Combination of Facial Landmarks for Robust Eye Localization Using the Discriminative Generalized Hough Transform

Ferdinand Hahmann, Dipl.-Inf.

Institute of Applied Computer Science
University of Applied Sciences Kiel

Darmstadt – 05.09.2013

Eye Localization

1. Introduction
2. Method
   - Discriminative Generalized Hough Transform (DGHT)
   - Combination of Landmarks
   - Modified Multi-Level-Approach (MLA)
3. Experiments
   - Database
   - Setup
   - Results
   - Discussion
4. Conclusion

Introduction

- Motivation
  - Eye positions required by many face processing algorithms
- State-of-the-Art
  - Face detector developed by Viola & Jones [VJ04]
  - Usage of a-priori knowledge for eye detection inside faces
- General object localization approaches:
  - Haar-Wavelets in a boosted cascade of classifiers [VJ04]
  - Hough Forests [GL09]
  - Discriminative Generalized Hough Transform (DGHT) [Sch07, Rup13]

[VJ04]: P. Viola and M.J. Jones. Robust real-time face detection. International journal of computer vision, 57(2):137–154, 2004.
[GL09]: J. Gall and V. Lempitsky. Class-specific hough forests for object detection. In Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
[Sch07]: H. Schramm, "Automatic 3-D object detection", Patent Pub. No. W0/2007/07/2391, 2007.
[Rup13] H. Ruppertshofen. Automatic Modeling of Anatomical Variability for Object Localization in Medical Images. PhD thesis, Otto-von-Guericke University Magdeburg, 2013.

Outline

1 Introduction

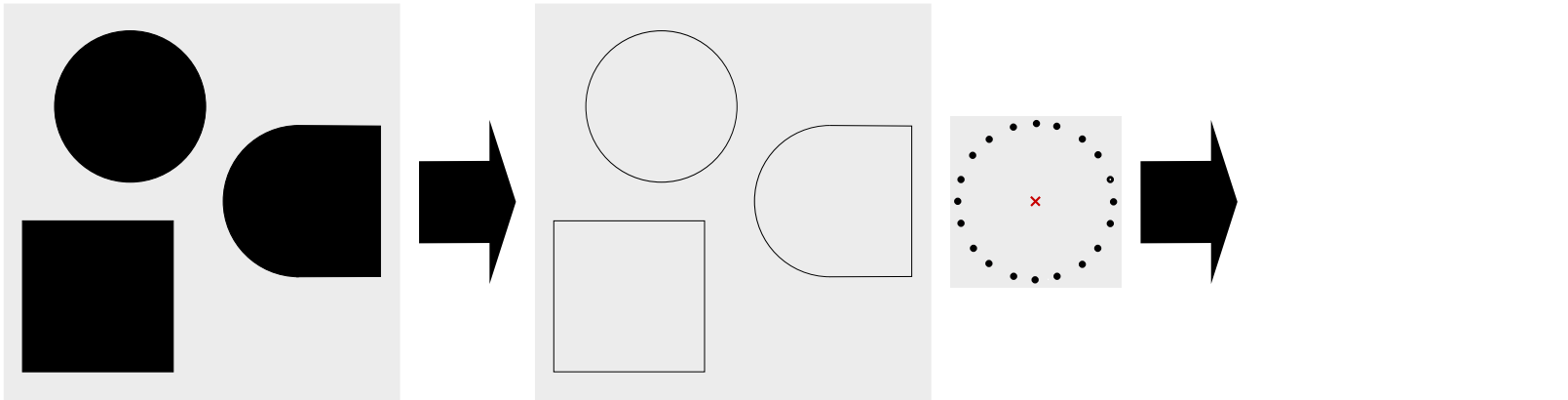
2 Method

3 Experiments

4 Conclusion

Generalized Hough Transform (GHT)

- Introduced by Ballard 1981 [Bal81]
- General model-based method for object localization



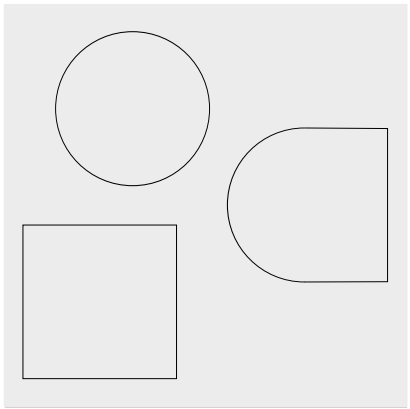
- No model transformations considered

[Bal81]: D. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recogniton 13(2), 1981

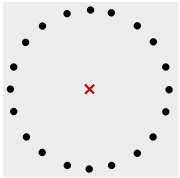
Discriminative Generalized Hough Transform (DGHT)

Motivation

- Discriminative GHT Model:

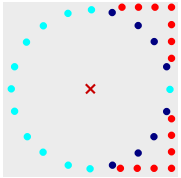


feature image



model

Hough space



weighted model

model point weights:
• = 0.5 • = 1 • =-1

weighted Hough space

- Model is composed of
  - geometric model layout
    $\rightarrow$ learned from training examples
  - individual weights of model points
    $\rightarrow$ discriminative weighting procedure

Model point weight estimation

- Generation of Hough spaces for training images, using initial GHT model
- Separation of Hough space votes coming from every single model point
- Weighted recombination of model point contributions into Maximum Entropy Distribution
- Optimization of introduced weights λ_j by using a Minimum Classification Error (MCE) approach

Model point weight estimation

- Generation of Hough spaces for training images, using initial GHT model
- Separation of Hough space votes coming from every single model point

$$f_j(c_i, X_n) = v_{i,j}$$

$v_{i,j}$: Votes of model point m_j in Hough cell c_i
X_n : Features of image n

- Weighted recombination of model point contributions into Maximum Entropy Distribution
- Optimization of introduced weights λ_j by using a Minimum Classification Error (MCE) approach

Model point weight estimation

- Generation of Hough spaces for training images, using initial GHT model
- Separation of Hough space votes coming from every single model point: $f_j(c_i, X_n)$
- Weighted recombination of model point contributions into Maximum Entropy Distribution

$$p_\Lambda(c_i|X_n) = \frac{\exp\left(\sum_j \lambda_j \cdot f_j(c_i, X_n)\right)}{\sum_k \exp\left(\sum_j \lambda_j \cdot f_j(c_k, X_n)\right)}$$

$\Lambda = [\lambda_1, \lambda_2, ..., \lambda_J]$: Individual model point weights

- Optimization of introduced weights λ_j by using a Minimum Classification Error (MCE) approach

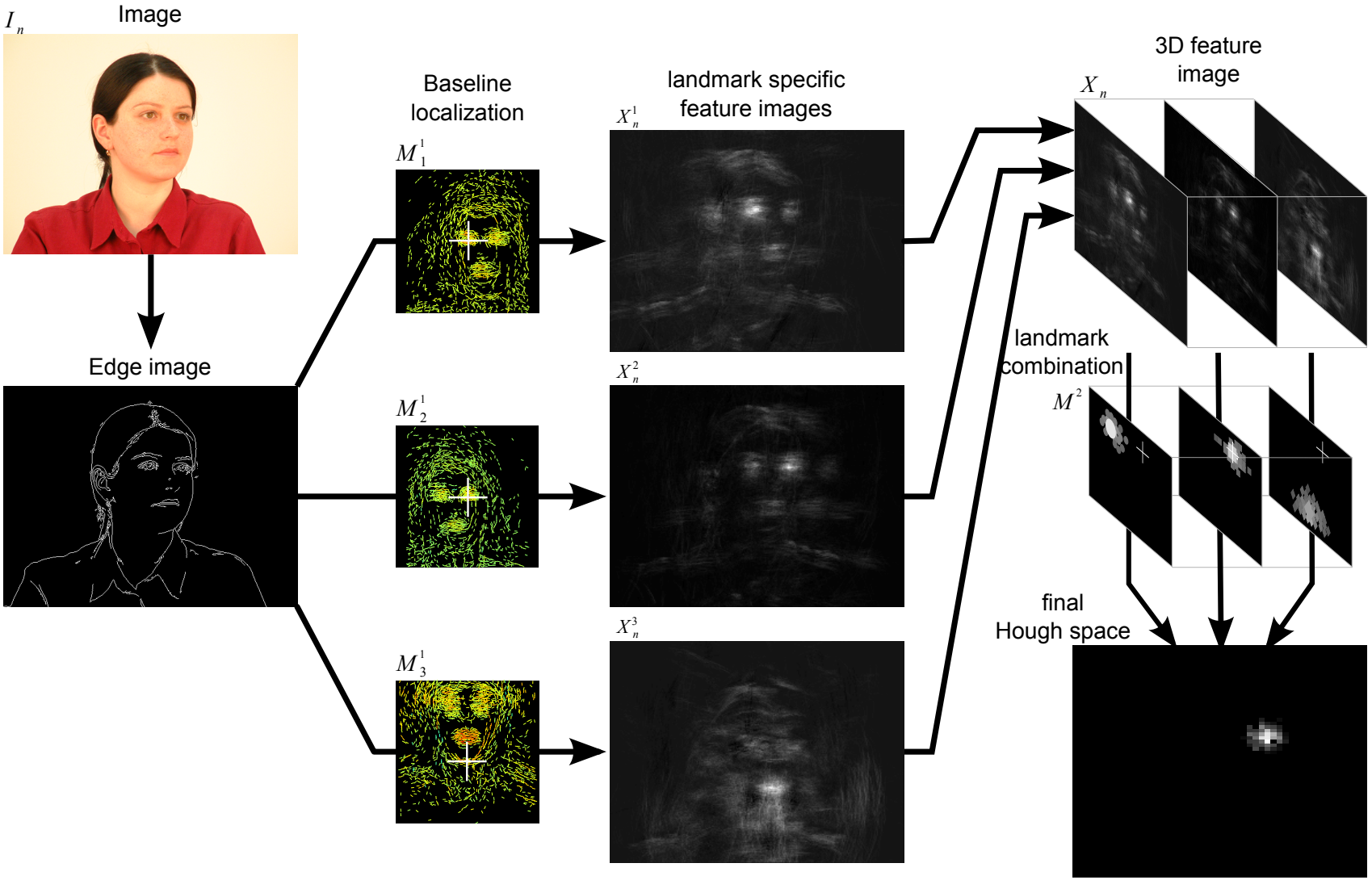
Model point weight estimation

- Generation of Hough spaces for training images, using initial GHT model
- Separation of Hough space votes coming from every single model point: $f_j(c_i, X_n)$
- Weighted recombination of model point contributions into Maximum Entropy Distribution : $p_\Lambda(c_i|X_n)$
- Optimization of introduced weights λ_j by using a Minimum Classification Error (MCE) approach

$$E(\Lambda) = \sum_{n=1}^{N} \sum_{i=1}^{I} \varepsilon(c_i, \tilde{c}_n) \cdot \frac{p_\Lambda(c_i|X_n)^\eta}{\sum_k p_\Lambda(c_k|X_n)^\eta}.$$

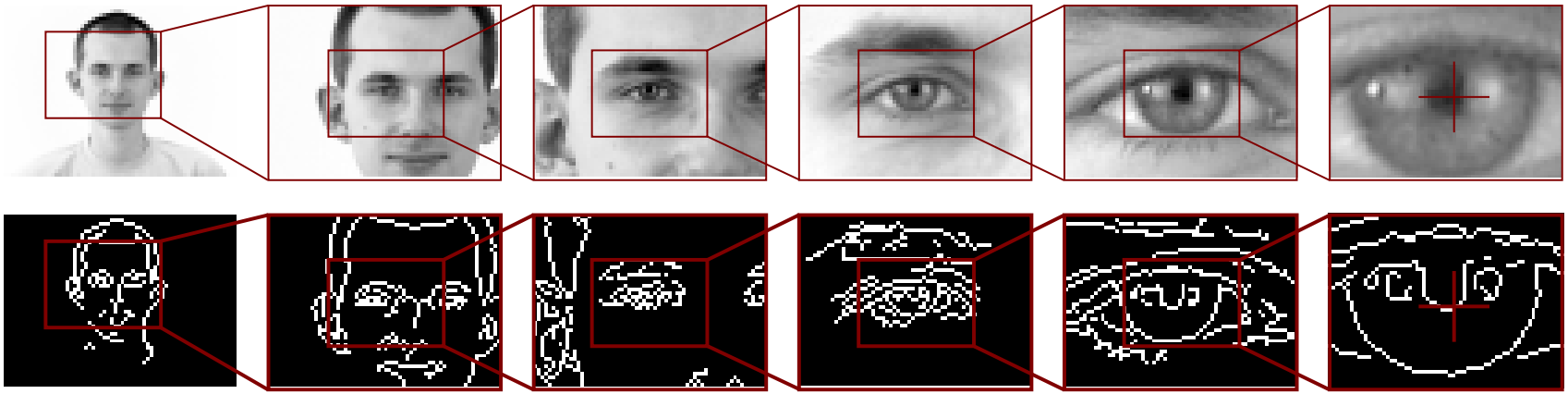
$\varepsilon(c_i, \tilde{c}_n)$:	Error measure (e.g. Euclidean Distance)
$\tilde{c}_n$:	Target cell in Hough space for image X_n
I:	Number of Hough cells
N:	Number of training images

Combination of Landmarks



Multi-Level-Approach

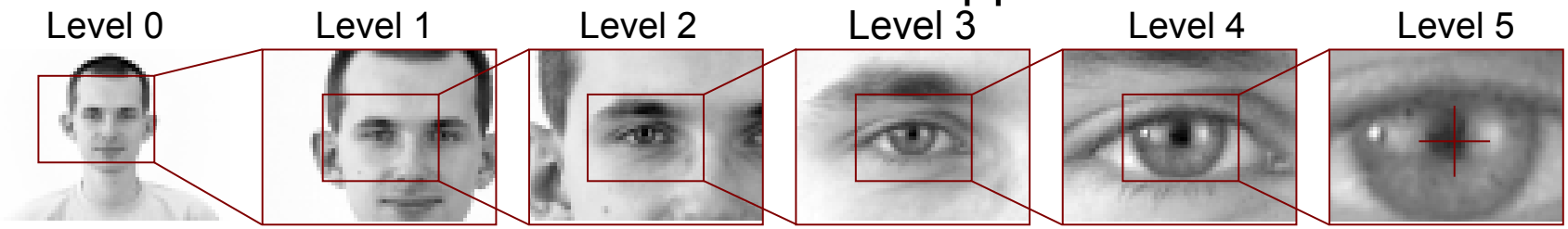
- Zooming-in strategy based on Gaussian pyramid
- Stepwise increase image resolution around detected point



- Specifically trained DGHT model on each level
  $\rightarrow$ Capturing of different level-specific details
- Good trade-off between model accuracy and reduced confusion with concurrent objects

Modified Multi-Level-Approach

Standard Multi-Level-Approach



Modified Multi-Level-Approach



Outline

1 Introduction

2 Method

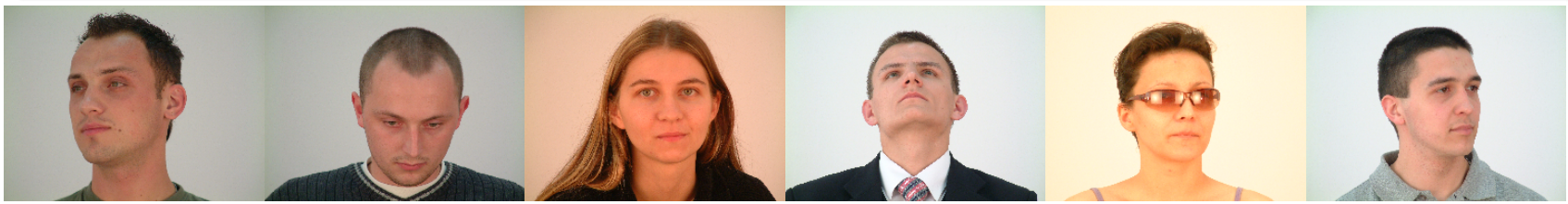
3 Experiments

4 Conclusion

Experimental Setup

Data:

- PUT Face Database [Kas08]
- 9971 face images of 100 subjects
- High resolution: 2048 x 1536
- Large variations in head pose
- Random separation into training and evaluation subjects
- Training on 600 training images from 60 subjects
- Evaluation on 3830 images from remaining 40 subjects



[Kas08]: A. Kasinski et al., The PUT face database, Image Processing and Communications 13(3-4), 2008

Experimental Setup

Setup:

- Modified Multi-Level-Approach with 2 zoom levels
- Feature generation for individual landmark detection:
  Canny edge detector [Can86]
- Combination of facial landmark in zoom level 0
- Validation of localization success [Jes01]:
  Worst result for both eyes, normalized with eye distance
- Error $e < 0.1$: both localization results inside the irises
- Error $e < 0.25$: both localization results inside the eyes

[Can86]: J. Canny, A computational approach to edge detection, Pattern Analysis and Machine Intelligence 8(6), 1986
[JKF01]: O. Jesorsky et al., Robust face detection using the hausdorff distance, AVBPA Conference, 2001

Results

Modified MLA with 2 zoom levels:

Localization rate of 97.2% within the iris achieved

	$\mathbf{e < 0.1}$	$\mathbf{e < 0.25}$
Kasinski et al. [KS10]	94.0%	-
Standard MLA with 6 zoom levels [HRB+12]	95.0%	96.5%
Standard MLA with 6 zoom levels and model interpolation [HRBS12]	96.6%	98.1%
Modified MLA with 2 zoom levels	97.2%	98.2%

[KS10]: A. Kasinski et al., The architecture and performance of the face and eyes detection system based on the haar cascade classifiers, Pattern Analysis & Applications 13(2), 2010
[HRB+12]: F. Hahmann, H. Ruppertshofen, G. Böer, R. Stannarius, and H. Schramm. Eye Localization Using The Discriminative Generalized Hough Transform. In DAGM-OAGM Joint Pattern Recognition Symposium, 2012.
[HRBS12]: F. Hahmann, H. Ruppertshofen, G. Böer, and H. Schramm. Model interpolation for eye localization using the Discriminative Generalized Hough Transform. In International Conference of the Biometrics Special Interest Group (BIOSIG), 2012.

Results

Modified MLA with 2 zoom levels and landmark combination:

Localization rate of 97.9% within the iris achieved

	$\mathbf{e < 0.1}$	$\mathbf{e < 0.25}$
Kasinski et al. [KS10]	94.0%	-
Standard MLA with 6 zoom levels [HRB+12]	95.0%	96.5%
Standard MLA with 6 zoom levels and model interpolation [HRBS12]	96.6%	98.1%
Modified MLA with 2 zoom levels	97.2%	98.2%
Modified MLA with 2 zoom levels and landmark combination	97.9%	99.1%

[KS10]: A. Kasinski et al., The architecture and performance of the face and eyes detection system based on the haar cascade classifiers, Pattern Analysis & Applications 13(2), 2010
[HRB+12]: F. Hahmann, H. Ruppertshofen, G. Böer, R. Stannarius, and H. Schramm. Eye Localization Using The Discriminative Generalized Hough Transform. In DAGM-OAGM Joint Pattern Recognition Symposium, 2012.
[HRBS12]: F. Hahmann, H. Ruppertshofen, G. Böer, and H. Schramm. Model interpolation for eye localization using the Discriminative Generalized Hough Transform. In International Conference of the Biometrics Special Interest Group (BIOSIG), 2012.

Processing Time

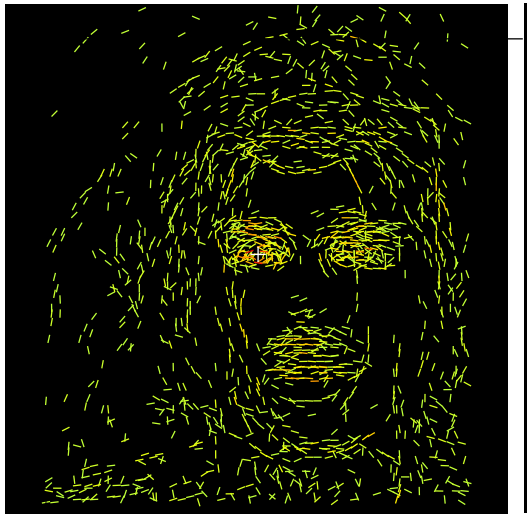
Processing Time

- Processing time for standard MLA with 6 zoom levels: 600 ms
- Processing time for modified MLA with 2 zoom levels: 970 ms
- No time measurement for landmark combination

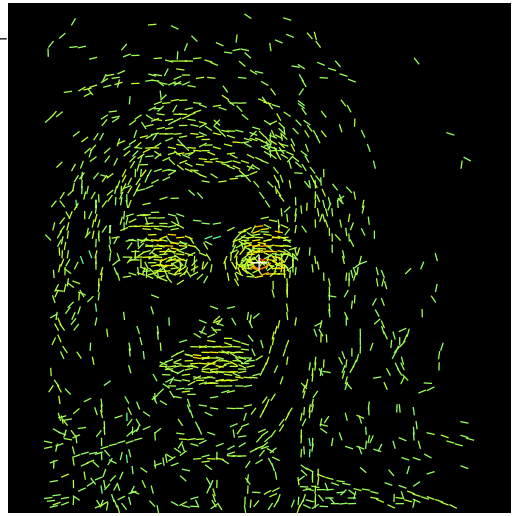
Discussion

Generated Models:

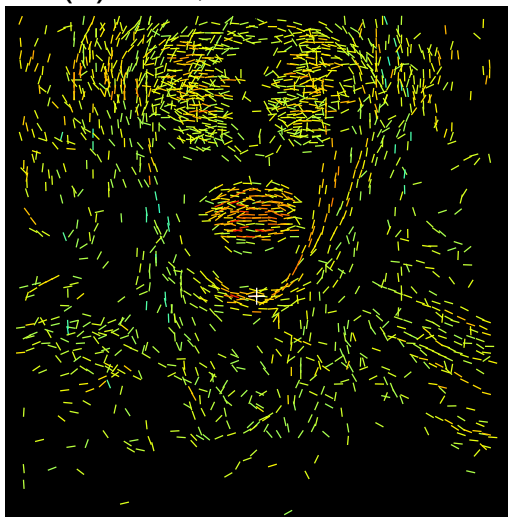
(a) right eye, zoom-level 0

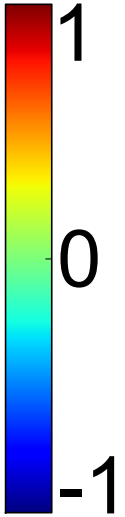


(b) left eye, zoom-level 0

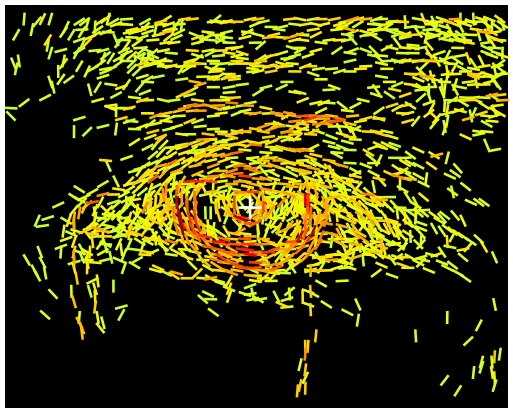


(c) chin, zoom-level 0

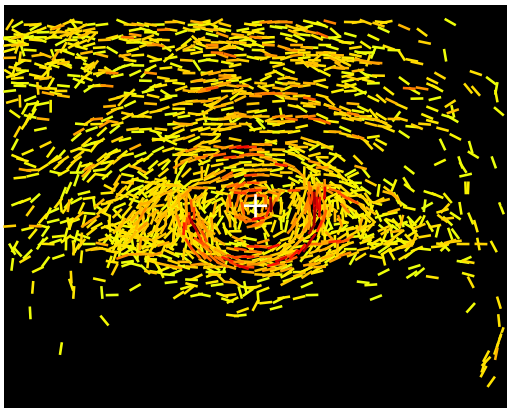




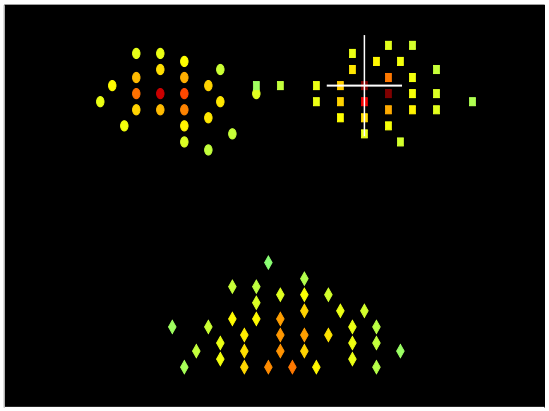
(d) right eye, zoom-level 1



(e) left eye, zoom-level 1

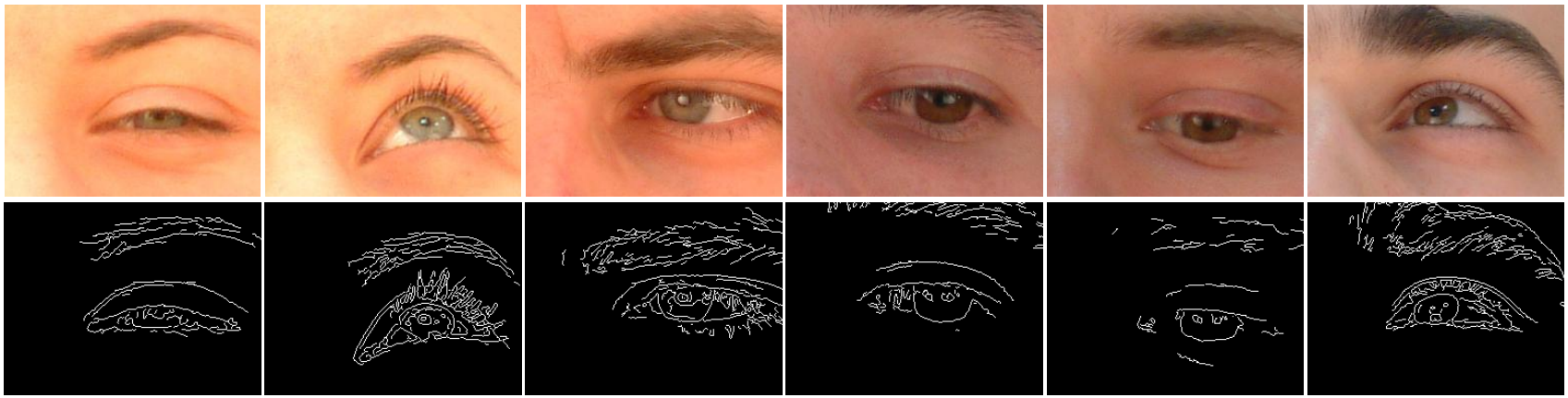


(f) left eye, higher level model



Discussion

Examples of image extracts in zoom level 1
with corresponding feature images



Conclusion

- Discriminative Generalized Hough Transform
  - can be applied to various facial landmarks
  - can be applied for facial landmarks combination
- Significant improvement of localization rate by
  - landmark combination and
  - task-specific adaptation of the Multi-Level-Approach
- Outperforms state-of-the-art results on same dataset
- Machine learning approach with minimal user interaction and no expert knowledge
- Outlook: Iterative selection of facial landmarks for combination

Thank you!