DA – Progress report 2 - Multi-view facial expression classification 16.12.2010

Nikolas Hesse

Motivation

- Facial expressions (FE) play an important role in **interpersonal communication**
- FE recognition can help to improve **natural human-machine-interaction**
- FE recognition systems exist, but they mainly concentrate on **frontal** faces
- To use FE recognition in **real world situations,** it is necessary to have systems that work for faces with **pose variations** as well

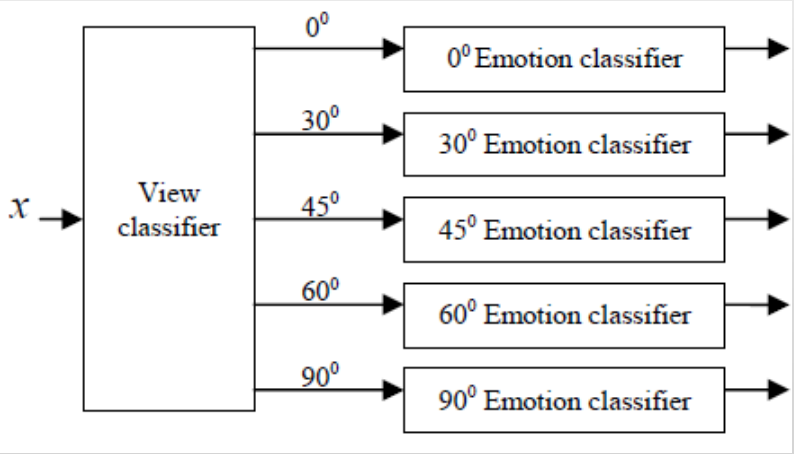
Outline

- Related work
- Multi-view facial expression classification using AAMs
  - Goals
  - Active Appearance Models
  - Features
  - Results
- Conclusion
- Future Work

Related Work

Related Work – Hu et. al 2008 – „Multi-View Facial Expression Recognition“

- Expression classification using different **appearance features** which were extracted from **facial key-points** and processed by different **feature selection** methods
- One classifier to detect the **view**, five **emotion** classifiers (one for each view)



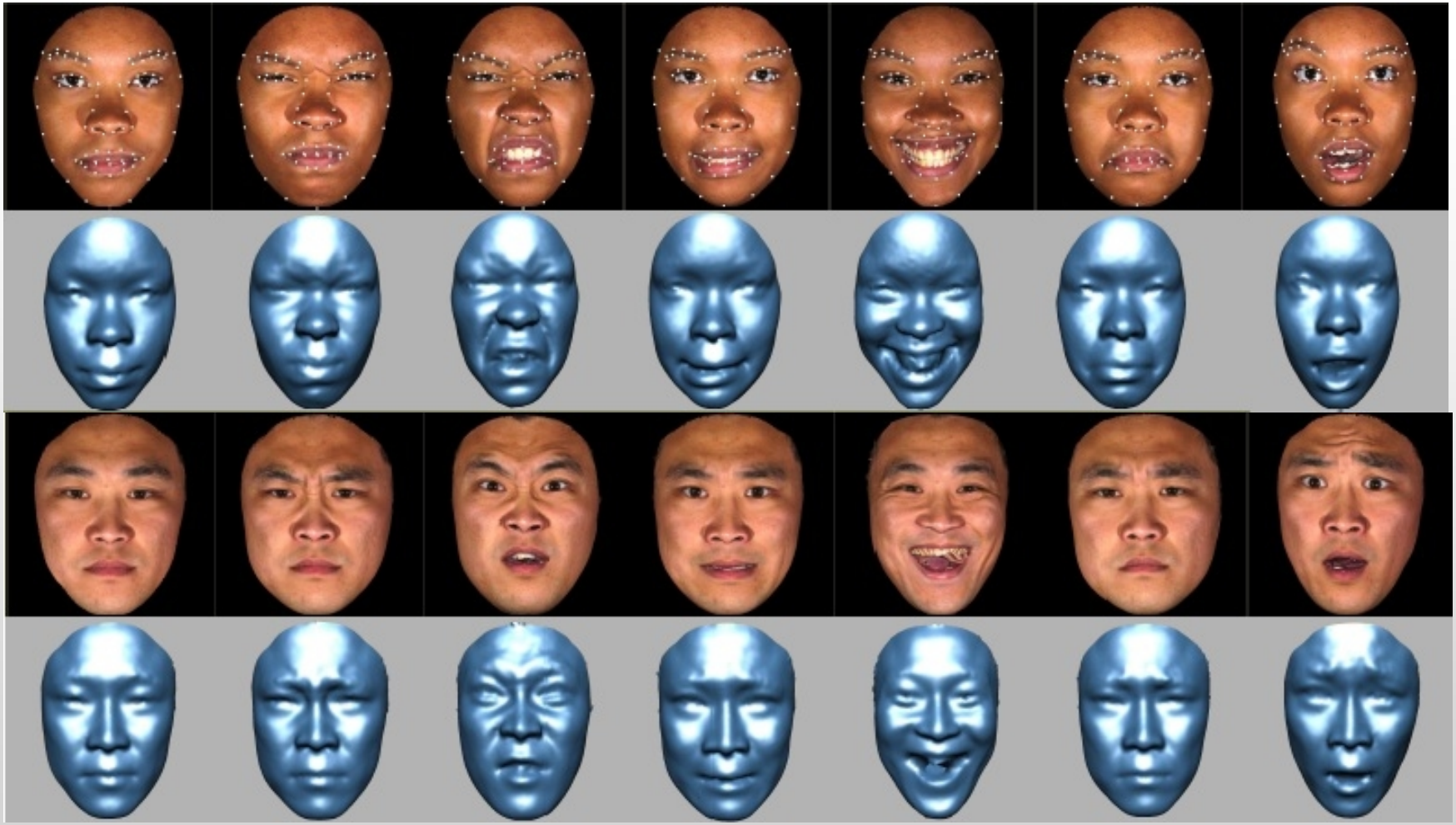
BU-3DFE Database

Data is taken from the Binghamton University 3D Facial Expression Database (BU-3DFE)

- Contains **3d models** (with texture and annotated facial landmarks (83 landmarks per face))
- Includes **100 subjects**: 56 female, 44 male, age from 18 to 70 years, different ethnicities
- Each subject shows **7 expressions** (anger, disgust, fear, happiness, neutral, sadness, surprise) with **4 levels of intensity** (except neutral)

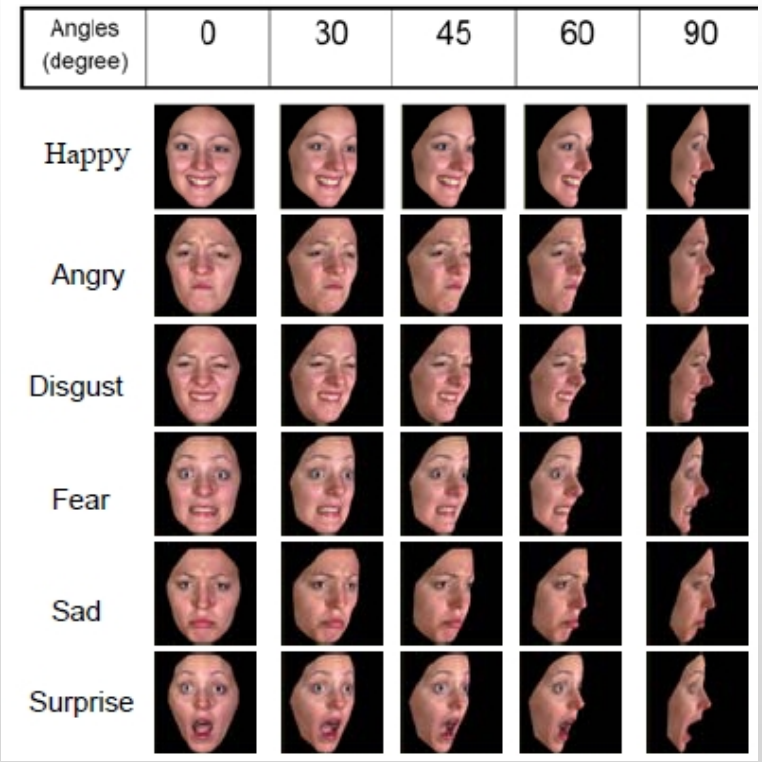
→ 2,500 3D facial expression models

BU-3DFE Database - Examples



Related Work – Data

- Images and shapes saved from **5 views**



Related Work – Features

- Features: Appearance descriptors, extracted at facial key points

- Feature Extraction:
  - **Local binary patterns** (LBP)
  - **Histogram of oriented gradients** (HoG)
  - **Scale-invariant feature transform** (SIFT)
- Feature Selection:
  - **Locality preserving projection** (LPP)
  - **Principal component analysis** (PCA)
  - **Linear discriminant analysis** (LDA)
- Classification: Nearest neighbor

Related Work – Results – Error Rates

- Raw appearance intensity:

	Original	LPP	PCA	LDA
0°	54,58	56,23	54,67	57,63
30°	54,98	61,98	54,88	62,82
45°	56,37	65,45	0,56	71,28
60°	58,48	69,62	58,5	69,5
90°	61,43	77,07	61,75	78,35
Average	**57,16**	66,07	57,18	67,91

- LBP:

	Original	LPP	PCA	LDA
0°	50,13	35,67	50,48	46,36
30°	51,32	34,88	51,5	45,38
45°	52,77	36,33	52,74	59,57
60°	51,5	34,82	51,56	49,08
90°	54,84	37,12	54,95	60,59
Average	52,11	**35,76**	52,25	52,2

Related Work – Results – Error Rates

- HoG:

	Original	LPP	PCA	LDA
0°	53,02	32,08	52,9	41,47
30°	53,73	31,55	53,63	40,04
45°	55,73	32,34	55,99	38,67
60°	56,07	31,98	55,75	37,16
90°	54,68	35,14	54,68	36,6
Average	54,65	**32,62**	54,59	38,79

- SIFT:

	Original	LPP	PCA	LDA
0°	43,68	27,24	43,93	40,84
30°	43,19	26,13	43,27	40,41
45°	44,54	26,65	44,23	55,18
60°	44,64	26,14	44,79	42,29
90°	43,84	28,55	43,89	48,1
Average	43,98	**26,94**	44,02	45,36

Multi-view facial expression recognition using AAMs

Goals

- Multi-view Facial expression recognition using Active Appearance Models (AAMs)
  - **Pose dependent AAMs**,

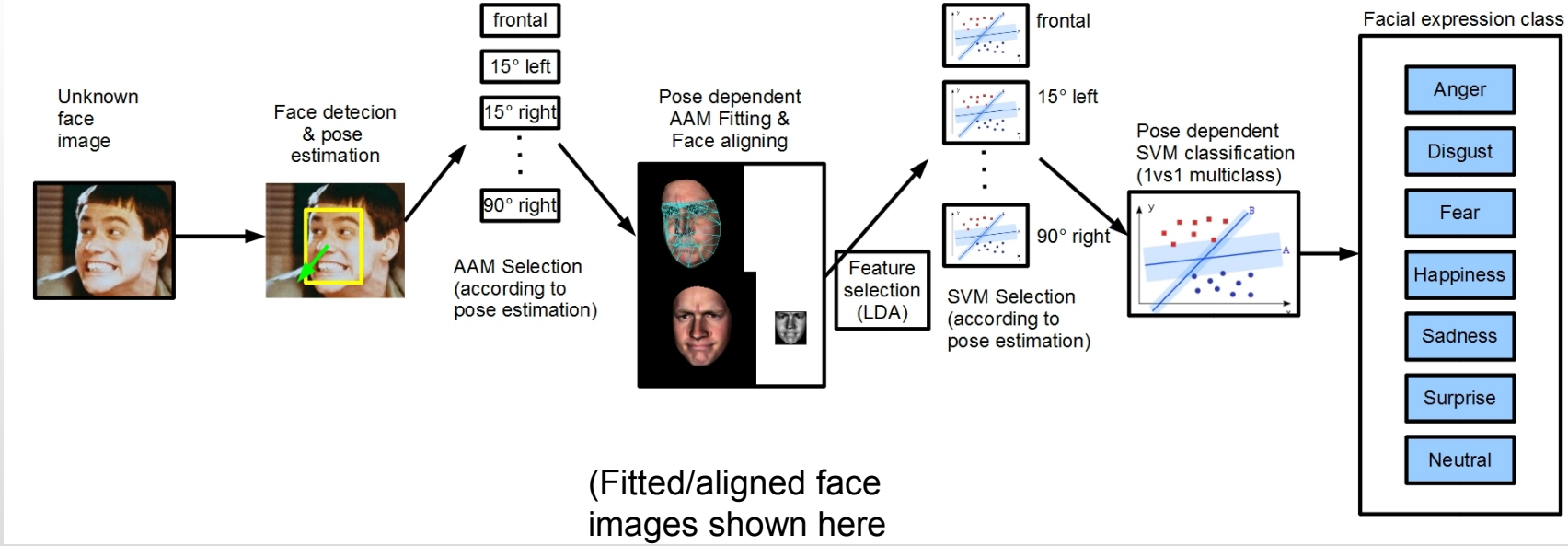
    1 AAM per pose (-90 to +90 degrees,

    15 degree steps → 13 models)
  - **Pose dependent SVM**

    1 SVM per pose (1vs1 multiclass classification)

    Expression classes: Anger, disgust, fear, happiness, sadness, surprise(, neutral)

Program overview



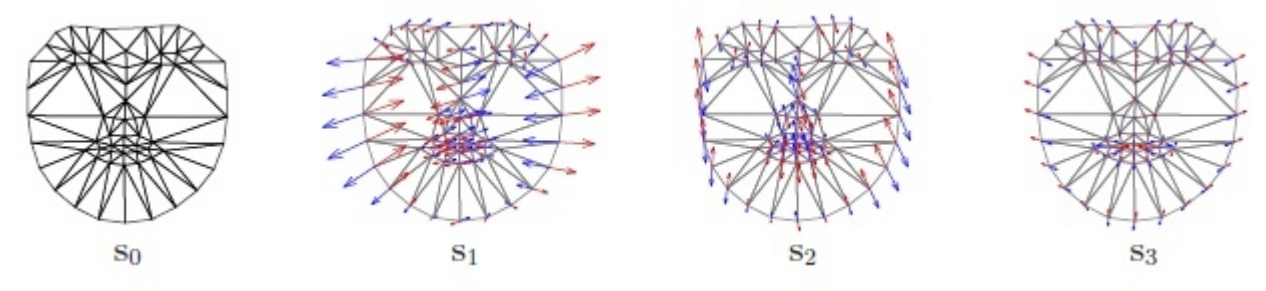
(Fitted/aligned face images shown here do not correspond to shown input image)

Active Appearance Models

Active Appearance Models - Shape

- **Shape**: defined by **mesh** and **vertex locations** of mesh
- Linear shape variation: base shape s_0 + linear combination of n shape vectors s_i; p_i: shape parameters

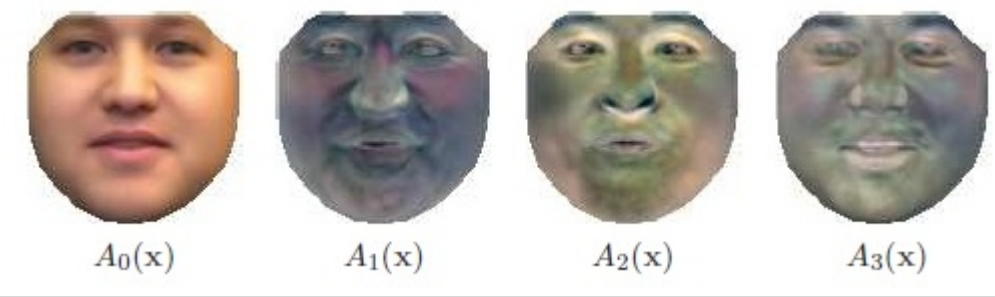
$$\mathbf{s = s_0 + \sum_{i=1}^{n} p_i s_i}$$



Active Appearance Models - Appearance

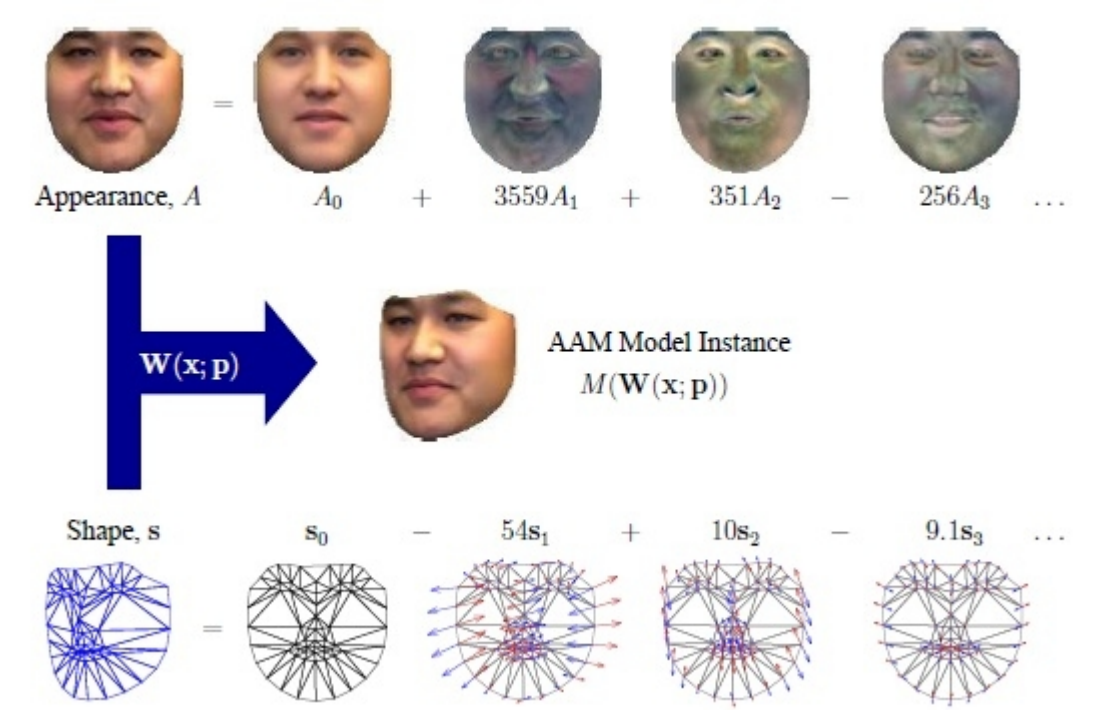
- **Appearance**: **image A(x)** defined over pixels **x ϵ s_0**
- Appearance: base appearance $A_0(x)$ + linear combination of m appearance images $A_i(x)$, λ_i appearance parameters

$$\mathbf{A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x)}$$



Active Appearance Models

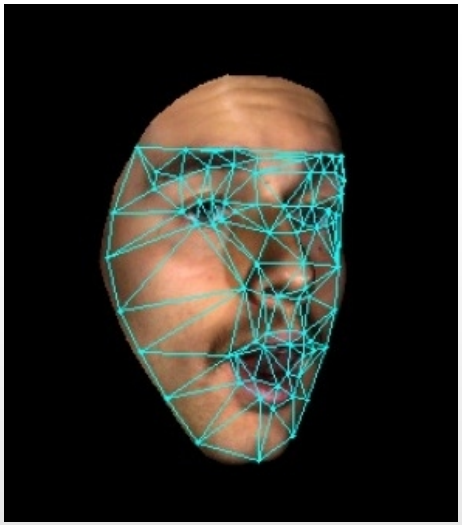
- AAMs are used for **matching a statistical model** of object shape and appearance **to a new image**
- AAM model instance **M(W(x;p))**: computed by **warping** the appearance **A from s_0 to s** using W(x;p)



Features

Features

- Shape coordinates
- Shape parameters
- Texture parameters
- SIFT descriptors

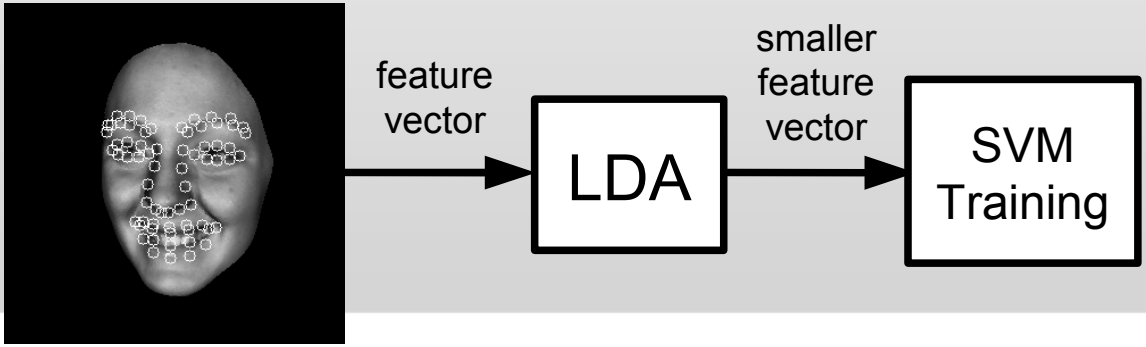


- Combination of features

SIFT – Scale Invariant Feature Transform

- Get '**interest-points**' from fitted AAM shape:

  frontal poses 68 points, side poses 36 points
- **Extract SIFT-descriptor** for each interest-point (128 values per interest-point) → frontal: 8704 values, side: 4608 values
- **Feature Selection** (LDA, afterwards: 6 values)
- Train SVM using data



Classification Results

Classification accuracy – average over all poses

Features used

Classes	Raw shape coordinates	Raw shape parameters	Raw texture parameters	Raw shape coordinates + texture parameters	Raw shape parameters + texture parameters	LDA shape parameters + LDA texture parameters	LDA SIFT-Features (without 90° left & right*)	LDA shape parameters + LDA texture parameters + LDA SIFT-Features
Anger	55,3	44,7	61,2	**64**	61,3	55,9	58,4	63,8
Disgust	61,2	54,5	61,2	67,8	**68,6**	66,1	67,3	65,5
Fear	29	28,9	33,7	39	**40,2**	33,7	37,3	33,5
Happiness	72,6	61,4	65,5	**73,1**	68,1	66,4	69,9	69,6
Sadness	49,4	49,8	60,2	63,7	**63,8**	59,6	50,1	61,6
Surprise	69,9	68	73,7	**76,2**	74	70,5	73	**76,2**
Neutral	9,4	9,1	26,8	**27**	25,4	21,2	12,8	12,9

*something wrong with those

Accuracy – average over all emotions

Features used

Angles	Raw shape coordinates	Raw shape parameters	Raw texture parameters	Raw shape coordinates + texture parameters	Raw shape parameters + texture parameters	LDA shape parameters + LDA texture parameters	LDA SIFT-Features	LDA shape parameters + LDA texture parameters + LDA SIFT-Features
90l	49,7	48,5	54,2	58,2	**58,9**	53,7	25,4*	53,6
75l	52,7	49,2	54,8	61,4	57,2	47,6	59	**61,8**
60l	52,7	51,3	54,6	**62**	61	53,2	54	56,3
45l	54,1	52,3	60,5	65,2	**65,3**	62,7	54	60,7
30l	54,3	51,8	58	62,6	**63,2**	62.1	57,4	62,6
15l	64,8	52,8	66,7	70,8	65,9	**71,5**	64,7	68,2
frontal	54,6	45	61,3	**64,8**	64,1	63,1	56	60,4
15r	58,6	48,2	58,8	**61,7**	59,1	52,4	58	60,4
30r	57,2	55,8	63,8	67,3	66,6	64,7	62	**67,4**
45r	56,5	52,3	58,8	**64,5**	63,3	58,5	58,8	61
60r	53	51,8	58,6	**61**	59,8	56,3	52,8	57,8
75r	49,8	50	54,5	58,9	**59,7**	49,2	54,1	57,3
90r	50,1	47,7	52,3	**56,8**	53,1	52,1	23,5*	52,3

*something wrong here

Classification results – training on high intensities – features: raw shape coordinates & texture parameters

Average accuracy over all poses

	All intensities used for training	Intensities 2, 3, 4 used for training	Intensities 3, 4 used for training
Anger	64	**65,3**	61,7
Disgust	**67,8**	66,5	62
Fear	39	39,3	**45,1**
Happiness	**73,1**	69,3	62,8
Sadness	**63,7**	60,2	53,9
Surprise	**76,2**	74,1	68,4
Neutral	27	36,4	**48,7**

Average accuracy over all poses & emotions

	All intensities used for training	Intensities 2, 3, 4 used for training	Intensities 3, 4 used for training
Intensity 1	**55,1**	52,1	44,7
Intensity 2	**65,5**	62,1	59,2
Intensity 3	**67**	66,7	64,2
Intensity 4	70,8	**71,6**	70,6

Classification results – training on high intensities – features: raw shape coordinates & texture parameters

Average accuracy over all emotions

	All intensities used for training	Intensities 2, 3, 4 used for training	Intensities 3,4 used for training
90l	58,2	**58,5**	53,1
75l	**61,4**	59,5	57,6
60l	**62**	61,4	55,9
45l	**65,2**	64	62,6
30l	62,6	**64**	60,3
15l	70,8	**72,5**	66,5
frontal	64,8	**65,5**	62,2
15r	**61,7**	59,5	57,5
30r	**67,3**	66,4	64
45r	**64,5**	63,3	58,1
60r	**61**	60,9	58,5
75r	**58,9**	50,2	53,5
90r	**56,8**	55,6	52,3

Comparison: confusion matrices: my best result – average related work

My 'best' result (highest average accuracy (72,5%)): Intensities 2, 3, 4 used for training – 15° left

Accuracy(%)		Recognized						
		Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral
Ground-truth	Anger	**71,9**	12,5	0	0	9,4	0	6,3
	Disgust	11,1	**80**	4,4	2,2	1,1	1,1	0
	Fear	10	14,3	**51,4**	8,6	4,3	5,7	5,7
	Happiness	1,7	3,4	3,4	**88,9**	0,9	0	1,7
	Sadness	14,3	3,9	9,1	1,3	**61**	0	10,4
	Surprise	5,1	3,8	10,1	1,3	5,1	**70,9**	3,8
	Neutral	6,3	6,3	0	0	18,8	0	**68,8**

Related work: average over combination of SIFT+LPP, LBP+LPP and HoG+LPP on 5 views

Accuracy(%)		Recognized					
		Anger	Disgust	Fear	Happiness	Sadness	Surprise
Ground-truth	Anger	**73,5**	5,6	3,9	0,8	15,4	0,9
	Disgust	10,2	**71**	7,7	3,7	4,1	3,4
	Fear	7	9,2	**55,4**	14,5	7,7	6,3
	Happiness	1,7	3,1	11,6	**81,7**	1,1	0,7
	Sadness	19,7	2,1	5,7	0,9	**71,4**	0,3
	Surprise	1,3	2,3	5,5	1,5	1,7	**87,7**

Conclusion

Conclusion

- Related work has better results, but:
  - groundtruth points used for extracting features
  - only 5 angles, bigger distances
  - neutral expression not included
- General Problems:
  - 'weak' expressions (fear, anger, sadness) have worse results than 'strong' (surprise, happiness) ones
  - 'neutral': few data, very bad results → will be dropped
  - low intensities of expressions hard to recognize, even by humans

Conclusion

- When training on high intensity data:
  - 'weak' expressions get better results, but others become worse
  - results improve for highest intensity (training on int. 2, 3, 4), but others are worse

Future Work:

- Drop 'neutral' class
- Try different feature selection methods
- Try linear SVMs for classification
- Cross validation

Thank you for your attention!

Any questions?

References

- Yuxiao Hu, Zhihong Zeng, Lijun Yin, XiaozhouWei, Jilin Tu, and T.S. Huang, „Multi-view facial expression recognition“, FG2008, 2008, ICPR 2008, 8th International Conference on Automatic Face and gesture Recognition, Sept. 2008.

Done until now

- Done fitting for rest of data
- Performed LDA on data
- Classified fitted test-data using: shape-coordinates, shape parameters, texture parameters, shape coordinates & texture parameters, shape & texture parameters, SIFT, shape & texture parameters & SIFT
- Compared results
- Analysis of influence of emotion-intensities on classification
- Trained SVMs with high intensity-data only (234, 34) + classification

Data - Example

Model rotated from -90° to 90°, 15° steps,
for each step image and shape are saved