handle occlusions;
	Stabilize facial region	Yes	Yes	spontaneous expressions;
	Extract geometric features	Yes	Yes	1st frame is frontal
	Rule-based classifier	Yes	Yes	and expressionless
UCSD S1 [33]	Detect face in each frame	Yes	Yes	Frontal view;
	Resize and crop face	Yes	Yes	recognize basic expressions;
	Extract appearance features	Yes	Yes	neutral face needed
	SVM classifier	Yes	Yes	
UCSD S2 [3]	Estimate 3D head pose	No	No	Handle large head motion;
	Stabilize facial region	Yes	Yes	neutral face needed;
	Extract appearance features	Yes	Yes	recognize single AUs;
	HMM-SVM classifier	Yes	Yes	spontaneous expressions
UIUC S1 [13]	Initialize face model	No	No	Handle limited head motion;
	Track face in 3D	Yes	Yes	recognize basic expressions;
	Extract geometric features	Yes	Yes	1st frame is frontal
	HMM classifier	Yes	Yes	and expressionless
UIUC S2 [88]	Initialize face model	No	No	Handle limited head motion;
	Track face in 3D	Yes	Yes	handle lighting changes;
	Extract geometric features	Yes	Yes	recognize basic expressions;
	Extract appearance features	No	Yes	1st frame is frontal
	NN-HMM classifier	Yes	Yes	and expressionless

Recent systems

Table 11.8. FACS AU or expression recognition of recent advances. SVM, support vector m chines; MLR, multinomial logistic ridge regression; HMM, hidden Markov models; BN, Bayesia network; GMM, Gaussian mixture model.

Systems	Recognition methods	Recognition rate	Recognized outputs	Databases
CMU S1 [81, 82, 83]	Neural network (frame)	95.5%	16 single AUs and their combinations	Ekman-Hager [26], Cohn-Kanade [43]
CMU S2 [14, 57]	Rule-based (sequence)	100%	Blink, nonblink	Frank-Ekman [36]
		57%	Brow up, down, and non-motion	
UCSD S1 [33]	SVM+MLR (frame)	91.5%	6 Basic expressions	Cohn-Kanade [43]
UCSD S2 [3]	SVM + HMM (sequence)	98%	blink, non-blink	Frank-Ekman [36]
		70%	brow up, down, and nonmotion	
UIUC S1 [13]	BN + HMM (frame & sequence)	73.22%	6 Basic expressions	Cohn-Kanade [43]
		66.53%	6 Basic expressions	UIUC-Chen [12]
UIUC S2 [88]	NN + GMM (frame)	71%	6 Basic expressions	Cohn-Kanade [43]

Example 1 (Tian et al.) →

Example 2 Bartlett et al. →

From Tian et al., 2003, “Facial Expression Analysis”

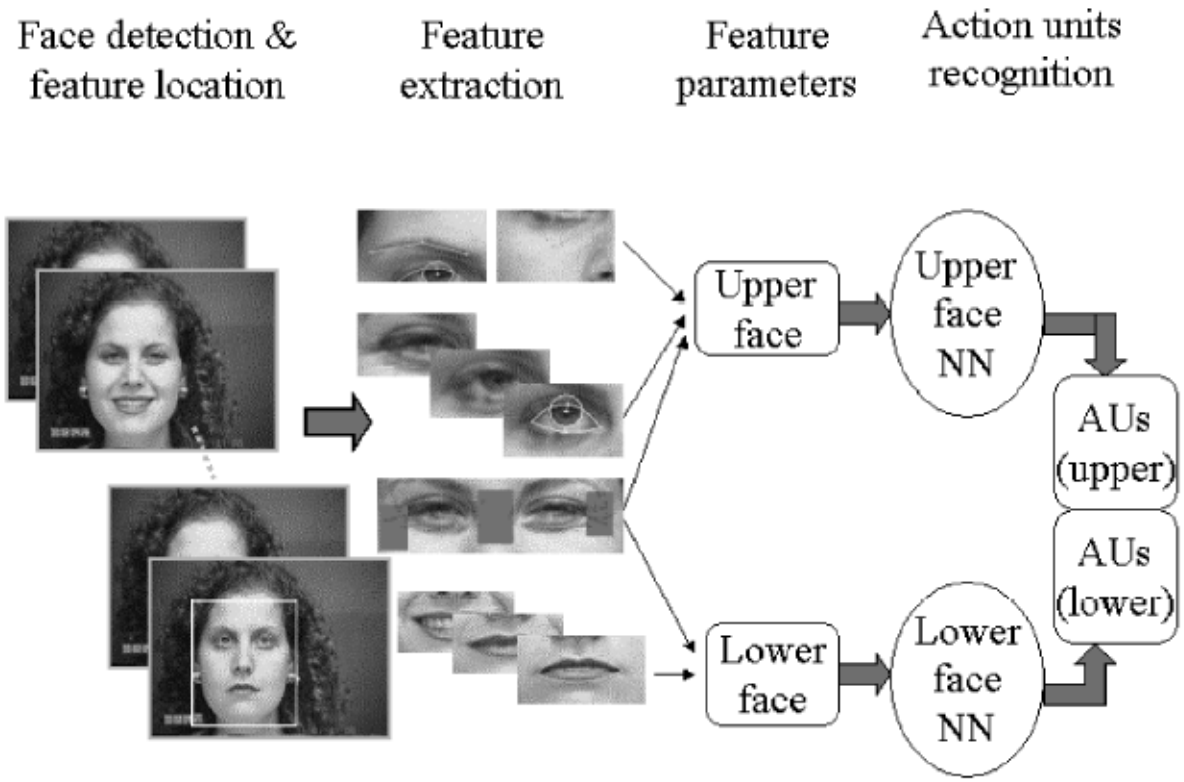
Recognizing Action Units for Facial Expression Analysis

Y. Tian, T. Kanade, J. Cohn
IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp. 97-115, Feb. 2001

- An Automatic Facial Analysis (AFA) system to analyze facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features (depending of facial furrows) in a nearly frontal-view image sequences.

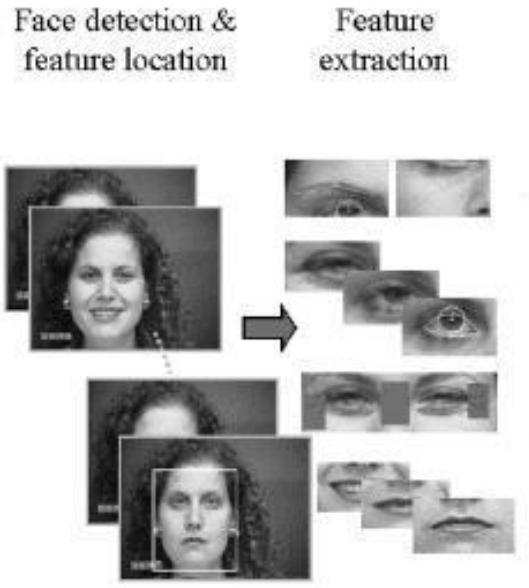
- A group of action units (neutral expression, six upper face AUs and 10 lower face AUs) are recognized whether they occur alone or in combinations.

Feature-based Automatic Facial Action Analysis (AFA) System



System Overview

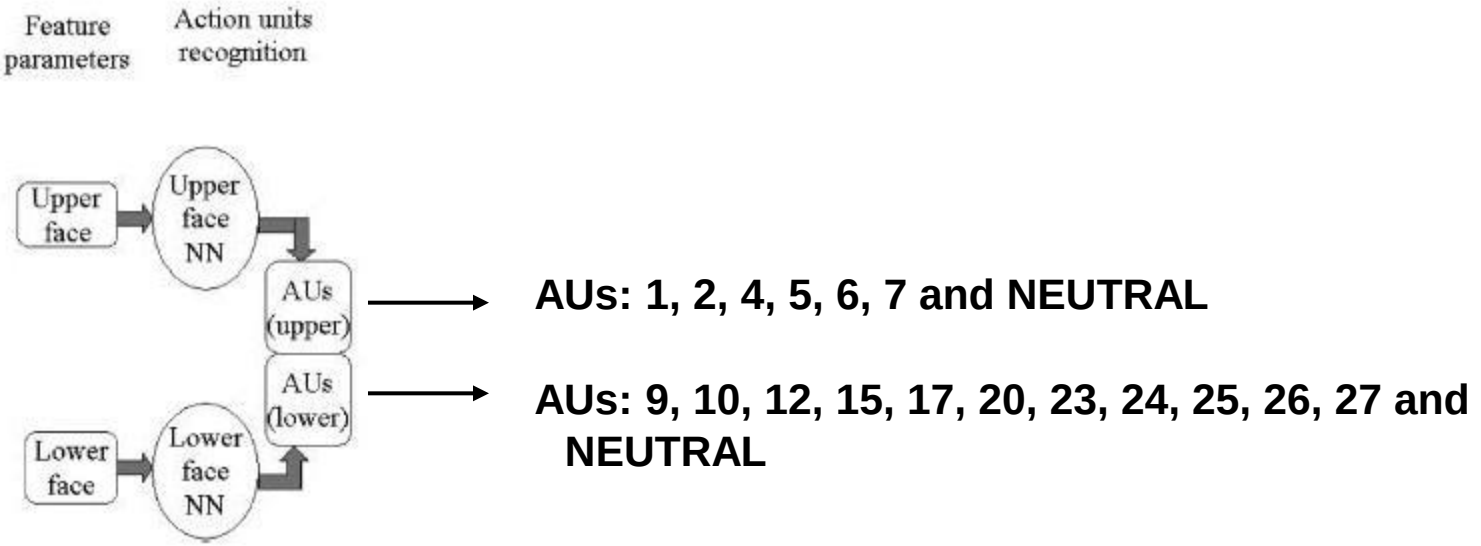
- Region of the face and location of individual face features detected automatically in the initial frame
  - → neural net based approach (Rowley et al.)
- Contours of face features and components adjusted manually in the initial frame.
- Face features are then tracked automatically
  - **permanent features** (e.g., brows, eyes, lips)
  - **transient features** (lines and furrows)



System Overview

- **Group facial features into separate collections of feature parameters**
  - 15 normalized upper face parameters (shape, motion, eye state, motion of brow and cheek, and furrows)
  - 9 normalized lower face parameters (shape, motion, lip state and furrows)

- Parameters fed to **two neural-network-based classifiers**

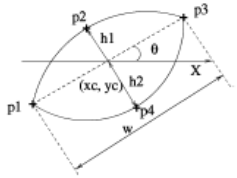
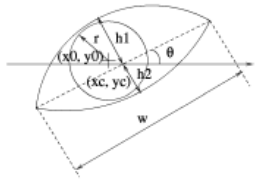
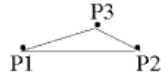
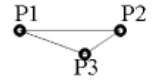
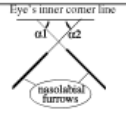


Facial Feature Extraction

- **Multistate Facial Component Models of a Frontal Face**

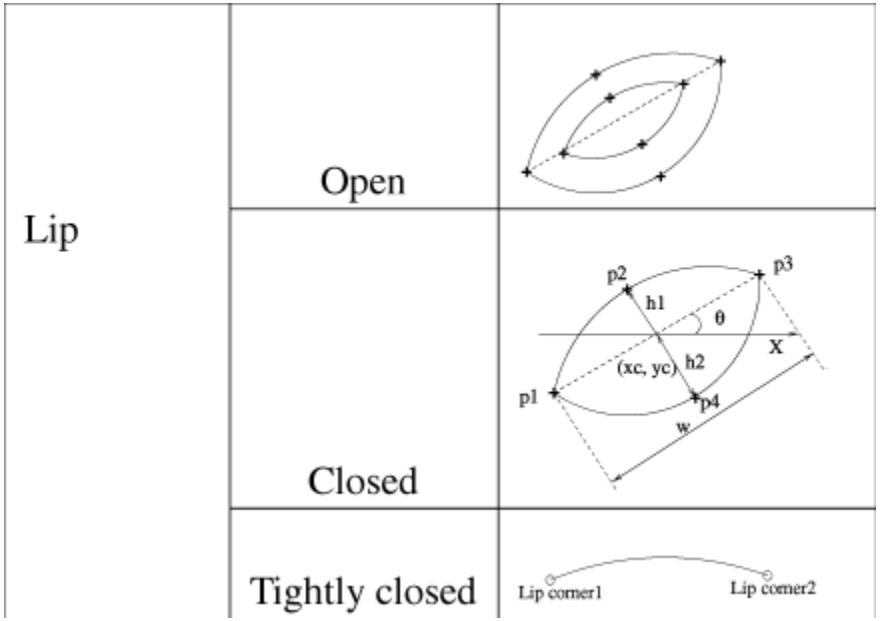
Permanent components

Transient component

Component	State	Description/Feature
Lip	Open	
	Closed	
	Tightly closed	
Eye	Open	
	Closed	
Brow	Present	
Cheek	Present	
Furrow	Present	
	Absent	

Permanent Features - Lip

- 3 **states** represent **open**, **closed** and **tightly closed**
- different lip contour templates for each lip state



➔ **two parabolic arcs with 6 parameters:**

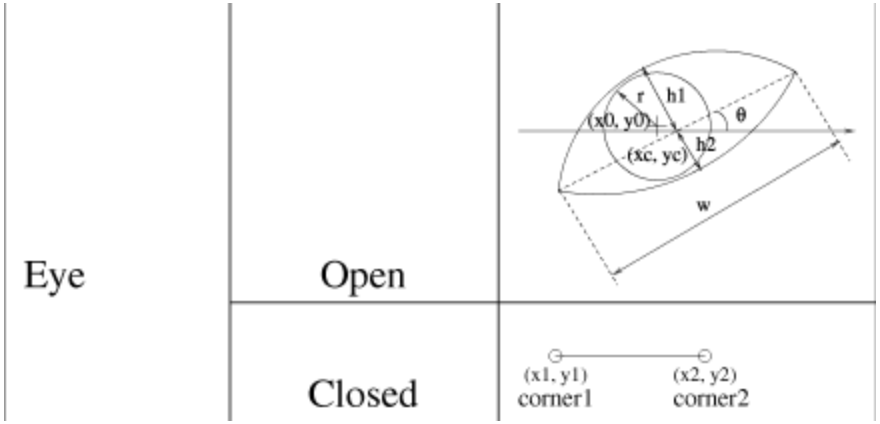
*lip center position (**xc, yc**)*
*lip shape (**h1, h2 and w**)*
*lip orientation (**θ**)*

➔ *dark mouth line*

- Tracking: uses color, shape, motion
  (see Tian et al., Asian Conf. on Computer Vision, 2000)

Permanent Features – Eye

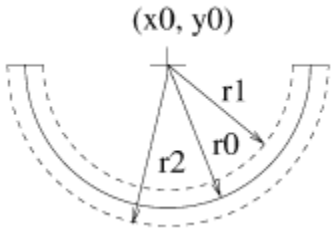
- **Open eye**
  - circle with 3 parameters to model the iris **(x0 , y0 , r)**
  - two parabolic arcs with 6 parameters **(xc , yc , h1 , h2 , w, θ)** to model boundaries of the eye
  - Adjusted manually in the first frame by moving six points for each eye



- **Closed eye**
  - template with 4 parameters

Permanent Features – Eye (iris)

- *half circle iris mask*
- *(x0 , y0) iris center*
- *r0 radius, r1 min, r2 max*



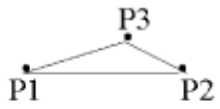
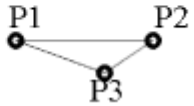
- r0 determined in the first frame
- in-/decreasing (δr): r0 to min (r0 – δr) and max (r0 + δr)
- **iris found conditions:**
  - *edges are at their max*
  - *average changing intensity < threshold*

- Tracking-Details: Tian et al., Int. Conf. on Face and Gesture Recognition, 2000

Permanent Features – Brows and cheeks

- triangular template with 6 parameters
  P1: **(x1, y1)**,
  P2: **(x2, y2)** and
  P3: **(x3, y3)**

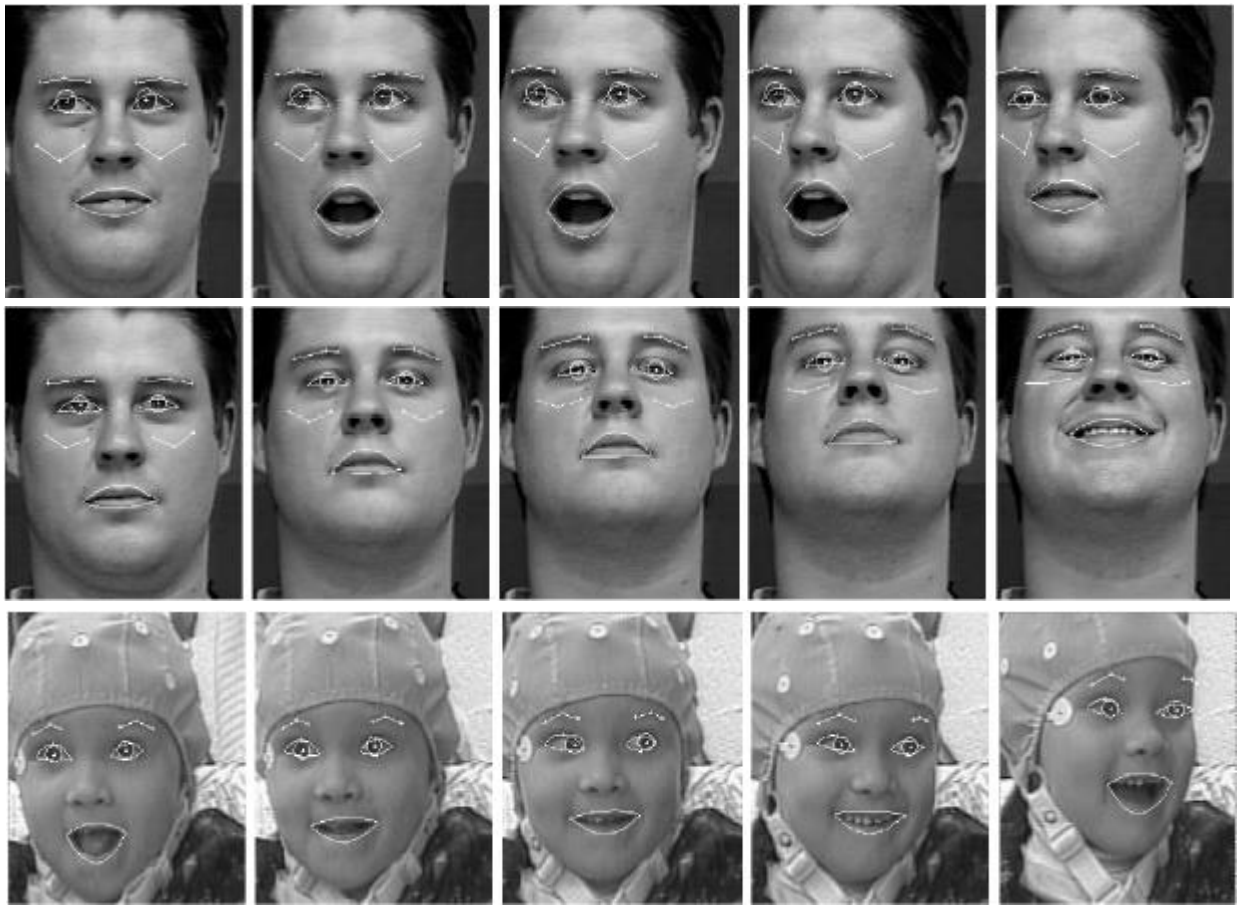
- tracking templates with Lucas-Kanade algorithm

Brow	Present	
Cheek	Present	

Permanent Feature Tracking Results for Different Expressions of Same Subject

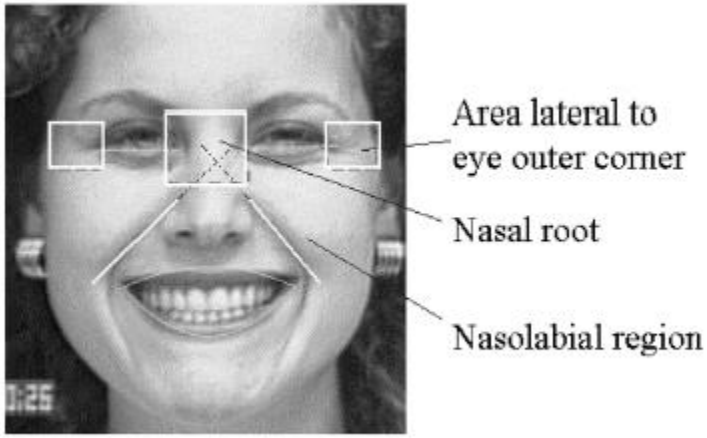
Permanent Feature Tracking Results for Different Subjects

Permanent Feature Tracking Results with Head Motions



Transient Features

- **furrows** and **wrinkles** appear **perpendicular** to the direction of the motion of the activated muscles



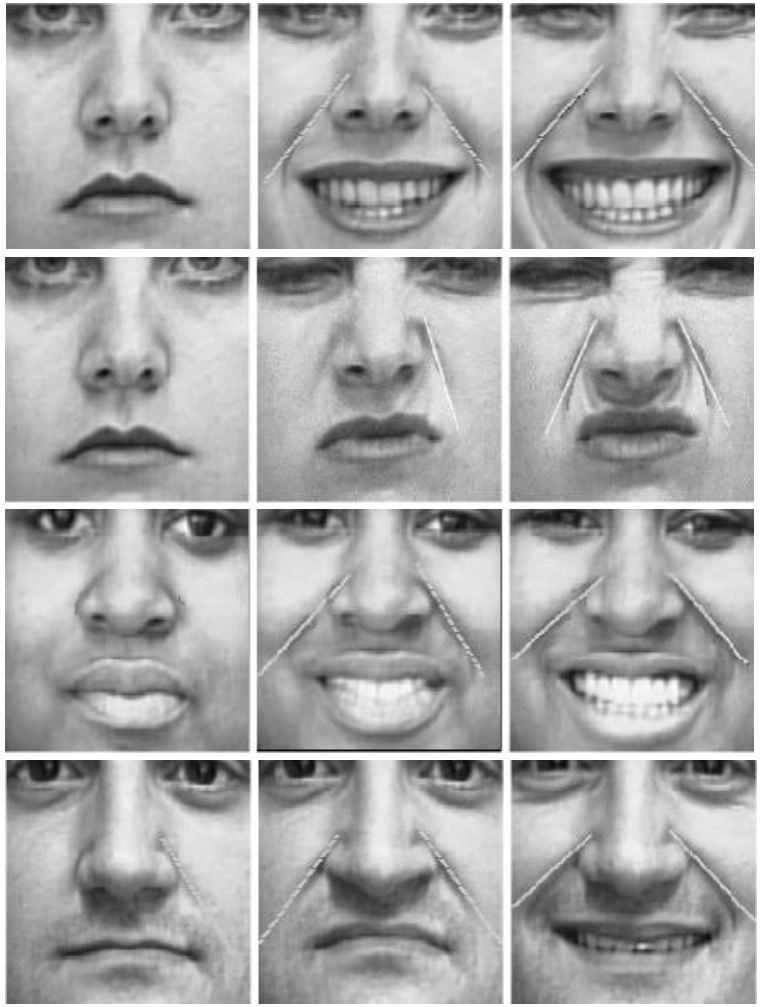
- (permanent with age) crow's-feet wrinkles around the outside corners of the eyes.

- **Classification:**
  - present (appear, deepen or lengthen)
  - absent

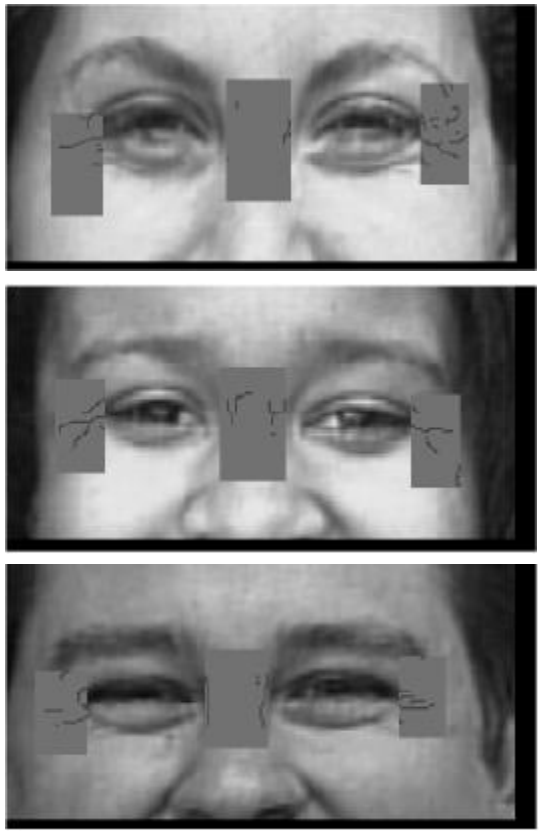
Transient Features Detection

- **Canny edge detector**
  - amount & orientation of furrows
- **Nasal root / crow's-feet wrinkles**
  - E = #edge pixels (current frame)
  - E0 = #edge pixels (first frame)
  - E/E0 > threshold ➔ furrows are present
- **Nasolabial furrows**
  - connected edge pixels > threshold ➔ furrows are present

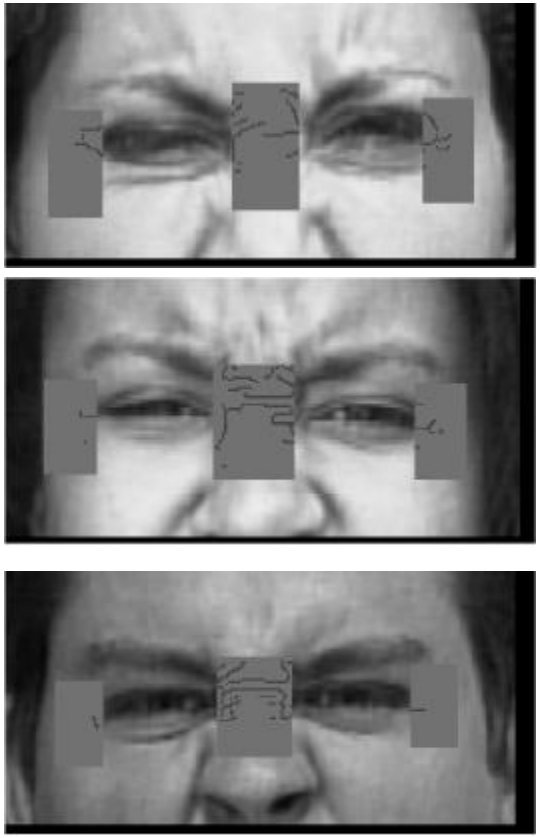
Nasolabial Furrow Detection Results



Nasal Root and Crow's-feet Wrinkle Detection



Crow's feet wrinkle detection



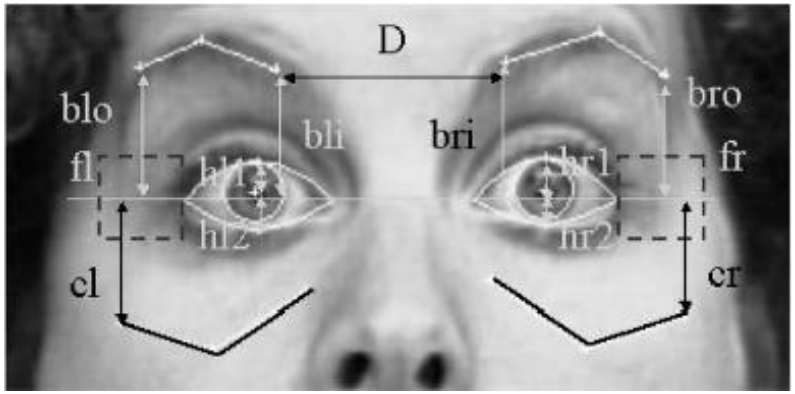
Nasal root wrinkles appear

Facial Feature Representation

- *Face coordinate system*
  - x = line between inner corners of eyes
  - y = perpendicular to x

- *Group facial features*
  - **upper face features**
    - *15 parameters*
  - **lower face features**
    - *9 parameters*

Upper Face Feature Representation



hl(hl1+hl2), hr(hr1+hr2) : height of left/right eye

fl / fr : left/right crow's-feet wrinkle areas

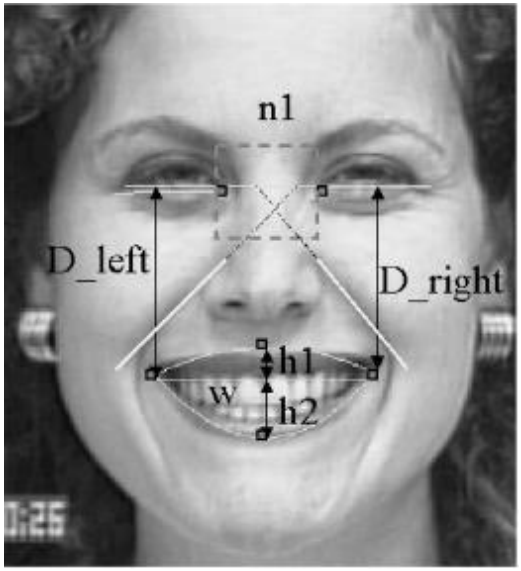
bli / bri : motion of inner part of left/right brow

blo / bro : motion of outer part of left/right brow

cl / cr : motion of left/right cheek

Permanent features (Left and right)			Other features
Inner brow motion (r_{binner})	Outer brow motion (r_{bouter})	Eye height ($r_{eheight}$)	Distance of brows (D_{brow})
$r_{binner} = \frac{bi-bi_0}{bi_0}$. If $r_{binner}>0$, Inner brow move up.	$r_{bouter} = \frac{bo-bo_0}{bo_0}$. If $r_{bouter}>0$, Outer brow move up.	$r_{eheight} = \frac{(h1+h2)-(h1_0+h2_0)}{(h1_0+h2_0)}$. If $r_{eheight}>0$, Eye height increases.	$D_{brow} = \frac{D-D_0}{D_0}$. If $D_{brow}<0$ Two brows drawn together.
Eye top lid motion (r_{top})	Eye bottom lid motion (r_{btm})	Cheek motion (r_{cheek})	crows-feet wrinkles $W_{left/right}$
$r_{top} = \frac{h1-h1_0}{h1_0}$. If $r_{top} > 0$, Eye top lid move up.	$r_{btm} = -\frac{h2-h2_0}{h2_0}$. If $r_{btm} > 0$, Eye bottom lid move up.	$r_{cheek} = -\frac{c-c_0}{c_0}$. If $r_{cheek} > 0$, Cheek move up.	If $W_{left/right} = 1$, Left/right crows feet wrinkle present.

Lower Face Feature Representation



h1, h2 : top/bottom lip heights

w : lip width

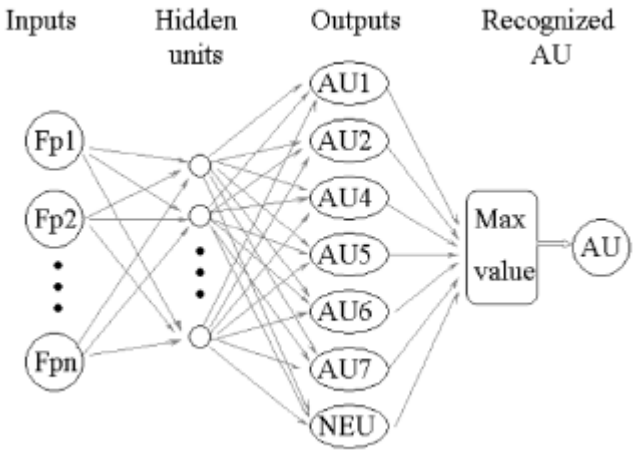
n1 : nasal root area

D_left, D_right : distance btw. the left/right lip corner and eye inner corners line

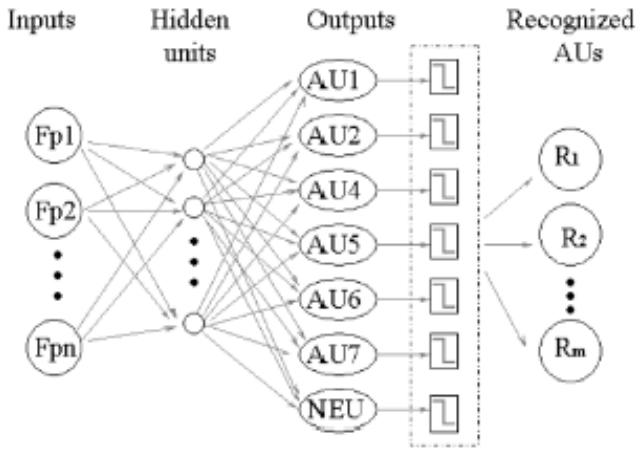
Permanent features		
Lip height (r_{height})	Lip width (r_{width})	Left lip corner motion (r_{left})
$r_{height} = \frac{(h1+h2)-(h1_0+h2_0)}{(h1_0+h2_0)}$. If $r_{height}>0$, lip height increases.	$r_{width} = \frac{w-w_0}{w_0}$. If $r_{width}>0$, lip width increases.	$r_{left} = -\frac{D_{left}-D_{left0}}{D_{left0}}$. If $r_{left}>0$, left lip corner moves up.
Right lip corner (r_{right})	Top lip motion (r_{top})	Bottom lip motion(r_{btm})
$r_{right} = -\frac{D_{right}-D_{right0}}{D_{right0}}$. If $r_{right}>0$, right lip corner moves up.	$r_{top} = -\frac{D_{top}-D_{top0}}{D_{top0}}$. If $r_{top}>0$, top lip moves up.	$r_{btm} = -\frac{D_{btm}-D_{btm0}}{D_{btm0}}$. If $r_{btm}>0$, bottom lip moves up.
Transient features		
Left nasolibial furrow angle (Ang_{left})	Right nasolibial furrow angle (Ang_{right})	State of nasal root wrinkles (S_{nosew})
Left nasolibial furrow present with angle Ang_{left}.	Left nasolibial furrow present with angle Ang_{right}.	If $S_{nosew} = 1$, nasal root wrinkles present.

AU Recognition by Neural Networks

- three layer neural networks (one hidden layer)
- standard back-propagation method
  - separate networks for upper- / lower face



single AUs in the upper face



for combinations of AUs in the upper face

Experiments – Ekman-Hager database

- upper face AUs occurring singly or in combination
  - **236** image sequences (23 subjects)
    - **99** sequences with **single AUs**
    - **137** sequences with **AU combinations**
  - **186** sequences for **training** (9 subjects)
  - **50** sequences for **testing** (14 subjects)
    - Testing was done with novel faces

Experiments – Cohn-Kanade database

- lower face AUs occurring singly or in combination
  - **463** image sequences
  - **400** sequences for **training**
  - **63** sequences for **testing**
    - **10 single AUs**
    - **NEUTRAL**
    - **11 AU combinations**
      - **(from 32 subjects)**
  - Testing was done with novel faces and limited head motion

Results

		Test databases		Train databases
		Cohn-Kanade	Ekman-Hager	
Recognition Rate	upper face	**93.2%**	96.4% (Table 9)	Ekman-Hager
	lower face	96.7% (Table 10)	**93.4%**	Cohn-Kanade

Automatic Analysis of Spontaneous Facial Behavior

M.S. Bartlett, B. Braathen, G. Littlewort-Ford, J. Hershey, I. Fasel, T. Marks, E. Smith, T.J. Sejnowski, J.R. Movellan
Tech. Report, UCSD MPLab TR 2001.08, Oct. 2001

- 3-D pose estimation and warping techniques to reduce image variability due to changes in pose.
  - (Not discussed today)

- Machine learning techniques performed directly on the warped images or on biologically inspired representations of these images.

Image Data

- 17 subjects from a deception study.
- 11 Caucasians, 3 African American, 3 Asians.
- Digitized video of 30 frames/sec. frame rate
- 300 gigabytes of 640x480 color images.
- Approximately one minute of video was FACS coded for each subject.

- → very challenging data
- spontaneous facial actions
- Imaging conditions not very controlled:
  - head orientation varies
  - Lower resolution than in Cohn-Kanade DB

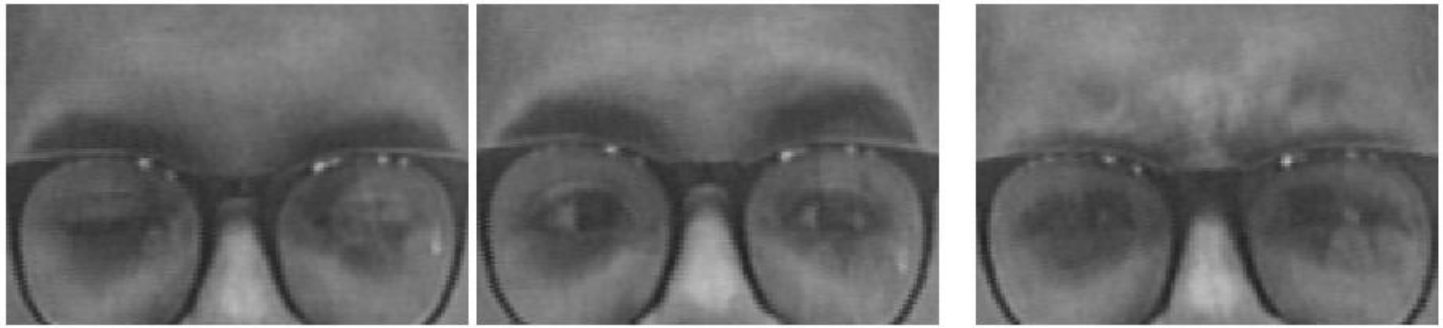
The “Theft“ Scenario

- For half the subjects, a drawer contained $50, for the other half the drawer was empty.
- If the drawer contained money, subjects were told that they had the choice of taking it or not.
- Afterwards they would have to convince an experimenter that either the drawer had no money or they did not take it.
- If they convince the experimenter, then they can keep the money.
- If the experimenter thought they were lying, then they would have to be subjected to a very uncomfortable loud noise for 1 minute.
- Subjects were given a sample of the noise at the beginning of the experiment. Subjects caught lying were not actually punished.

- → data contains **spontaneous** facial movements
- (Frank & Ekman: „deception can be reliably detected from facial actions“)

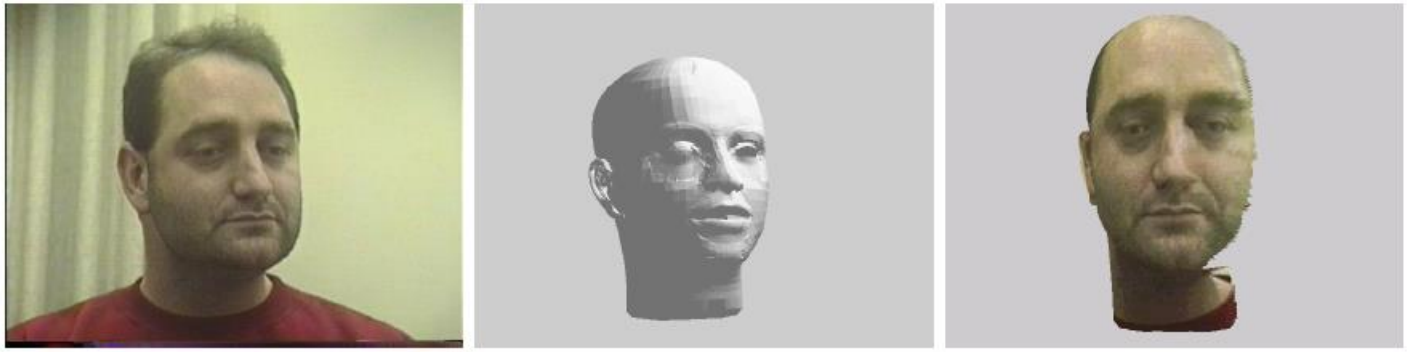
The Task

- Detection and discrimination of
  - Action unit 45 (blinks)
  - Action units 1+2 (brow raise)
  - Action unit 4 (brow lower)



Example of action unit 45 (left), 1+2 (center) and 4 (right)

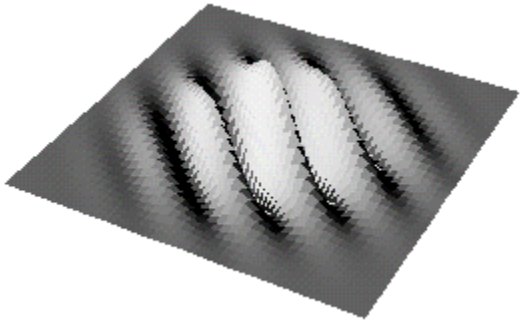
Generation of Image Representation

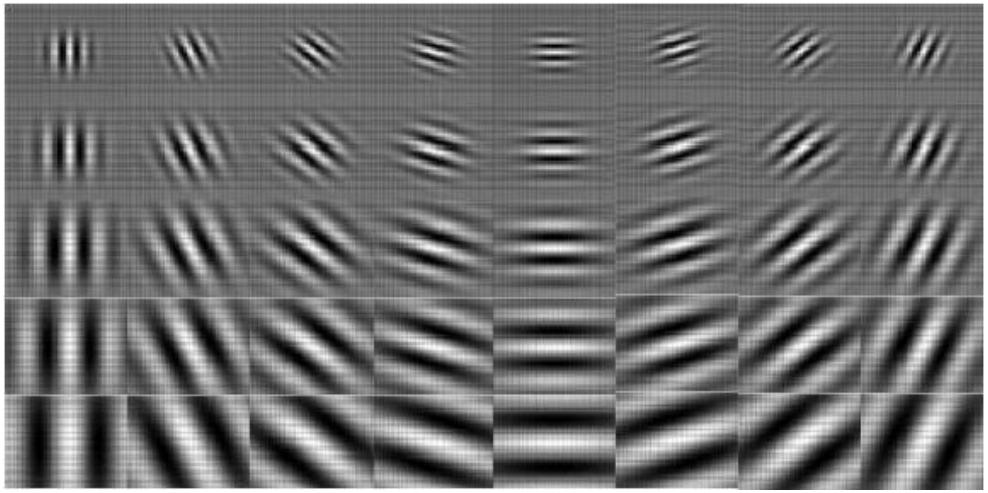


- Once 3D pose was estimated, faces were rotated to frontal and warped to a canonical face geometry.
- Images were scaled and aligned to 192x132 pixels resolution with 105 pixels between the eyes.
- *Difference images* were obtained by subtracting a neutral expression frame from the subsequent frames.
- The difference images were passed through the *Gabor filters* at 5 spatial frequencies and 8 orientations

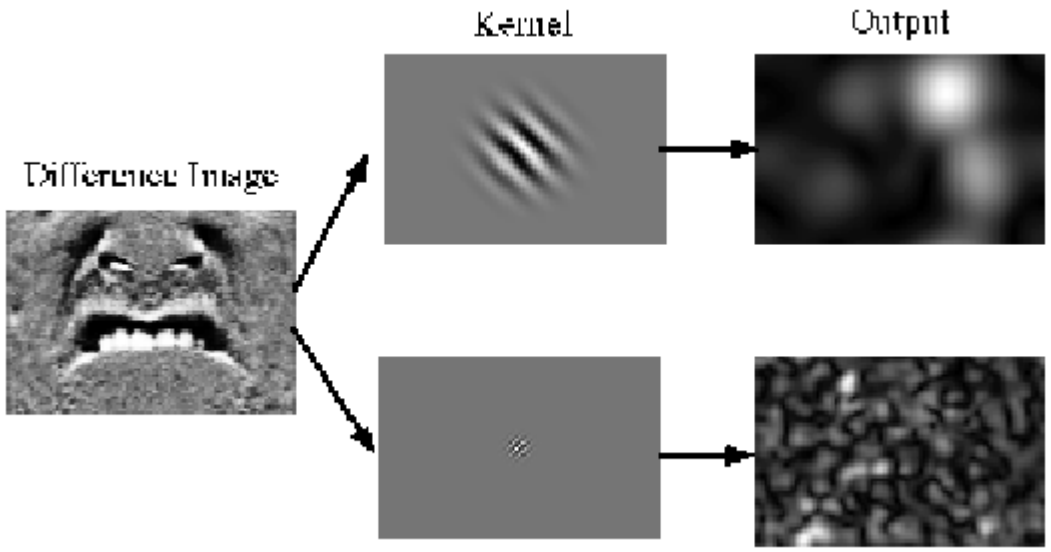
Gabor Filters

- 2-D sine waves modulated by a Gaussian envelope.
- Good models of the receptive fields found in simple cells of the primary visual cortex.
- 2D Gabor filter –Spatial domain:





Example Image Decomposition



Classification

- Gabor representations comprised the input to the Support Vector Machines (SVMs).

- The output of the classifiers are fed to a bank of Hidden Markov Models (HMMs) to take advantage of dynamic information.

Results (1)

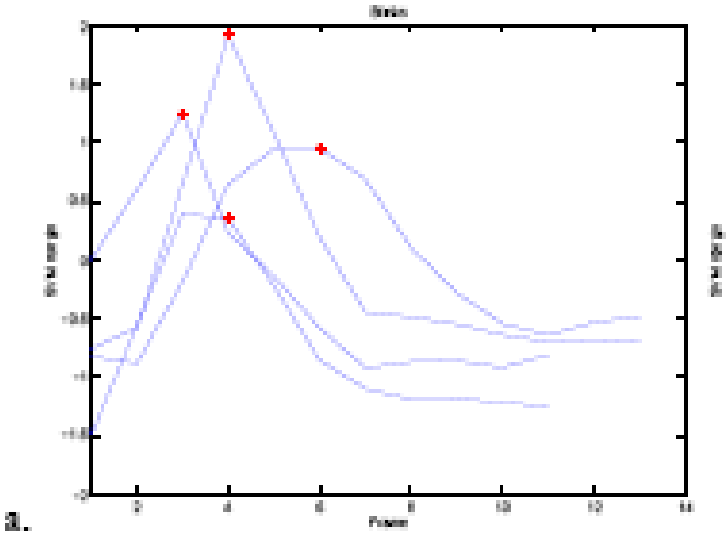
- SVM trained to discriminate blinks vs. non-blinks
  - Input: Gabor representation / difference images
  - Kernel: Linear vs. non-linear
  - Trained and tested on peak frames
  - Results using Gabor-representation: 94.3 % correct
  - Results using Difference images: 95.9 % correct

	Gabor-Features	Difference-Images
Non-linear kernel	94.3 %	**95.9** %
Linear kernel	93.5 %	78.3 %

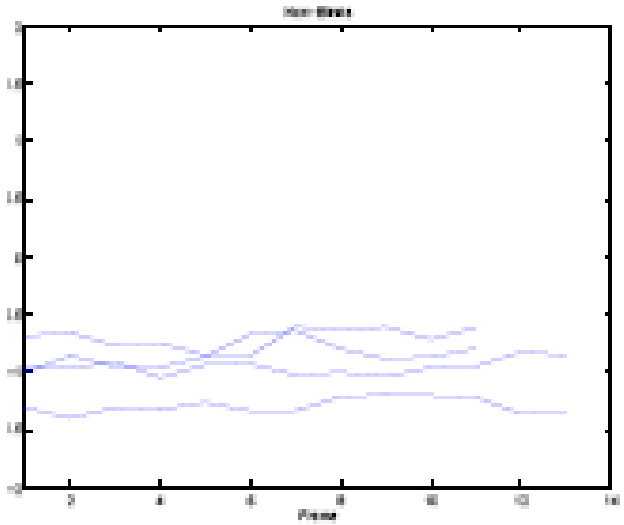
Blink vs. non-blink detection

Example SVM Outputs

Blinks



Non-Blinks



- SVM margins over time, for blinks and non-blinks
- Examples from one subject
- The star indicates the manually coded AU peak

HMMs trained on SVM-output

- Confusion matrix for HMM trained on SVM outputs. Overall agreement: 66.9%. Agreement omitting Brow Lower: 89.5%.

	Automated System		
Manual Coder	Brow Raise	Brows Lower	Matched Random Sequences
Brow Raise	39	5	4
Brow Lower	2	6	6
Matched Random Sequences	5	19	38

- Summary of Results

	HMM	
	on Gabor	on SVM
Blink vs. Non-blink	95.7	98.1
Brow Raise vs. Matched Random Sequences	-	90.6
Brow Lower vs. Matched Random Sequences	-	75.0
Brow Raise vs. Brow Lower	-	93.5
Brow Raise vs. Lower vs. Random	70.2	66.9

Summary

- What is facial expression analysis
- Typical structure of recognition system
- Levels of description
  - Six basic emotions vs. Facial Action Units
- Used features
  - Geometric features vs. appearance-based features
- Types of classification
  - Frame based: ANN, SVM
  - Sequence based: HMMs, Rule-based
  - Hybrid
- Data
  - Acted vs. spontaneous expressions
  - Controlled vs. uncontrolled scenario (illumination, pose, etc.)

References

- **Facial Expression Analysis**
  Y. Tian, T. Kanade, J. Cohn
  Handbook of face recognition, S.Z. Li & A.K. Jain, ed., Springer, Oct. 2003.

- **Recognizing Action Units for Facial Expression Analysis**
  Y. Tian, T. Kanade, J. Cohn
  IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp. 97-115, Feb. 2001

- **Automatic Analysis of Spontaneous Facial Behavior**
  M.S. Bartlett, B. Braathen, G. Littlewort-Ford, J. Hershey, I. Fasel, T. Marks, E. Smith, T.J. Sejnowski, J.R. Movellan
  Tech. Report, USCD MPLab TR 2001.08, Oct. 2001

Weitere Tips (für die Freizeit)

- Paul Ekman und Wallace V. Friesen, **Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions**, Malor Books
  - Und andere …
- Malcom Gladwell, **Blink: The Power of Thinking Without Thinking**, Little, Brown and Company
  - Bspw. Kapitel “The Naked Face”
  - siehe auch http://www.gladwell.com/2002/2002_08_05_a_face.htm
- Interviews with Paul Ekman (googlen, YouTube, …)
  - http://globetrotter.berkeley.edu/people4/Ekman/ekman-con0.html
  - YouTube: **„Conversations with History: Paul Ekman“**