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

Ferdinand Hahmann, Gordon Böer, Hauke Schramm

Institute of Applied Computer Science
University of Applied Sciences Kiel
Grenzstraße 3, 24149 Kiel
Ferdinand.Hahmann@FH-Kiel.de

Abstract: The Discriminative Generalized Hough Transform (DGHT) is a general and robust automated object localization method, which has been shown to achieve state-of-the-art success rates in different application areas like medical image analysis and person localization. In this contribution the framework is enhanced by a novel facial landmark combination technique which is theoretically introduced and evaluated for an eye localization task on a public database. The technique applies individually trained DGHT models for the localization of different facial landmarks, combines the obtained Hough spaces into a 3D feature matrix and applies a specifically trained higher-level DGHT model for the final localization based on the given features. In addition to that, the framework is further improved by a task-specific multi-level approach which adjusts the zooming-in strategy with respect to relevant structures and confusable objects. With the new system it was possible to increase the iris localization rate from 96.6% to 97.9% on 3830 evaluation images. This result is promising, since the variation of the head pose in the database is quite large and the applied error measure considers the worst of a left and right eye localization attempt.

1 Introduction

Automatic landmark localization in face images is an important first step for various computer vision applications like person recognition and tracking, gender classification or facial expression analysis. This underlines the relevance of this task, which has attracted wide scientific interest in recent years. Especially for the eyes as the most important facial landmarks, a large number of localization approaches have been proposed. Many of these techniques have in common that they were specifically developed for the given task using expert knowledge about the object's positioning, appearance and individual adjustments.

Several eye localization approaches employ the Viola & Jones face detector [VJ04] in a first step, which uses Haar-wavelets in a boosted cascade of classifiers. Although this technique is a general object localization approach it requires additional shape constraints to be successful on the task of eye localization [CC03]. Those constraints might be either manually defined [KS10] or automatically learned [CCS04]. Frequently, the method of

Viola & Jones is only used to localize a bounding box around the face in order to perform a subsequent eye localization inside the box, using a specifically developed approach. In this second localization level the given bounding box allows for a rough determination of the eye positions [KHM08] or at least an additional restriction of the search space [KS10].

A popular eye detection approach, sometimes applied in previously located bounding boxes, is to search for circular structures, representing the pupil or the iris [TB11, DLCD04, VG08, NG12]. Despite the technical differences of these approaches, they all make use of expert knowledge about the appearance of the target object and therefore cannot be directly transferred to other localization tasks.

A general and well-known object localization method is the Generalized Hough Transform (GHT) [Bal81]. This technique uses a voting procedure to transform an image into a transformation parameter space, called Hough space, in order to determine the degree of matching between a transformed shape model and the image content. An extension of this approach are Hough Forests [GL09], which learn a direct mapping between the appearance of image patches and the votes in the Hough-space. Hough Forests have already been used in different application scenarios like mouth localization for audio-visual speech recognition [FGVG09] or classification of facial expressions [FYN+12]. In both cases, however, the eyes were localized by searching for circular structures by the method of [VG08].

The idea of splitting the target object into different parts is utilized in various object localization methods [Oka09, LLS08]. Furthermore, in [CCS04] a procedure for eye localization is presented, which detects 17 different facial features using the method of Viola & Jones and learns their relative positions in a geometric model.

The success of the GHT heavily depends on the applied shape models. Therefore, the Discriminative Generalized Hough Transform (DGHT) [RBS08, Rup13] extends the GHT by a fully automated and general learning method for model generation. In [HRB+12] the DGHT was successfully used for eye localization and in [HRBS12] improved results for this task have been achieved by combining the localizations of both eyes with prior knowledge about the expected eye distance vector. In this contribution, we present a novel method for combining different landmarks in two hierarchical DGHT-based localization levels. In the first level the standard DGHT technique is used to determine the localization probabilities for different facial landmarks which are afterwards combined in a 3D feature matrix. On these features the DGHT training approach (Section 2.3) is applied to train a 3D DGHT localization model which learns the relative positioning of each of the landmarks given in level one. In addition to that, a modified multi-level approach is introduced in this work, which achieves an improved robustness by replacing the gradual reduction of the search space in [HRBS12] with a direct zooming into the eye region (Section 2.2).

Both changes have been evaluated on the public PUT Face Database (Section 3) and led to a significant improvement over the standard method (Section 3.3). The paper closes with a discussion of the experimental results (Section 4) and a conclusion (Section 5).

2 Method

2.1 Discriminative Generalized Hough Transform

The Generalized Hough Transform (GHT) is a general method for object localization. It is based on a geometric model which stores model points representing features of the searched-for object in relation to a reference point. The GHT transforms an image space into a model transformation parameter space, from which the optimal object transformation into the given image can be derived. Although, the transformation, considered in the GHT, is in general not restricted, object localization can be based on translation parameters only by determining the position with the highest degree of matching between the model and the feature image. This restriction, used throughout this contribution, allows for quick processing and works well if the model sufficiently represents the object's variability.

A cell c_i of the quantized parameter space, also called Hough space H, represents an image position and reflects the degree of matching between model M and feature image X_n by the number of corresponding feature points $\mathbf{e}_k$ and model points $\mathbf{m}_j$. The Hough space is generated by an efficient voting procedure and can be formalized by

$$H(\mathbf{c}_i|X_n) = \sum_{\forall \mathbf{e}_k \in X_n} \sum_{\forall \mathbf{m}_j \in M} \begin{cases} 1, & \text{if } \mathbf{c}_i = \mathbf{e}_k - \mathbf{m}_j \text{ and } d(\mathbf{e}_k, \mathbf{m}_j) \leq \vartheta \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

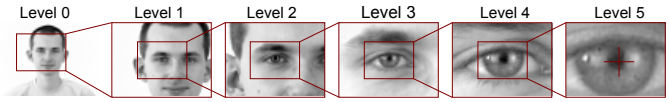
whereas $d(\mathbf{e}_k, \mathbf{m}_j)$ specifies the distance between the value of the feature and model point, which has to be lower than a threshold ϑ. In case of using edge features, the values of the feature and model points are usually given by the gradient direction φ_k and model point orientation φ_j. Thus, the distance is determined by $d(\mathbf{e}_k, \mathbf{m}_j) = |\varphi_k - \varphi_j|$.

Since the result of equation (1) highly depends on the quality of the model, the DGHT comprises an automatic training procedure to generate optimal models. This training procedure uses the Hough space, resulting from the explained voting procedure, to extract the model point specific contributions $f_j(c_i|X_n)$ which is the number of votes from model point m_j into the Hough cell c_i. These model point specific votes are recombined into a Maximum-Entropy distribution [Jay57] to ensure maximum objectivity:

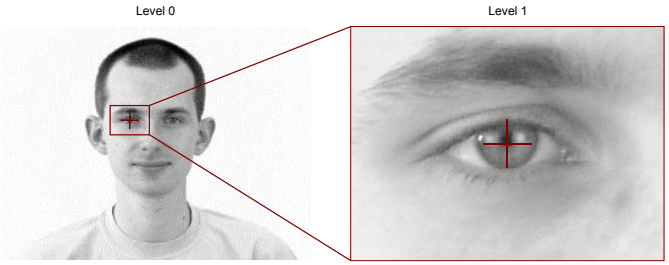
$$p_\Lambda(c_i|X_n) = \frac{\exp\left(\sum_j \lambda_j \cdot f_j(c_i|X_n)\right)}{\sum_k \exp\left(\sum_j \lambda_j \cdot f_j(c_k|X_n)\right)} \tag{2}$$

The introduced model point specific weights $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_J\}$, which may be negative, are optimized to reflect the importance of a model point for the correct localization and for the distinction of similar objects. For further details on the DGHT method, we refer the reader to [HRBS12].

Note that the DGHT can be used for 2D and 3D images. Although edge images are used most of the time, other features may also be utilized.



(a) Standard Multi-Level-Approach with 6 zoom levels



(b) Modified Multi-Level-Approach with 2 zoom levels

Figure 1: Comparison of the standard Multi-Level-Approach with the modified Multi-Level-Approach

2.2 Modified Multi-Level-Approach

The Multi-Level-Approach (MLA) is a zoom-in strategy, in which the resolution is gradually increased around the suspected target point. By decreasing the considered image extract and increasing the resolution in each zoom level the visible structures range from global and coarse to local but fine structures. Since the different DGHT models, applied in the MLA, are specifically trained on the respective image extracts they learn relevant and discriminative structures in each zoom level. Therefore, the MLA is a good trade-off between keeping sufficient target object details and suppressing noise and confusing objects.

The MLA presented in prior publications [RKL+11] doubled the resolution and halved the size of the image extract in each zoom level, therefore keeping the number of pixel constant. For the task of eye localization on the public PUT Face Database this procedure was used with 6 zoom levels in [HRBS12] (Figure 1(a)).

It could be shown in [HRB+12], that the standard MLA procedure is prone to a confusion of the eyes in zoom levels, where both eyes might be visible while important discriminating structures are missing. Consequently, the modified MLA uses a higher resolution in the first zoom level in order to ensure a more accurate target localization than the standard approach. This especially aims at a reliable distinction between both eyes. In the second zoom level of the modified MLA the image extract is already restricted to a region containing only a single eye which excludes a confusion with the other eye. This image extract already has the full resolution and is used for the final localization (Figure 1(b)).

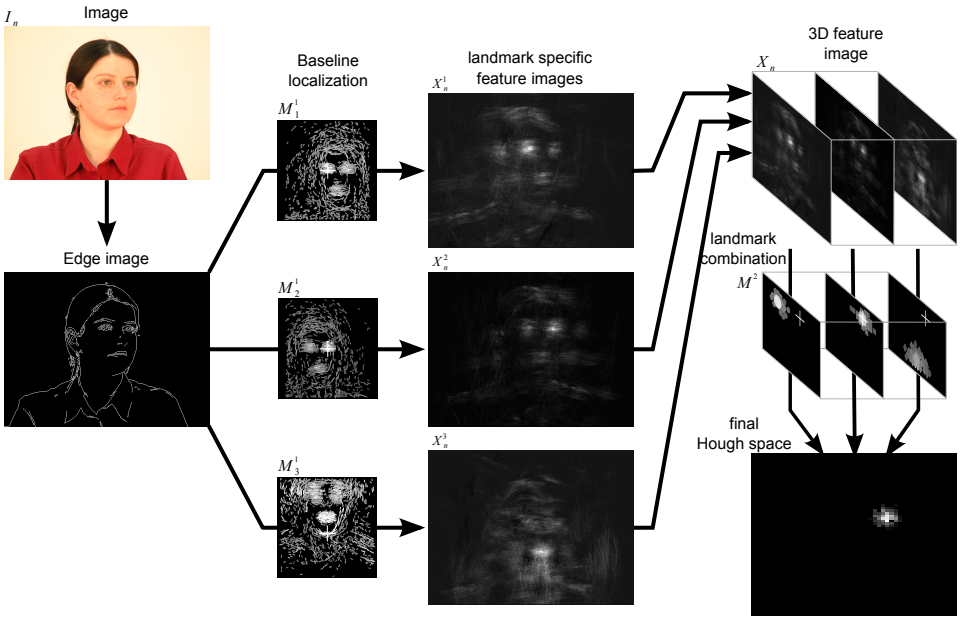


Figure 2: Illustration of the process of landmark combination: According to the standard procedure for landmark localization the image is transformed into a feature image (edge image). Subsequently, the edge-based DGHT models M_1^1, M_2^1, M_3^1 are utilized for single localization of both eyes (M_1^1 and M_2^1) and the chin (M_3^1). The thereby generated probability distributions X_n^1, X_n^2, X_n^3 are combined in a 3D feature image X_n. On this 3D feature image, a discriminatively trained 3D model M^2 is applied for the final localization. Hence, M^2 combines the information about the probable position of the individual facial landmarks related to the target landmark.

2.3 Landmark combination

The landmark combination occurs in two levels. In the first level, for each landmark l special DGHT models M_l^1 are trained by using the standard DGHT procedure (section 2.1) and canny edge images [Can86] as features. By applying these models to new images, individual probability distributions X_n^l (see Equation (2)) of target localizations are generated. Since (i) with the distribution of a landmark (e.g. left eye), the position of another landmark (e.g. right eye) can be estimated and (ii) the DGHT is neither restricted to edge images nor to 2D images, these landmark specific distributions are combined in a new 3D feature image $X_n = \{X_n^1, ..., X_n^L\}$ for the next localization level. For a given set of N training images, the corresponding 3D features $X_1, ..., X_N$ are used to train a higher-level 3D DGHT model M^2 in the second level utilizing the standard DGHT training approach (section 2.1). This model captures the relative position of the landmarks to each other and provides the final localization result.

The feature value of a point $\mathbf{e}_k$ in X_n^l specifies the probability $p_l(\mathbf{e}_k|I_n)$ (calculated by Equation (2)) of landmark l being localized at position $\mathbf{e}_k$ for the given image I_n and model M_l^1. Thus, it represents the certainty of the underlying localizer in level one. This

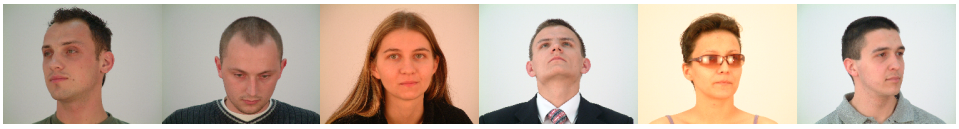


Figure 3: Illustration of the large head position variability contained in the PUT database.

important source of information should be directly incorporated into the GHT voting procedure of level two in order to increase the influence of areas with high localization reliability. Therefore, the standard voting procedure (Equation (1)) is adapted to directly vote with the feature value $p_l(\mathbf{e}_k|I_n)$ instead of voting with the value 1. In addition to that, a summation over the L landmarks has to be done in order to combine the results from the different landmark localizations in level one. This leads to the following modified voting procedure for the GHT in level two:

$$H(\mathbf{c}_i|X_n) = \sum_{l}^{L} \sum_{\forall \mathbf{e}_k \in X_n^l} \sum_{\forall \mathbf{m}_j \in M_l^1} \begin{cases} p_l(\mathbf{e}_k|I_n), & \text{if } \mathbf{c}_i = \mathbf{e}_k - \mathbf{m}_j \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Note that the standard DGHT training approach (see section 2.1) is used for optimizing the models of both described localization levels.

3 Experiments

3.1 Data

The experiments were conducted using the public PUT Face Database [KFS08] in training and evaluation, which includes 9971 images from 100 subjects. The high resolution (2048×1536 pixels) color images were taken under controlled lighting conditions in front of a uniform background. Since 30 facial landmarks are provided for each image in this corpus it is very well suited for investigating the presented landmark combination technique. Despite of the neutral background, the corpus is challenging due to the strong variability of head positions (see Figure 3).

As in [HRB$^+$12, HRBS12], the 100 different subjects in the corpus were divided into a training set, containing 60 subjects, and an evaluation set with the remaining 40 subjects. For better comparability the evaluation corpus is identical to [HRB$^+$12, HRBS12] and includes 3830 images. The training was performed on 600 images which have been randomly selected from the training set.

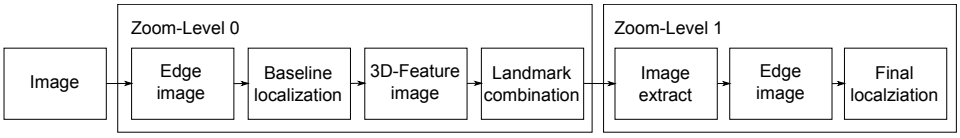


Figure 4: System overview of modified multi-level approach with landmark combination.

3.2 Setup

In the modified MLA (Section 2.2), the resolution is reduced by a factor of eight in zoom level 0 (see Figure 1(b)). Around the target point, localized in this level, an image extract with original resolution and the size of one-eight of the complete image is taken for the second and final localization step. The system works with Canny edge features [Can86] and applies a standard DGHT training procedure for generating the specific GHT models for the two localization levels. All described experiments have been performed using a 64 bit system with an Intel Xeon W3520 with 2.66 GHz and 24 GB RAM.

To further enhance the robustness of the modified MLA in zoom level 0, a combination of three landmarks (both eyes and chin) is applied by the landmark combination procedure described in Section 2.3: Using standard DGHT models, based on Canny edge features, three probability distributions for the landmark locations are generated (see Section 2.3). These distributions are combined into a 3D feature image X_n, ignoring values of less than 0.01 in order to decrease the processing time and to reduce noise. With a specifically trained 3D DGHT model a robust target localization in zoom level 0 is performed using the modified voting procedure (equation (3)) and the result is handed over to zoom level 1. Figure 4 gives an overview of the system with the modified MLA and landmark combination.

To determine the localization rate, the measurement explained in [JKF01] is used, in which the larger localization error of both eyes is normalized with the eye distance. An error of less than 0.1 / 0.25 therefore corresponds to a localization result approximately located within the iris / eye. Due to slightly inaccurate annotations, provided by the PUT Face database, an error distance less than 0.1 is not meaningful since the inaccuracy would be higher than the error distance.

3.3 Results

By using the modified MLA a success rate of 97.2% for a localization within in the iris could be achieved on the evaluation corpus. This is an improvement of 0.6% compared to the previously best published result and a gain of 2.2% to the published result obtained with a standard method (Table 1). A good indicator for the localization robustness of zoom level 0 of the MLA is given by the number of target points lying outside the optimal image extract. In comparison to the standard MLA approach and a comparable image extract, this number could be reduced from 130 to 50 by applying the described modifications.

Table 1: Experimental results comparing different systems for different fault tolerances.

	$\mathbf{e < 0.1}$	$\mathbf{e < 0.15}$	$\mathbf{e < 0.2}$	$\mathbf{e < 0.25}$
Kasinski et al. [KS10]	94.0%	-	-	-
Standard MLA with 6 zoom levels [HRB$^+$12]	95.0%	95.4%	96.0%	96.5%
Standard MLA with Model interpolation [HRBS12]	96.6%	97.1%	97.6%	98.1%
Modified MLA with 2 zoom levels	97.2%	97.6%	98.0%	98.2%
Modified MLA with landmark combination	97.9%	98.5%	98.9%	99.1%

A further improvement of the localization robustness in zoom level 0 of the modified MLA could be achieved by using the described landmark combination technique for three facial landmarks. This measure reduced the number of target points lying outside the optimal image extract to 20 and therefore improved the error rate to 97.9% for iris localization. Considering a less restricted fault tolerance, a localization inside the eye was achieved in 99.1% (Table 1). The generated landmark localization models M_l are shown in Figure 5 (a) to (e). The model points are represented as lines to visualize their orientation while the gray value illustrates their weight. Figure 5 (f) displays the 3D DGHT model of zoom level 0. Here, the symbol of a model point indicates the corresponding landmark and the gray value represents again the individual weight as obtained by the discriminative training process.

4 Discussion

The significant improvement of the modified MLA can be mostly explained by a better discrimination between both eyes. This is due to an improved localization robustness in zoom level 0 which may be assigned to a better and more detailed DGHT model with a strong focus on both eyes (e.g. see Figure 5(a)). Comparing the models of the standard and the modified MLA, it is noticeable that the average number of model points has substantially increased from 357 to 1807. This rise results from the higher resolution in the modified MLA which leads to an increase of feature points and shape variation, compensated by a larger number of model points.

It is interesting to note that only a few model points of a given localization model are relevant for a single image. Therefore, the percentage of model points, voting for the best Hough cell, is only 11% on average for the standard MLA. For the modified MLA, however, this number is even smaller and amounts to only 5% which underlines the fact that the overall size of the model results from the large variation over all images.

The higher number of feature and model points also explains an increase of the processing time from about 600 ms for the standard MLA to 970 ms for the modified approach. Note, the system has not been optimized for runtime performance yet.

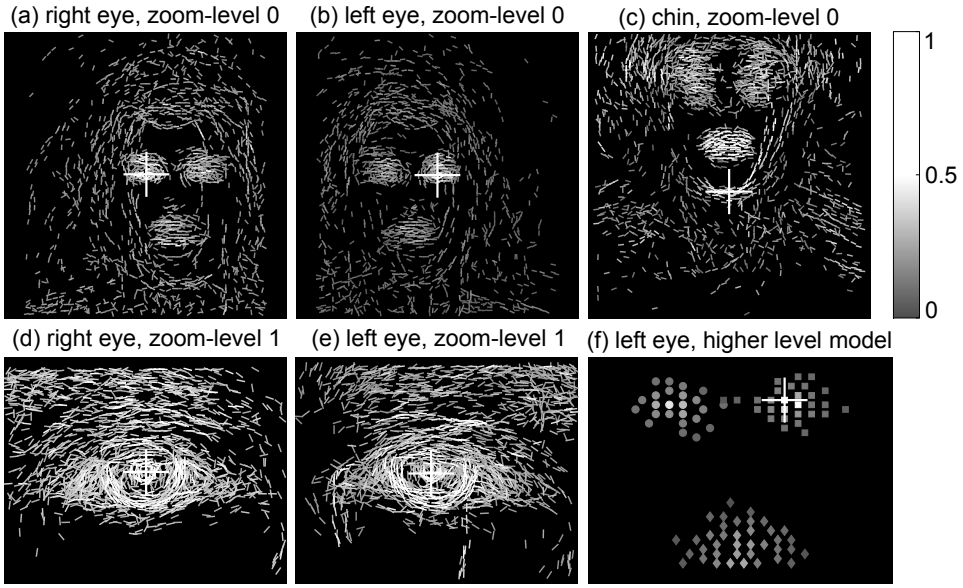


Figure 5: (a) to (e): DGHT models used for baseline landmark localization, where the gray value denotes the individual model point weight. (f): 3D DGHT model used for landmark combination. The symbols illustrate the corresponding landmark layer (circle: right eye layer, square: left eye layer, diamond: chin layer) and the gray value represents the model point weight. Note, that model points with negative weights, which ensure a better discrimination of similar object, are not shown for clarity since they only play a minor role in these experiments.

A clear advantage of the DGHT approach in comparison to most other state-of-the-art localization techniques is the visual interpretability of the models, which reveal the shape of the most discriminative structures as well as the importance of each individual model point. In the localization models of zoom level 0 (Figure 5 (a) to (c)), for example, it can be seen that the localization heavily relies on both eyes and the mouth. The nose, is hardly represented by model points since it is a facial structure which is rarely visible in the feature images and, in addition to that, highly variable (Figure 6(b)). Another interesting aspect, which can be seen in the model images, is that they represent different head positions at the same time to cope with the strong head pose variation contained in the PUT database. For demonstrating this aspect, Figure 6 shows (a) some original images with overlayed model, (b) the corresponding edge feature images, and (c) the model points which voted for the best localization hypotheses.

In zoom level 1 (Figure 5 (d) und (e)), the eye localization models clearly display two concentric circles, representing the iris and the pupil respectively. This search structure has also been integrated in many other systems by using expert knowledge [TB11, DLCD04, VG08, NG12] which demonstrates that the DGHT may learn and incorporate this kind of knowledge fully automatically without the need for a detailed insight into the localization problem. Other model points, contained in the localization model, represent the eyebrows and eyelids, which have different positions depending on the viewing direction, and reflec-

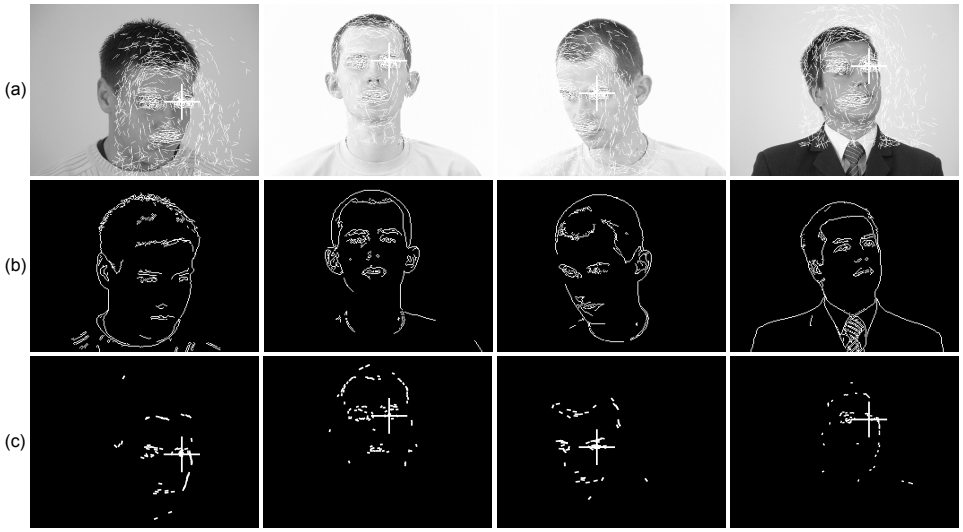


Figure 6: (a) Original images with overlayed model, (b) corresponding feature images, (c) model points which voted for the best localization hypotheses in the respective image. The used model is identical to Figure 5(b).

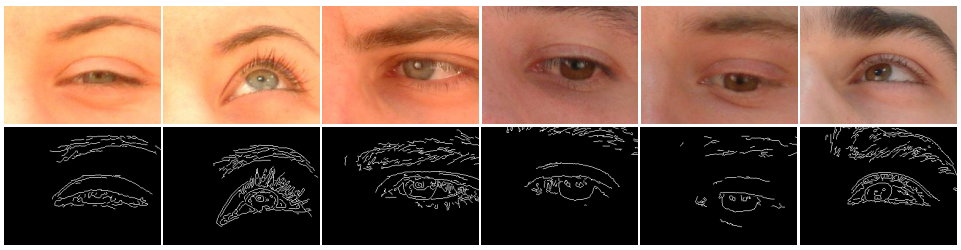


Figure 7: Examples of image extracts in zoom level 1 with corresponding feature images

tions of the flash on the eyeball (see Figure 7). This also contradicts the common modeling assumption that the sclera is always brighter than the iris, which in turn is brighter than the pupil.

When studying the model for the landmark combination (Figure 5 (f)), it is apparent that model points of the chin have a large scattering and very similar weights while the important points, representing the eye, are much more focussed. This is because of the lower reliability of the chin localizer, which has a mean error of 49 pixels in comparison to 21 and 23 pixels for left and right eye, respectively. It is also worth mentioning that the increased robustness of the landmark combination goes together with a loss in accuracy since the model is more blurred. The increase of the eye localization mean error to 29 and 31 pixels for the left and right eye in zoom level 0 after the landmark combination is compensated by the more precise edge based localization model applied in zoom level 1. In zoom level 1, the mean error was reduced to 12 and 10 pixel.

5 Conclusion

In this contribution two novel techniques for an improved eye localization in portrait images based on the Discriminative Generalized Hough Transform have been presented. By using a task-specific multi-level strategy and a novel facial landmark combination technique it was possible to increase the iris localization rate from 96.6 to 97.9%. This result is promising, since the variation of the head pose in the used public PUT face database is quite large and the applied error measure considers the worst of a left and right eye localization attempt.

The general standard MLA, which gradually zooms into the target object by halving the search space in each level, could be shown to be suboptimal. A more task-specific approach, adjusting the zooming strategy with respect to the relevant structures and confusable objects, may significantly improve the success rate. For the given task of eye localization, with two very confusable objects, a good strategy is an early limitation of the search space to a region, covering only a single eye.

The novel approach for facial landmark detection, which has been introduced in this paper, could be combined with the modified MLA and further increases the robustness of the system in the first zoom level. With this framework, it could be shown for the first time that the DGHT is applicable for both, the individual localization of various landmarks and combined usage in a higher-level localization model. This comes together with the possibility to visually interpret the generated DGHT models in the different stages unveiling discriminative structures and important model parts.

Although in this contribution only three facial landmarks, both eyes and the chin, have been combined with the novel method, the approach may theoretically incorporate an unlimited number. Since the applied discriminative training procedure identifies and penalizes model points of weak landmarks, not supporting the localization, it is possible to select the most discriminative ones from a large set of candidates. A systematic evaluation of this idea, selecting optimal landmarks in an iterative training procedure as well as evaluation on other databases and comparision with other methods will be done by our group in the next future.

Acknowledgments. This work is partly funded by the Innovation Foundation Schleswig-Holstein under the grant 2010-90H. Additionally, the authors are grateful to the anonymous reviewers for their valuable comments.

References

[Bal81] D.H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[CC03] D. Cristinacce and T. Cootes. Facial feature detection using adaboost with shape constraints. In *British Machine Vision Conference (BMVC)*, 2003.

[CCS04] D. Cristinacce, T. Cootes, and I. Scott. A multi-stage approach to facial feature detec-

tion. In *British Machine Vision Conference (BMVC)*, 2004.

[DLCD04] T. D'Orazio, M. Leo, G. Cicirelli, and A. Distante. An algorithm for real time eye detection in face images. In *International Conference on Pattern Recognition (ICPR)*, 2004.

[FGVG09] G. Fanelli, J. Gall, and L. Van Gool. Hough transform-based mouth localization for audio-visual speech recognition. In *British Machine Vision Conference (BMVC)*, 2009.

[FYN+12] G. Fanelli, A. Yao, P. Noel, J. Gall, and L. Van Gool. Hough forest-based facial expression recognition from video sequences. In *Trends and Topics in Computer Vision*, 2012.

[GL09] J. Gall and V. Lempitsky. Class-specific hough forests for object detection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[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.

[Jay57] E.T. Jaynes. Information theory and statistical mechanics. *The Physical review*, 106(4):620–630, 1957.

[JKF01] O. Jesorsky, K. Kirchberg, and R. Frischholz. Robust face detection using the hausdorff distance. In *International Conference on Audio-and Video-Based Biometric Person Authentication (AVBPA)*, 2001.

[KFS08] A. Kasinski, A. Florek, and A. Schmidt. The PUT face database. *Image Processing and Communications*, 13(3-4):59–64, 2008.

[KHM08] B. Kroon, A. Hanjalic, and S.M.P. Maas. Eye localization for face matching: is it always useful and under what conditions? In *International Conference on Content-based image and video retrieval (CIVR)*, 2008.

[KS10] A. Kasinski and A. Schmidt. The architecture and performance of the face and eyes detection system based on the Haar cascade classifiers. *Pattern Analysis & Applications*, 13(2):197–211, 2010.

[LLS08] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1):259–289, 2008.

[NG12] O. Nikisins and M. Greitans. Local binary patterns and neural network based technique for robust face detection and localization. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2012.

[Oka09] R. Okada. Discriminative generalized hough transform for object dectection. In *Conference on Computer Vision (ICCV)*, 2009.

[RBS08] A. Recuero, P. Beyerlein, and H. Schramm. Discriminative optimization of 3D shape models for the Generalized Hough transform. *International Conference and Workshop on Ambient Intelligence and Embedded Systems (AMIES)*, 2008.

[RKL+11] H. Ruppertshofen, D. Künne, C. Lorenz, S. Schmidt, P. Beyerlein, Z. Salah, G. Rose, and H. Schramm. Multi-Level Approach for the Discriminative Generalized Hough Transform. In *Computer- und Roboterassistierte Chirugie (CURAC)*, 2011.

[Rup13] H. Ruppertshofen. *Automatic Modeling of Anatomical Variability for Object Localization in Medical Images*. PhD thesis, Otto-von-Guericke University Magdeburg, 2013.

[TB11] F. Timm and E. Barth. Accurate eye centre localisation by means of gradients. In *Conference on Computer Vision Theory and Applications (VISAPP)*, 2011.

[VG08] R. Valenti and T. Gevers. Accurate eye center location and tracking using isophote curvature. In *Computer Vision and Pattern Recognition (CVPR)*, 2008.

[VJ04] P. Viola and M.J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.