Karlsruhe Institute of Technology

FG 2011 Facial Expression Recognition and Analysis Challenge (FERA2011)

Tobias Gehrig | February 3, 2011

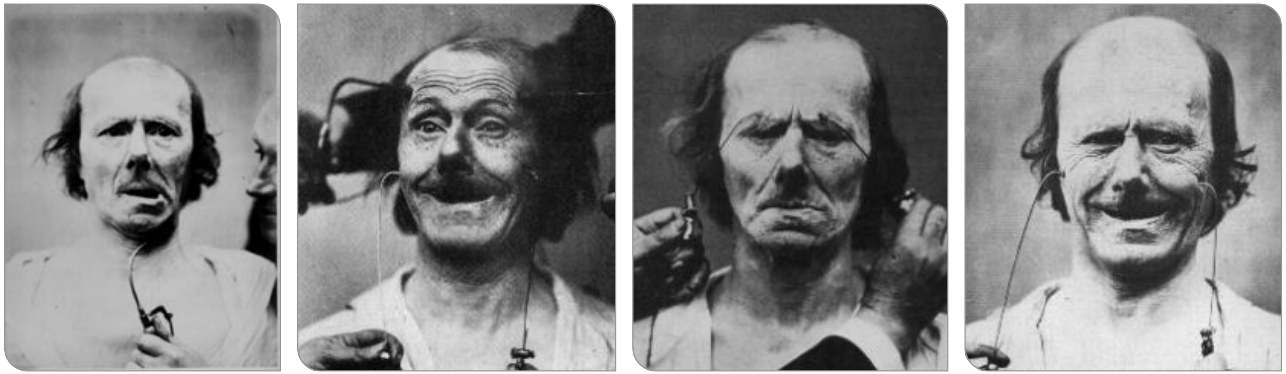
INSTITUTE FOR ANTHROPOMATICS, FACIAL IMAGE PROCESSING AND ANALYSIS YIG

Overview

- Motivation
- FERA 2011
  - Tasks
  - Challenge Data
- Approaches
  - LBP-based baseline system
  - Our DCT-based system
- Results
- Conclusion and Future Work

Motivation

Why facial expression analysis?

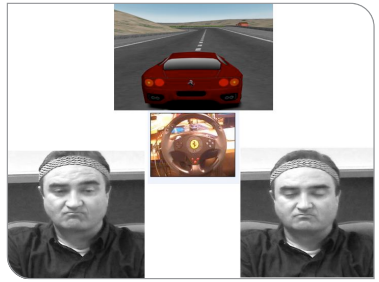


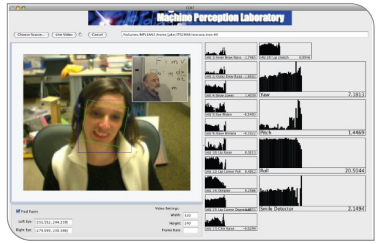
- facial expressions are naturally used by humans to communicate their emotions, feelings, opinions, intentions, and cognitive states with each other
- Expression of emotion through faces is faster then describing the affective state with words

Motivation

Applications [2]

- human-computer interaction
- psychological studies
- driver safety [4]
- online tutoring systems [5]
- pain or stress detection
- assistance systems for autistic persons [3]





Institute for Anthropomatics
Facial Image Processing and Analysis YIG

Motivation

Problems of previous work

- similar processing systems
- data sparseness
- lack of standardized evaluation procedures
- low comparability of results

FG 2011 Facial Expression Recognition and Analysis Challenge (FERA2011)

Goals

- standardize evaluation procedures for automatic facial expression analysis
- make results comparable

Tasks

- Emotion recognition
- FACS action unit detection

Website

- `http://sspnet.eu/fera2011`

Tasks

Emotion recognition

- recognize emotions on video basis
- 5 discrete, mutually-exclusive emotion categories:
  - Anger
  - Fear
  - Joy
  - Relief
  - Sadness

Tasks

FACS action unit detection

- detect 12 action units on frame-by-frame basis

Upper Face Action Units				
AU 1	AU 2	AU 4	AU 6	AU 7
				
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Cheek Raiser	Lid Tightener

Lower Face Action Units						
AU 10	AU 12	AU 15	AU 17	AU 18	AU 25	AU 26
						
Upper Lip Raiser	Lip Corner Puller	Lip Corner Depressor	Chin Raiser	Lip Puckerer	Lips Part	Jaw Drop

Institute for Anthropomatics
Facial Image Processing and Analysis YIG

Challenge Data

GEMEP-FERA

- derived from the GEneva Multimodal Emotion Portrayals (GEMEP) database [1] [1]
- 10 professional French-speaking actors (5 males, 5 females)
- displaying a range of expressions, while uttering a meaningless phrase, or the word 'Aaah'
- directed by and interacting with a professional stage director
- videos of 720×576 at 25 fps
- between 1 and 4 seconds long sequences
- data provided as a strictly divided training and test set
- training set: 7 subjects (3 male, 4 female)
- test set: 6 subjects (3 male, 3 female) including 3 subjects (1 male, 2 female) from training set

[1] [1] Tanja Bänziger and Klaus R. Scherer. "Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus", pages 271–294. Oxford University Press, Oxford, England, 2010.

Challenge Data

Emotion detection

- interlaced data
- training data:
  - 155 videos
  - labels per video
- test data:
  - 134 videos
  - same kind as training data
  - half of the subjects appear in training data

training sample statistics

	pos.	neg.
anger	1686	7334
fear	1467	7553
joy	1895	7125
relief	1854	7166
sadness	2118	6902

Challenge Data

FACS action unit detection

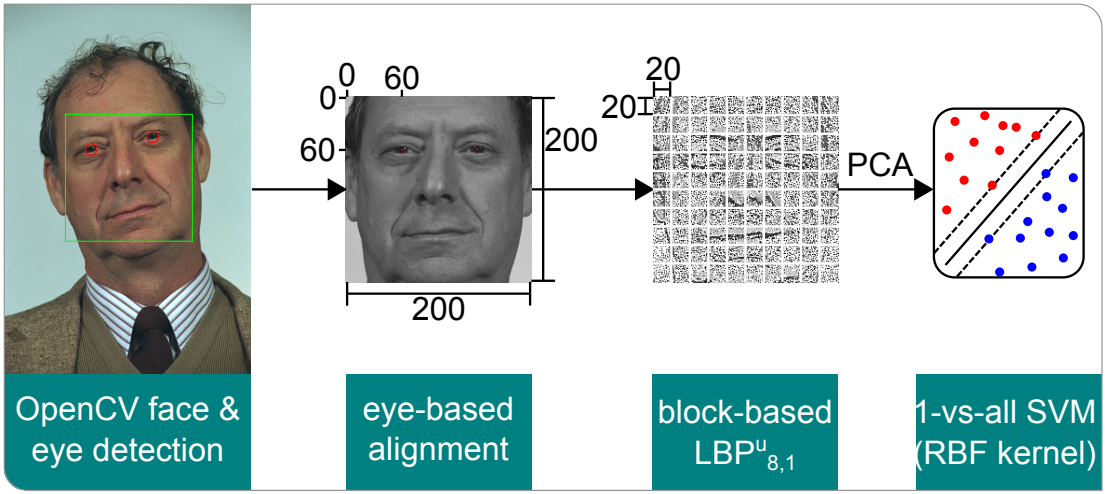
- non-interlaced data
- training data:
  - 87 videos (5264 frames)
  - labelled frame-by-frame
  - AU25 and AU26 are not labeled if there is speech (AD50)
- test data:
  - 71 videos
  - same kind as training data
  - half of the subjects appear in training data

training sample statistics

	pos.	neg.
AU1	1564	3700
AU2	1619	3645
AU4	1315	3949
AU6	1846	3418
AU7	2124	3140
AU10	2034	3230
AU12	2726	2538
AU15	998	4266
AU17	822	4442
AU18	419	4845
AU25	913	514
AU26	567	827
AD50	3937	

Baseline system

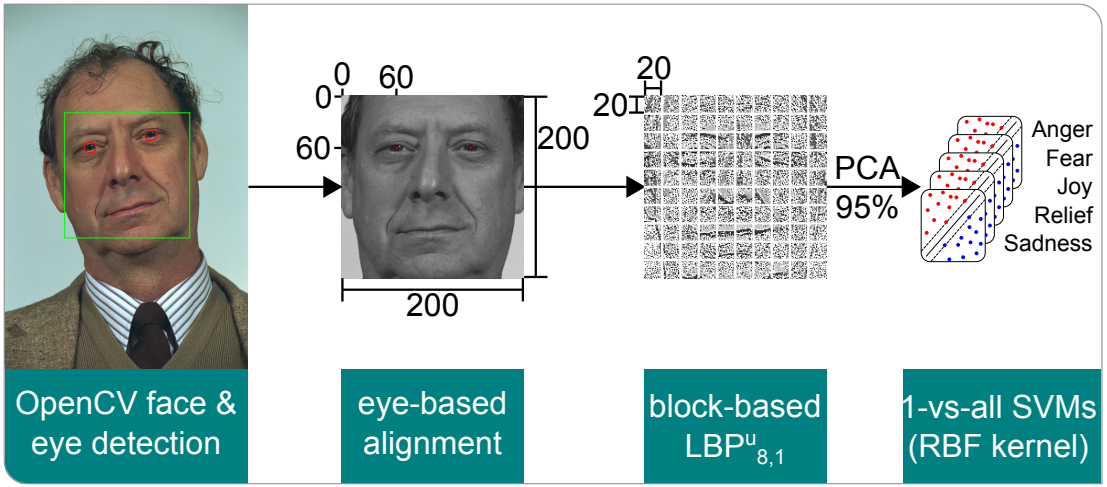
Common setup



- training:
  - pre-processed images were manually verified and removed if incorrect
  - grid search over C and γ using 5-fold subject-independent cross-validation
- testing:
  - if no eyes were found, unaligned faces are used

Baseline system

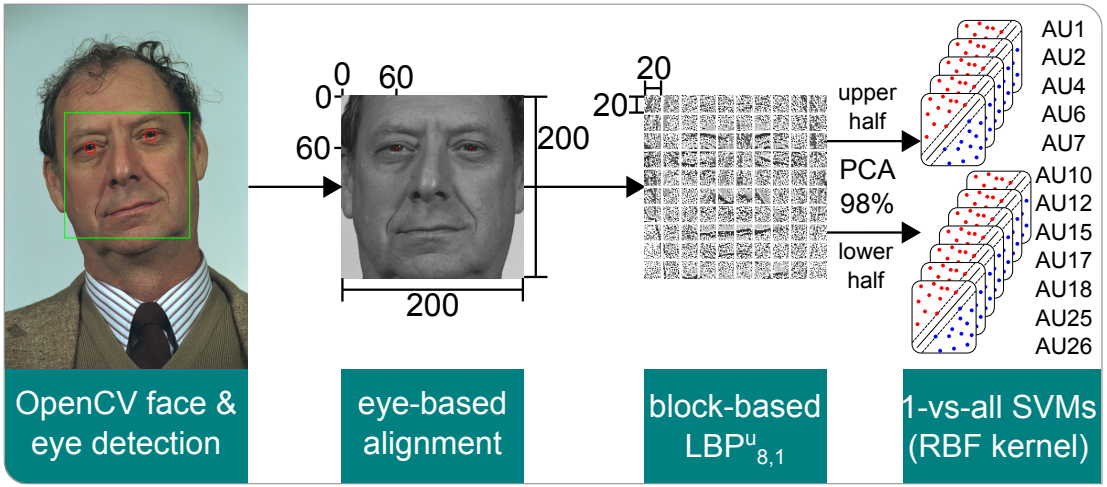
Emotion detector



- training:
  - positive samples: all frames of videos labeled with corresponding emotion
  - negative samples: the rest
- testing:
  - per frame: emotion corresponding to classifier with highest decision function
  - per video: majority voting over frame classifications

Baseline system

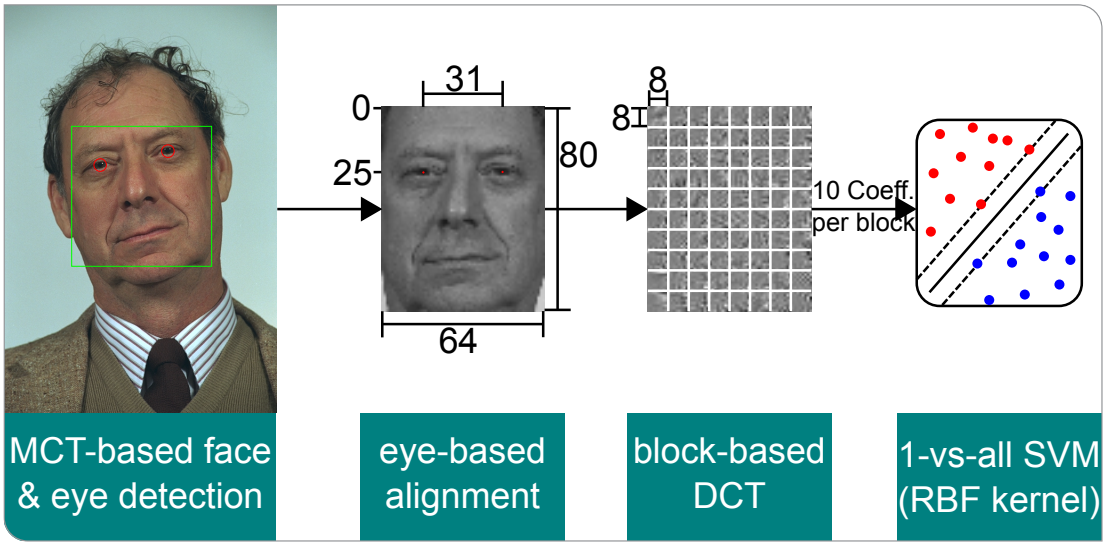
AU detector



- training:
  - from each video only one frame per AU combination present was used
- testing:
  - AUs set to be absent if no face was found

Our DCT-based system

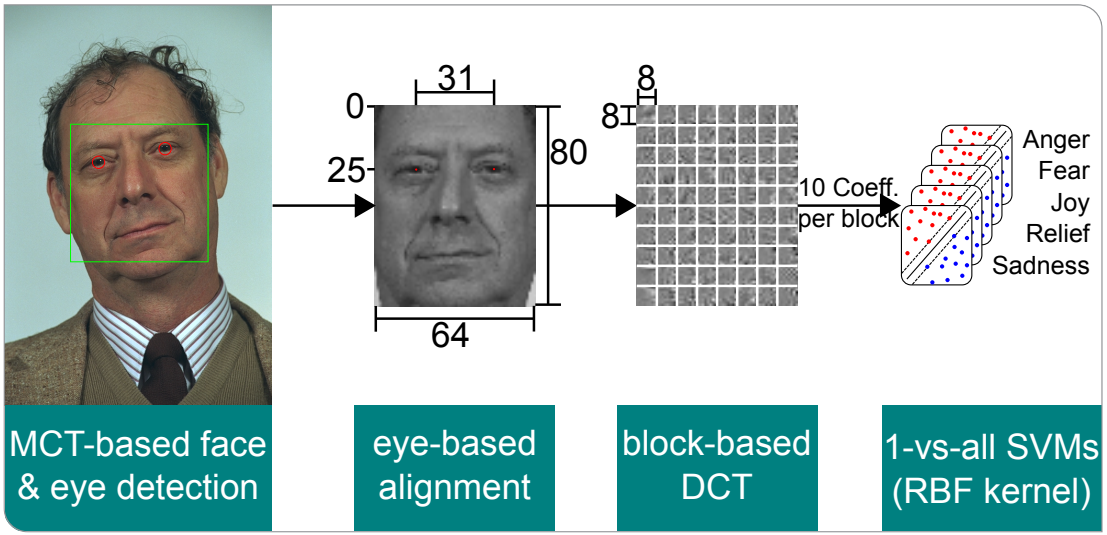
Common setup



- training:
  - only frames with valid face and eye detections
  - grid search over C and γ using 5-fold subject-independent cross-validation
- testing:
  - if no eyes were found, unaligned faces are used

Our DCT-based system

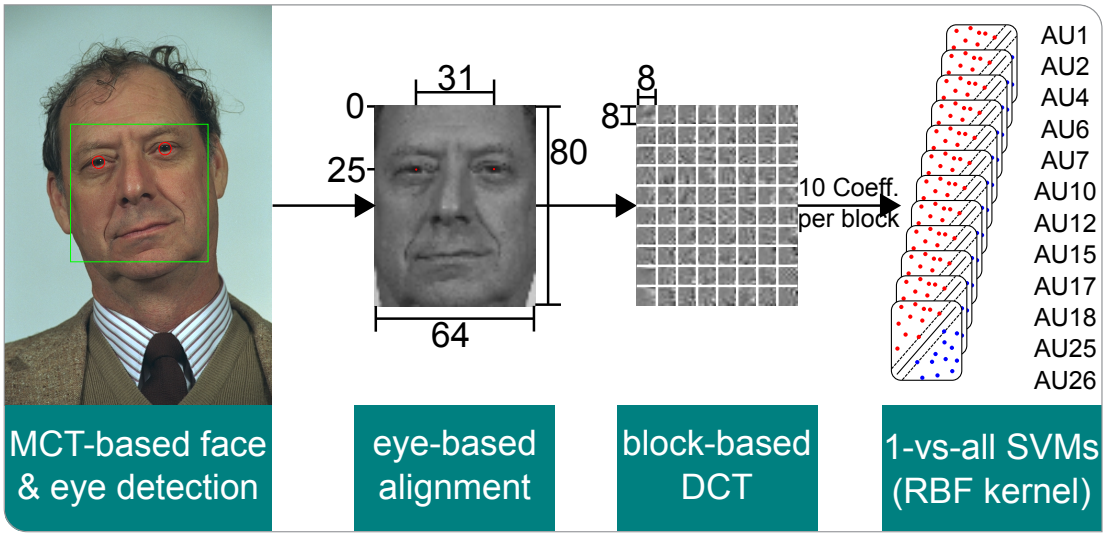
Emotion detector



- training:
  - model for probability estimates
  - grid search: $C = 2^k$ and $\gamma = 2^l$ with $k = -3, \ldots, 1$ and $l = -16, \ldots, -7$
  - positive samples: all frames of videos labeled with corresponding emotion
  - negative samples: the rest
- testing:
  - per frame: emotion corresponding to classifier with highest probability
  - per video: majority voting over frame classifications

Our DCT-based system

AU detector



- training:
  - grid search: $C = 10^{-8} \cdot 2^k$, $\gamma = 2^l$ with $k = 0, \ldots, 31$, $l = -15, \ldots, 1$
  - positive samples: all frames for which AU is labeled as active
  - negative samples: the rest
  - balanced training set
- testing:
  - AUs set to be absent if no face was found

Baseline Results

Classification rates for emotion recognition

	Person independent	Person specific	Overall
anger	86%	**92%**	89%
fear	7%	**40%**	20%
joy	70%	**73%**	71%
relief	31%	**70%**	46%
sadness	27%	**90%**	52%
Average	44%	**73%**	56%

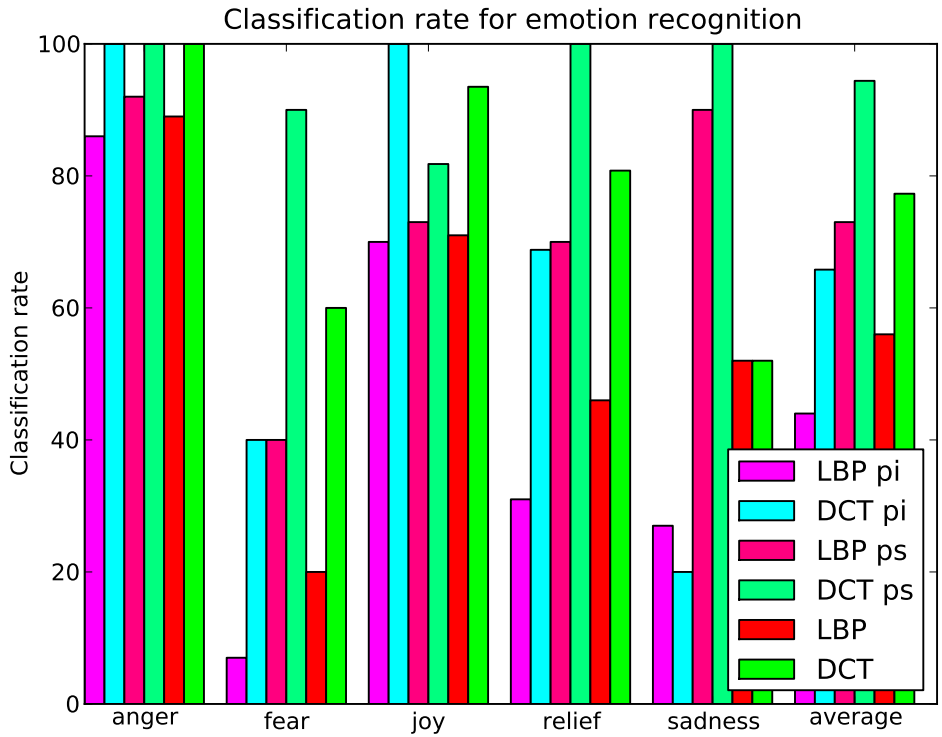
DCT Results

Classification rates for emotion recognition

	Person independent	Person specific	Overall
anger	**100.0%**	**100.0%**	**100.0%**
fear	40.0%	**90.0%**	60.0%
joy	**100.0%**	81.8%	93.5%
relief	68.8%	**100.0%**	80.8%
sadness	20.0%	**100.0%**	52.0%
Average	65.8%	**94.4%**	77.3%

Results

Comparison of classification rates for emotion recognition



DCT Results

Confusion matrix for emotion recognition - person independent

prediction \ truth	Anger	Fear	Joy	Relief	Sadness
Anger	14	4	0	1	11
Fear	0	6	0	2	1
Joy	0	5	20	2	0
Relief	0	0	0	11	0
Sadness	0	0	0	0	3

- confusions of fear with anger are also hard to distinguish for a human
- confusion of relief with anger is possibly due to the actor looking up from a downward oriented head
- confusions of relief with fear are possibly due to an actor starring at the camera
- confusions of relief with joy is also hard to distinguish for a human from just the face

DCT Results

Confusion matrix for emotion recognition - person independent

prediction \ truth	Anger	Fear	Joy	Relief	Sadness
Anger	14	4	0	1	11
Fear	0	6	0	2	1
Joy	0	5	20	2	0
Relief	0	0	0	11	0
Sadness	0	0	0	0	3

- confusions of sadness with anger and fear seems to be due to relatively neutral faces

DCT Results

Confusion matrix for emotion recognition - person specific

prediction \ truth	Anger	Fear	Joy	Relief	Sadness
Anger	13	1	1	0	0
Fear	0	9	0	0	0
Joy	0	0	9	0	0
Relief	0	0	1	10	0
Sadness	0	0	0	0	10

- confusion of fear with anger is due to extrem yaw changes and false face detection results
- confusion of joy with anger is due to occlusions of the face and false face detection results (those frames with correct eye positions are also correctly classified as joy)
- confusion of joy with relief is due to extrem pitch changes, which most often occur in relief videos

DCT Results

Confusion matrix for emotion recognition - overall

prediction \ truth	Anger	Fear	Joy	Relief	Sadness
Anger	27	5	1	1	11
Fear	0	15	0	2	1
Joy	0	5	29	2	0
Relief	0	0	1	21	0
Sadness	0	0	0	0	13

Baseline Results

F1 measures for AU detection

	Person independent	Person specific	Overall
AU1	**87.1%**	49.0%	77.6%
AU2	**84.2%**	57.2%	76.5%
AU4	49.8%	**66.0%**	57.7%
AU6	**85.4%**	55.3%	79.1%
AU7	**68.5%**	68.0%	68.3%
AU10	56.4%	**61.5%**	58.6%
AU12	85.9%	**86.4%**	86.0%
AU15	14.6%	**41.1%**	25.2%
AU17	**65.2%**	41.3%	55.9%
AU18	42.3%	**51.8%**	47.5%
AU25	79.9%	**82.8%**	81.1%
AU26	50.2%	**63.6%**	55.9%
Average	**64.1%**	60.3%	64.1%

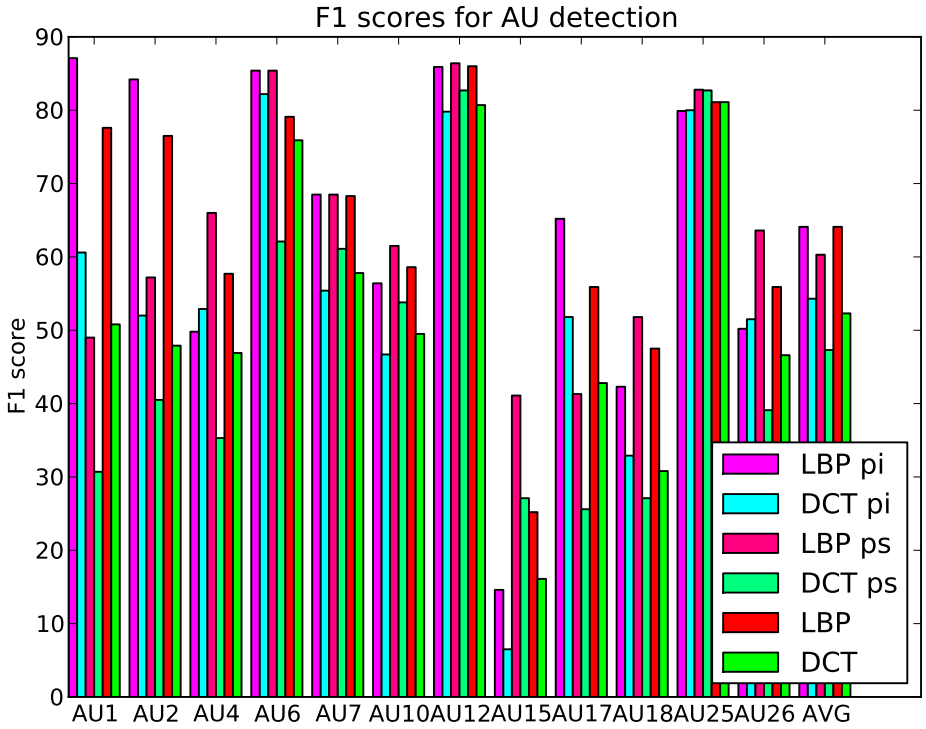
DCT Results

F1 measures and areas under the ROC curve for AU detection

	Person independent		Person specific		Overall	
	F1	2AFC	F1	2AFC	F1	2AFC
AU1	**60.6%**	51.2%	30.7%	**53.1%**	50.8%	52.6%
AU2	**52.0%**	61.3%	40.5%	**65.6%**	47.9%	63.2%
AU4	**52.9%**	**58.1%**	35.3%	47.6%	46.9%	54.9%
AU6	**82.2%**	**88.1%**	62.1%	82.2%	75.9%	85.2%
AU7	55.4%	**68.8%**	**61.1%**	64.6%	57.8%	67.3%
AU10	46.7%	49.9%	**53.8%**	**57.9%**	49.5%	53.3%
AU12	79.8%	81.9%	**82.7%**	**90.4%**	80.7%	84.7%
AU15	6.5%	39.5%	**27.1%**	**73.3%**	16.1%	55.5%
AU17	**51.8%**	**62.0%**	25.6%	60.6%	42.8%	61.1%
AU18	**32.9%**	**79.9%**	27.1%	59.3%	30.8%	71.8%
AU25	80.0%	**62.5%**	**82.7%**	58.4%	81.1%	60.4%
AU26	**51.5%**	**66.7%**	39.1%	50.8%	46.6%	58.1%
Average	**54.3%**	**64.2%**	47.3%	63.7%	52.3%	64.0%

Results

Comparison of F1 scores for AU detection



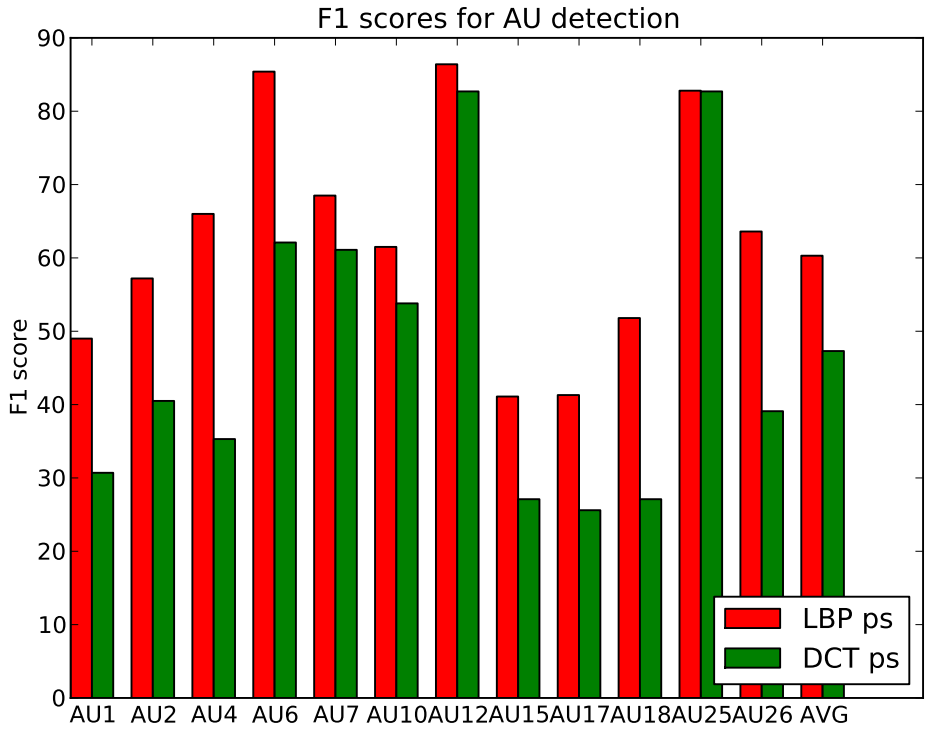
Results

Comparison of F1 scores for person independent AU detection

Institute for Anthropomatics
Facial Image Processing and Analysis YIG

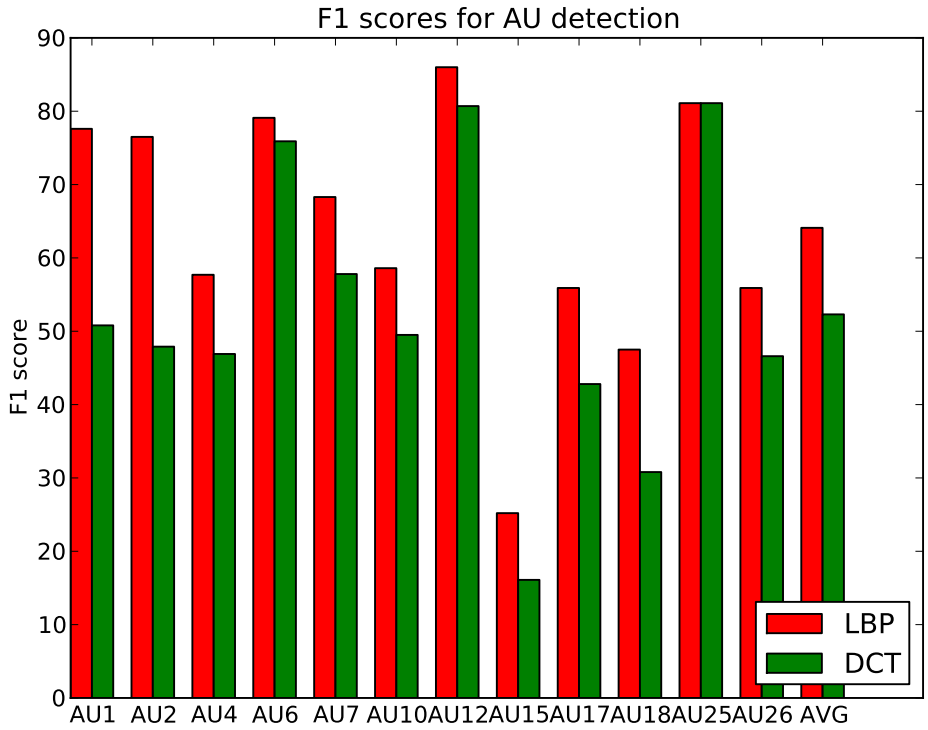
Results

Comparison of F1 scores for person specific AU detection



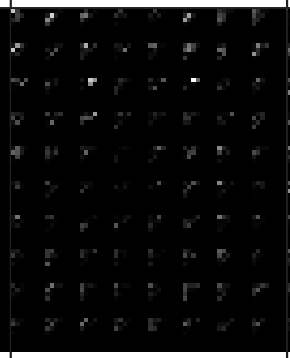
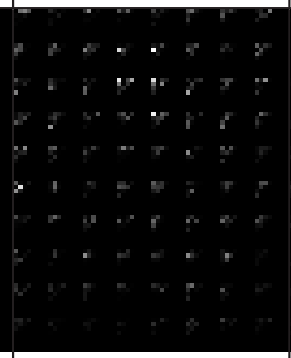
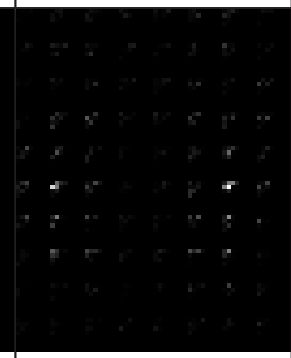
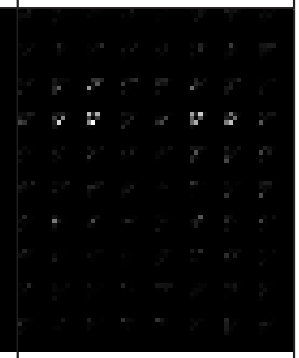
Results

Comparison of F1 scores for overall AU detection



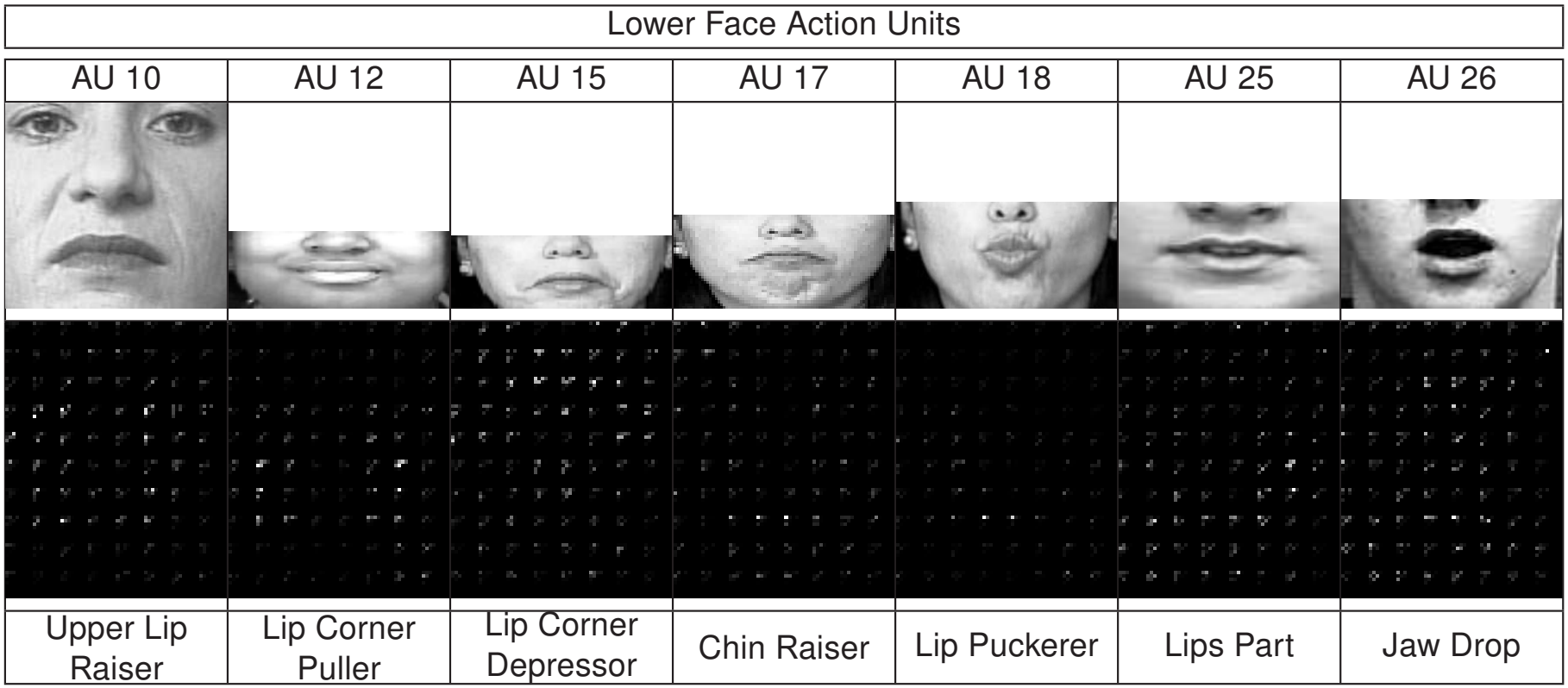
Results

DCT feature importance analysis for AU detection

Upper Face Action Units					
	AU 1	AU 2	AU 4	AU 6	AU 7
					
	Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Cheek Raiser	Lid Tightener

Results

DCT feature importance analysis for AU detection



Conclusion

- Motivation for FERA Challenge
- FERA 2011 tasks and challenge data
- Official LBP-based baseline system
- Our DCT-based approach
- Results
  - Our DCT-based approach outperfoms the LBP-based baseline for emotion recognition
  - For AU detection LBP works better than DCT
    $\Rightarrow$ still lots of things to do about that

Future Work

Upcoming - Secondary emotion recognition test

- organisers have decided to follow up emotion sub-challenge with a second test, where participants do not get to see data
- two options to perform this test:
  1. send end-to-end program to organisers
     - programs must be in by the 3rd of February 2011
  2. bring end-to-end program to FG conference and perform test on site with own hardware
     - test videos will be provided on memory stick (with same format as first test set)
     - results have to generated within 4 hours in the presence of the organizers
     - results have to be mailed to organizers, who will reply with the scores and update the ranking accordingly
- size of secondary test set is approximately half of the original test set
- final score will be weighted sum of primary and secondary emotion recognition test results
- weights of the sets will be proportional to the number of videos in each test set $\Rightarrow \frac{1}{3}$ for second set

Future Work

- improve misclassifications for emotion recognition
- find features that are more appropriate for AU detection
- incorporate time information for AU detection

Questions?

Thank you for your attention!

References I

Tanja Bänziger and Klaus R. Scherer.
Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus, pages 271–294.
Oxford University Press, Oxford, England, 2010.

Marian Stewart Bartlett and Jacob Whitehill.
Automated facial expression measurement: Recent applications to basic research in human behavior, learning, and education.
Oxford University Press, 2010.

References II

Miriam Madsen, Rana el Kaliouby, Matthew Goodwin, and Rosalind W. Picard.

Technology for Just-In-Time In-Situ Learning of Facial Affect for Persons Diagnosed with an Autism Spectrum Disorder.

In *Proceedings of the 10th ACM Conference on Computers and Accessibility (ASSETS), October 13-15, 2008*, Halifax, Canada, 2008.

Esra Vural, Mujdat Cetin, Aytul Ercil, Gwen Littlewort, Marian Bartlett, and Javier Movellan.

Drowsy Driver Detection Through Facial Movement Analysis.

In *ICCV 2007 Workshop on Human Computer Interaction*, 2007.

References III

Jacob Whitehill, Marian Bartlett, and Javier Movellan.
Automatic Facial Expression Recognition for Intelligent Tutoring Systems.
In *Workshop on CVPR for Human Communicative Behavior Analysis, IEEE Conference on Computer Vision and Pattern Recognition.*, 2008.