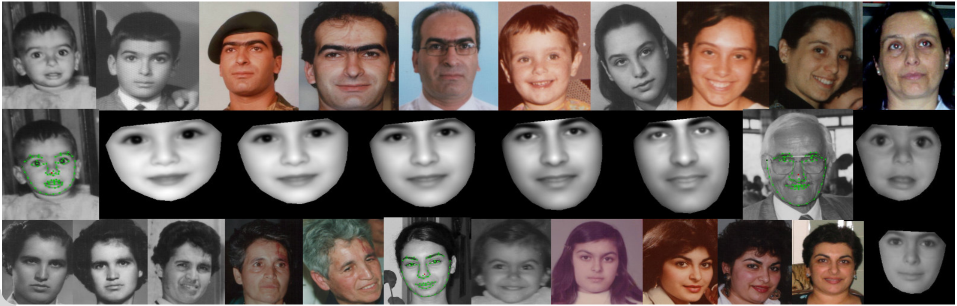
Karlsruhe Institute of Technology

Facial Image-based Age Estimation

Matthias Steiner | December 9, 2010

INSTITUTE FOR ANTHROPOMATICS, FACIAL IMAGE PROCESSING AND ANALYSIS



Outline

Introduction

The face is an important indicator of a person's age

Sample applications

- Observation of age restrictions
- Age adapted user interfaces
- Simulation of the aging process

Challenges

- Every person ages differently

- Collecting sufficient amount of training data
- Influence of facial expressions, head pose, gender and ethnicity

Active Appearance Model I

Basics

- Objective: Describe a face with a set of parameters
- Idea: Parameters describe the differences from the mean face

$\Rightarrow$ A statistical model of the variation of shape and texture is learned

$\Rightarrow$ Training images with the annotated shape are needed

1. Shape model

- PCA is applied to all shapes

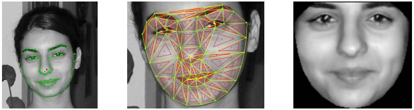
$\Rightarrow$ Orthogonal modes of variation P_s

$\Rightarrow$ Shape Model: $s = \overline{s} + P_s b_s$

Active Appearance Model II

2. Texture model

- Mean shape is used to warp every face into a shape free version



- The texture is normalized in the shape area
- PCA is applied to all warped textures

$\Rightarrow$ Texture Model: $t = \bar{t} + P_t b_t$

3. Combined model

- The shape and texture vector are concatenated: $b = \binom{W_s b_s}{b_t}$
- A third time PCA is applied to these vectors

$\Rightarrow$ Combined Model: $b = Q b_c$

Support Vector Machine I

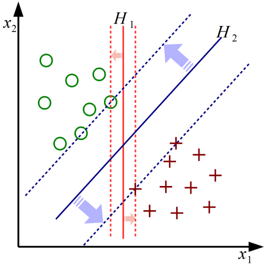
Basic Problem

$$d = \{(x_i, y_i) | x_i \in \mathbb{R}^m, y_i \in \{-1, +1\}\}$$

Objective

Find a optimal hyperplane between the these classes

$\Rightarrow$ Lowest separation error and the best generalization



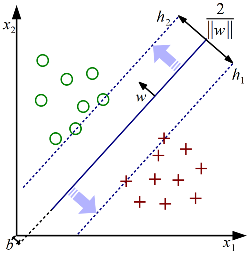
Support Vector Machine II

Approach

- Hyperplane: $\langle w, x \rangle + b = 0$
- Scale parameters to fulfill $|\langle w, x \rangle + b| \geq 1$

$\Rightarrow$ Points closest $|\langle w, x \rangle + b| = 1$ on h_1, h_2

$\Rightarrow$ Maximize distance h_1, h_2 : $\frac{2}{\|w\|}$



Basic Optimization Problem

$$\Rightarrow \min_{w,b} \|w\| \text{ s.t. } y_i * (\langle w, x_i \rangle + b) \geq 1, \forall i$$

Support Vector Machine Extensions I

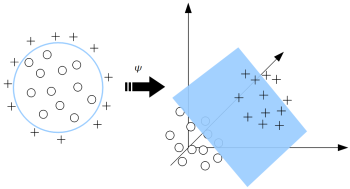
The Kernel-Trick

- The non-linear separable data is transformed into a higher dimensional space

⇒ Complex calculation of the scalar product

⇒ Replace the scalar product with a kernel function

  - polynomial, radial basis function (RBF),...

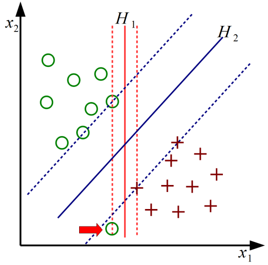


Support Vector Machine Extensions II

Soft Margin

- Allows to regulate between separation error and generalization

$$\min_{w,b,\xi} \|w\| + C\sum_{i=1}^{n}\xi_i \text{ s.t. } y_i * (\langle w, x_i\rangle + b) \geq 1 - \xi_i, \forall i, \xi_i \geq 0$$



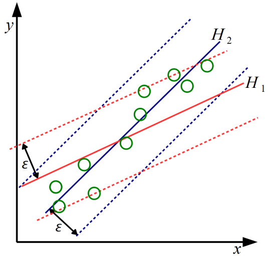
Support Vector Machine Extensions III

Karlsruhe Institute of Technology

(epsilon) SV-Regression

- y_i can be any real number $\Rightarrow$ regression problem
- Find a function that has at most ϵ deviation

$$\min_{w,b} \|w\| \text{ s.t. } \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \epsilon \\ \langle w, x_i \rangle + b - y_i & \leq \epsilon \\ \forall i & \geq 0 \end{cases}$$

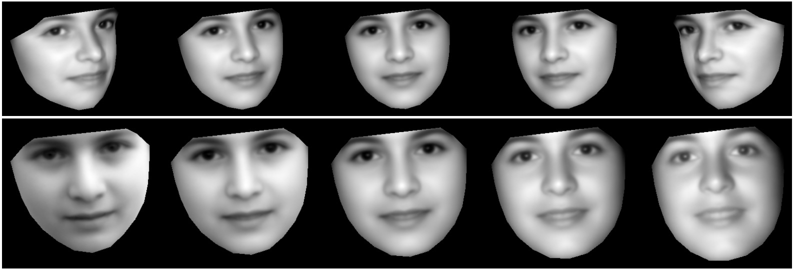


AAM Building I

Configuration

- Convex hull model
- 95% of the variation is described
- The texture size is halved

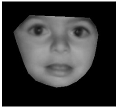
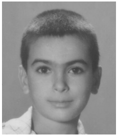
Variation of the first 2 parameters

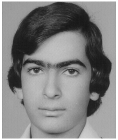


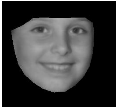
AAM Building II

Variation of the parameters 3-5

AAM Fitting

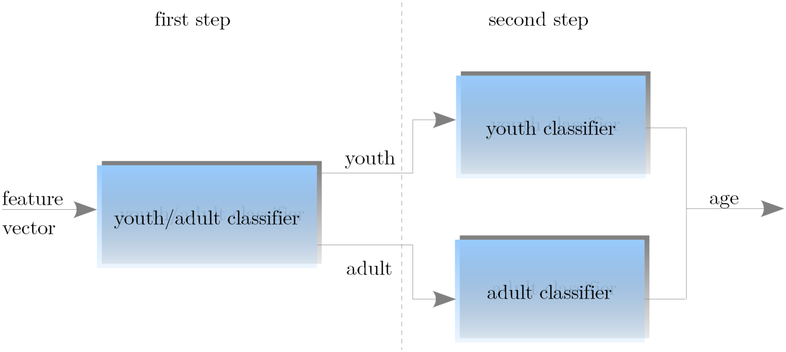








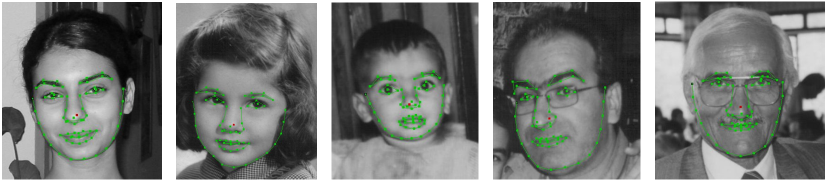
The classifier



youth $\widehat{=}$ age$\leq$20 , adult $\widehat{=}$ age$>$20

The FG-NET database

- 1002 mixed color and greyscale images
- 68 landmark points for every sample
- 6-18 images for each of the 82 subjects (age: 0-69)
- Image resolution varies about 400 $\times$ 500 pixels
- Uncontrolled conditions
- Ethnicity: White people



Evaluation Method

Leave One Person Out Evaluation (LOPO)

In each fold:

- The pictures of one person are hold out for testing
- All remaining images are used for training

$\Rightarrow$ On the FG-NET database this leads to 82 folds

- Training set: AAM building $\rightarrow$ feature extraction $\rightarrow$ SVM training
- Testing set: feature extraction $\rightarrow$ age estimation

Parameter optimization

- Cross validation: The subjects instead of the single images are randomly divided into training and testing set

$\Rightarrow$ Prevents that ”intra personal” relations are learned

Performance Measurements

Mean Absolute Error (MAE)

- is the mean difference between the real and the predicted age

$$MAE = \frac{\sum_{i=0}^{n} |EA_i - RA_i|}{n}$$

- where EA_i is the estimated and RA_i the real age for the i^{th} of n tested samples

Cumulative Score (CS)

- Let d an age error in years
- CS is % of estimations having an estimation error $\leq d$

$$CS(d) = \frac{N|EA_i - RA_i|_{\leq d}}{n} \times 100$$

Karlsruhe Institute of Technology

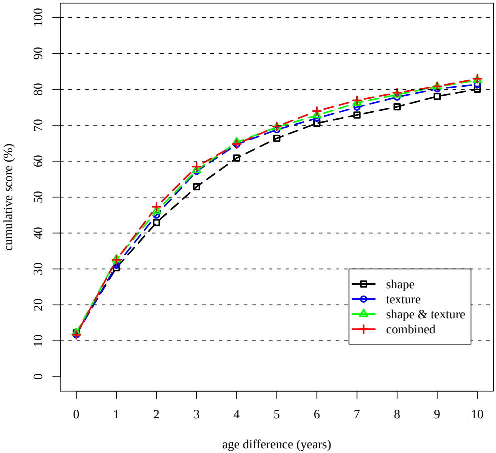
Mean Absolute Error

For the MAE calculation of the second step classifiers the miss classifications of the first step are ignored

features used	dim.	overall result	first step	second step	
			youth/adult classifier	youth classifier	adult classifier
shape only	27	6.16	20.44%	2.32	7.77
texture only	102	5.84	19.24%	2.15	7.55
shape & texture	129	5.71	18.84%	2.16	7.55
combined	47	5.58	18.50%	2.11	7.56

Baseline Version Evaluation II

Cumulative Score



DCT Features Approach

Alignment Version

1. Use the eye coordinates to align the face

2. Fit the AAM to get the shape free face

Extraction

1. Scale image to 64×64 pixels
2. DCT is performed on blocks of 8×8 pixels
3. 5 coefficients in zig-zag order are kept for each block

$\Rightarrow 8 \times 8 \times 5 = 320$ dimensional feature vector

DCT Features Evaluation

KIT Karlsruhe Institute of Technology

Mean Absolute Error

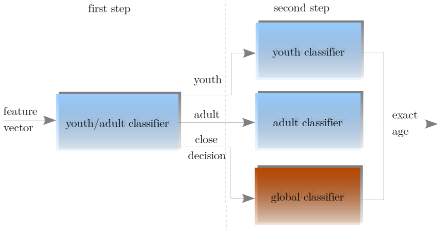
features used	dim.	overall result	first step	second step	
			youth/adult classifier	youth classifier	adult classifier
combined	47	5.58	18.50%	2.11	7.56
DCT v1	320	5.91	17.42%	2.75	7.85
DCT v2	320	5.55	17.61%	2.48	7.78
DCT v2 & shape	347	5.08	15.65%	2.19	7.80

Soft Binary Approach

- Use the decision value to identify close decisions

⇒ Close decisions are given to a global classifier

(overall MAE: combined: 5.50 years, DCT v2 & shape: 5.08 years)



- combined: 5.21 years (before 5.58), (youth/adult error 5.69%)
- DCT v2 & shape: 4.77 years (before 5.08), (youth/adult error 4.59%)

Final Evaluation

KIT Karlsruhe Institute of Technology

Age Range Analyses

- Estimations are grouped according to their real age

features used	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE
combined	2.28	5.01	7.29	8.31	14.11	23.54	33.3
combined(sb)	2.41	3.88	5.78	9.92	15.83	25.64	35.46
DCT&shape	1.99	4.04	7.12	9.37	13.62	21.79	28.66
DCT&shape(sb)	2.19	3.67	4.97	9.10	15.09	22.60	32.13
image count	371	339	144	79	46	15	8

- Upper age ranges are trained badly
- Classifier prefers ages with many samples

$\Rightarrow$ The estimation performance varies widely

Current Studies

Objective:

- Build a balanced age estimation system

Approach:

- Balance the number of images per age
  - ⇒ Add data from the MORPH database
  - ⇒ Use flipped images
  - ⇒ Limit images per age
- AAM fitting is complex

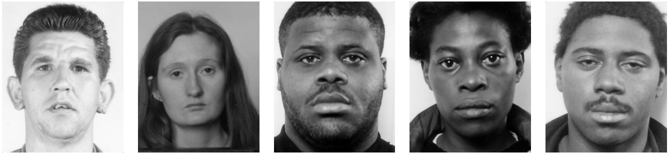
  ⇒ Use eye alignment only (DCT v1)
- The youth/adult classification influences the MAE of the age ranges

  ⇒ Drop the youth/adult classification step

The MORPH database

Album 1

- 1690 greyscale images
- eye coordinates for every sample
- 631 subjects of the age from 15-68
- Image resolution: 400 $\times$ 500 pixels
- Ethnicity: black 1253 images white 434 images



The MORPH database

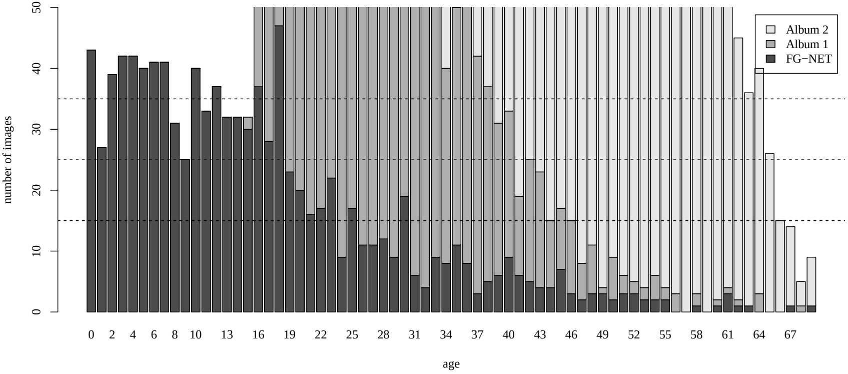
Album 2

- 55608 color images
- 13673 subjects of the age from 16-99
- Image resolution: 200×240 - 400×480 pixels
- Ethnicity: black 42897 images white 10736 images
- No eye coordinates

  $\Rightarrow$ MCT detector of the okapi library is used

Evaluation Set

Age Distribution FG-NET & MORPH



Evaluation I

Karlsruhe Institute of Technology

FG-NET MAE Age Range Analysis

training age limit	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE	overall MAE
30	7.93	6.44	4.94	8.73	13.87	25.47	27.38	7.75
50	6.84	5.68	4.80	10.00	14.76	27.07	29.00	7.25
70	6.51	5.33	4.81	9.98	14.87	27.73	30.25	6.99
#images	371	339	144	79	46	15	8	1002

Evaluation II

Album 1 MAE Age Range Analysis

training age limit	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE	overall MAE
30	-	12.31	8.81	5.37	6.87	12.8	16.71	8.60
50	-	9.50	6.76	4.64	7.40	13.48	19.29	6.98
70	-	7.90	5.67	4.74	7.94	13.92	20.14	6.24
#images	0	343	763	428	124	25	7	1690

Karlsruhe Institute of Technology

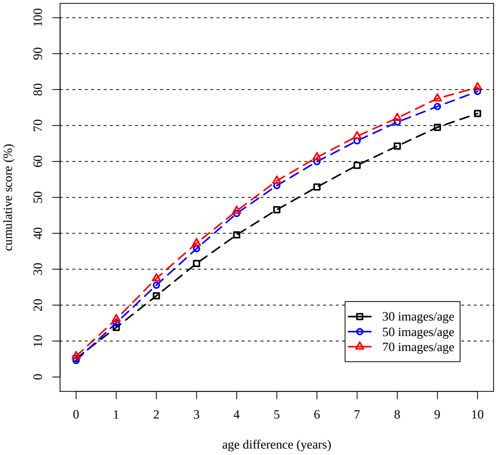
MAE Age Range Analysis

- Testing samples are collected from the whole evaluation set
- Testing samples per age are limited to 50

training age limit	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE
30	8.19	7.03	6.56	6.21	7.19	6.47	10.69
50	7.14	6.12	5.33	5.76	6.70	6.71	10.81
70	6.78	5.67	4.65	5.78	6.63	6.86	11.05

Evaluation IV

Cumulative Scores



Conclusion

- On the FG-NET database a human reaches the following MAE:
  - Whole picture: 6.23 years
  - Face only: 8.13 years
- Our best results: 4.77 years

⇒ AAM or DCT combined with SVM is suited for age estimation

- The youth/adult classification is very challenging
  - ⇒ Using DCT features slightly reduces the classification error
- An age balanced training set provides a quite balanced age estimation
- Ethnicity should probably be considered

Questions?

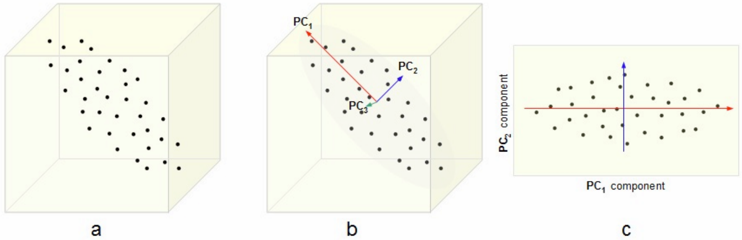
References

- T.F.Cootes, G.J. Edwards and C.J.Taylor, Active Appearance Models, *in Proc. European Conference on Computer Vision, Vol. 2*, pp. 484-498, 1998.
- V. Vapnik, A. Lerner Pattern recognition using generalized portrait method, *Automation and Remote Control, vol. 24*, pp. 774-780, 1963.
- B. Stegmann, The AAM-API, `http://www2.imm.dtu.dk/~aam/`.
- Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, `http://www.csie.ntu.edu.tw/~cjlin/libsvm`, 2001.
- Xin Geng, Thi-Hua Thou, Yu Thang, Gang Li, Honghui Dai, Learning from Facial Aging Patterns for Automatic Age Estimation, *In Proc. of 14th ACM Int'l Conf. Multimedia*, pp. 307-316, 2006.

Principal Component Analysis (PCA)

1. Direction of the greatest variance is detected
2. The basis is changed
3. The dimension is reduced

Example:



Implementation Additionals I

Training:

1 All samples to train the youth/adult classifier

2a Samples with age ≤ 20 to train the youth classifier

2b Samples with age > 20 to train the adult classifier

SVM type:

1. Support Vector Classification for the first step classifier
   - $\rightarrow$ CSVC with the RBF kernel from the *LIBSVM* library
2. Support Vector Regression for the second steps classifiers
   - $\rightarrow$ *Epsilon-SVR* with the RBF kernel also from *LIBSVM*

Implementation Additionals II

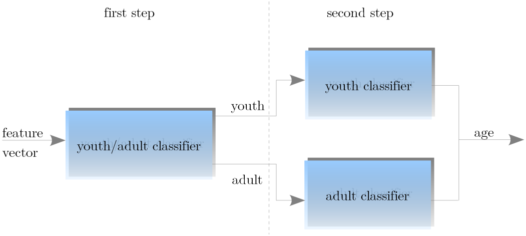
Building procedure

1. Build the AAM on the training set
2. Extract the feature vectors for all training images
3. Scale the parameters of the feature vectors to $\{-1,1\}$
4. Train the three classifiers using the respective data and optimize the SVM parameters using a 5 fold cross validation

Implementation Additionals III

Estimation procedure

1. Extract the feature vectors for all testing images
2. Scale the feature vectors
3. Estimate the age by applying the two steps of classifiers



AAM Fitting

Algorithm

1. Start the fitting with the mean or landmark shape
2. Calculate the initial error between synthesized and real face
   Mahalanobis distance

---

3. Compute the next displacement using a regression matrix
4. Estimate the new combined vector with step size 1
5. Calculate the new error
   a Accept the estimation if the error has improved
   b Otherwise go to step 4 and try a smaller step
6. Go to step 3 until the error is not further reduced or the maximum of iterations is reached

Basic Version Evaluation IV

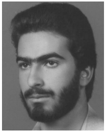
MAE age range analysis

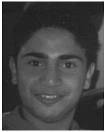
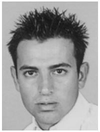
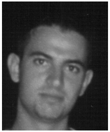
All estimations are grouped according to their real age

features used	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE
shape only	2.77	6.19	7.61	8.18	14.13	23.27	33.50
texture only	2.73	5.65	7.22	8.17	14.15	23.28	34.79
shape & tex	2.52	5.54	7.08	8.03	13.83	23.15	34.00
combined	2.28	5.01	7.29	8.31	14.11	23.54	33.30
image count	371	339	144	79	46	15	8

Baseline Version Evaluation IV

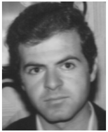
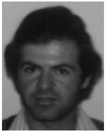
Errors of the youth/adult classifier



real age: 19 10 11 18 20

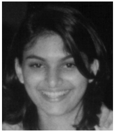


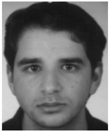




real age: 23 26 22 23 21

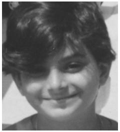
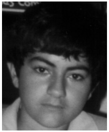
Baseline Version Evaluation V

Large estimation errors of the youth classifier





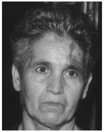
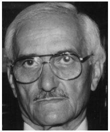

real age:	15	7	20	12
estimated age:	9	13	14	5

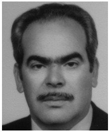
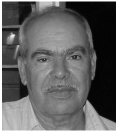

real age:	16	10	7	20
estimated age:	10	4	14	8

Baseline Version Evaluation VI

Large estimation errors of the adult classifier

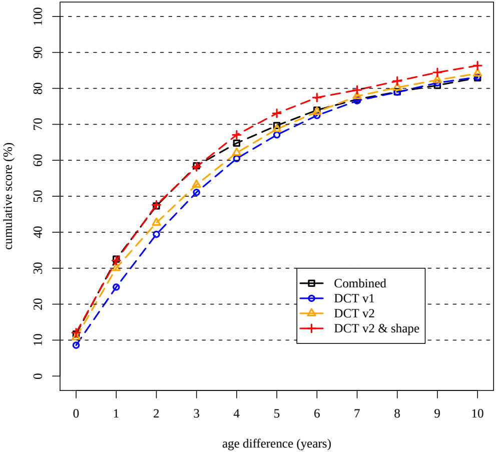
real age:	58	60	62	63
estimated age:	36	32	35	37

real age:	49	52	61	69
estimated age:	24	29	30	39

DCT Features Evaluation II

Cumulative Score



Global Classifier Approach

Background

Without any youth/adult error the overall MAE would be:

- Combined feature vector: 3.74 years
- DCT & shape feature vector concatenated: 3.7 years

$\Rightarrow$ Idea: Get rid of the youth/adult classification

$\Rightarrow$ Only one global classifier (SVR) for the whole age range 0-69

Global Classifier Approach Evaluation

Karlsruhe Institute of Technology

Overall MAE

- Combined: 5.50 years (before 5.58)
- DCT & Shape: 5.08 years (before 5.08)

Age range analyses

features used	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE
combined	2.28	5.01	7.29	8.31	14.11	23.54	33.30
combined(G)	3.12	3.58	4.96	10.85	18.32	27.75	38.46
DCT&shape	1.99	4.04	7.12	9.37	13.62	21.79	28.66
DCT&shape(G)	3.09	3.69	3.80	8.89	16.76	24.89	35.38
image count	371	339	144	79	46	15	8

Soft Binary Evaluation

Overall MAE

- Combined: 5.21 years (before 5.58), (youth/adult error 5.69%)
- DCT & Shape: 4.77 years (before 5.08), (youth/adult error 4.59%)

Age range analyses

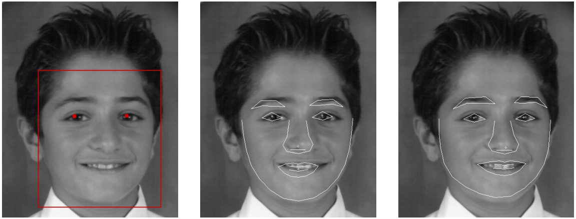
features used	0-9 MAE	10-19 MAE	20-29 MAE	30-39 MAE	40-49 MAE	50-59 MAE	60-69 MAE
combined	2.28	5.01	7.29	8.31	14.11	23.54	33.3
combined(sb)	2.41	3.88	5.78	9.92	15.83	25.64	35.46
DCT&shape	1.99	4.04	7.12	9.37	13.62	21.79	28.66
DCT&shape(sb)	2.19	3.67	4.97	9.10	15.09	22.60	32.13
image count	371	339	144	79	46	15	8

Automatic Initialized Fitting Approach

Problem and Solution

- In real life there are no landmarks to initialize the AAM fitting
- Worst Solution: Try any possible size and position
- Better Solution: Use the MCT detector for face and eye detection

Example



Automatic Initialized Fitting Evaluation

Karlsruhe Institute of Technology

Mean Absolute Error

The face and eyes detection failed for 185 images

⇒ They are excluded from testing

features used	overall result	first step youth/adult classifier	second step youth classifier	second step adult classifier
combined landmark	4.87	5.27	2.07	7.23
combined automatic	6.47	9.49	2.91	7.40
DCT v1 landmark	5.08	5.21	2.56	7.35
DCT v1 automatic	7.04	10.04	3.96	7.77
DCT v2 & shape landmark	4.43	5.08	2.14	7.34
DCT v2 & shape automatic	5.67	9.14	2.73	7.61