Model Interpolation for Eye Localization Using the Discriminative Generalized Hough Transform

Ferdinand Hahmann, Heike Ruppertshofen, Gordon Böer, Hauke Schramm

Institute of Applied Computer Science
University of Applied Sciences Kiel
Grenzstraße 3
24149 Kiel
Ferdinand.Hahmann@FH-Kiel.de
Heike.Ruppertshofen@Philips.com
Gordon.Boeer@FH-Kiel.de
Hauke.Schramm@FH-Kiel.de

Abstract: The Discriminative Generalized Hough Transform (DGHT) is a general method for the localization of arbitrary objects with well-defined shape, which has been successfully applied in medical image processing. In this contribution, the framework is used for eye localization in the public PUT face database. The DGHT combines the Generalized Hough Transform (GHT) with a discriminative training procedure to generate GHT shape models with individual positive and negative model point weights. Based on a set of training images with annotated target points, the individual votes of model points in the Hough space are combined in a maximum-entropy probability distribution and the free parameters are optimized with respect to the training error rate. The estimated model point specific weights reflect the important model structures to distinguish the target object from other confusable image parts. Additionally, the point weights allow for a determination of irrelevant parts in the model, which can be eliminated to make space for new model point candidates from training images with high localization error. The iterative training procedure of weight estimation, point elimination, testing on training images, and incorporation of new model point candidates is repeated until a stopping criterion is reached. Furthermore, the DGHT framework incorporates a multi-level approach, in which the searched region is reduced in 6 zooming steps, using individually trained shape models. In order to further enhance the robustness of the method, the DGHT framework is, for the first time, extended by a linear model interpolation for the trained left and right eye model. An evaluation on the PUT face database has shown a success rate of 99% for iris detection in frontal-view images and 97% if the test set contains a large head pose variability.

1 Introduction

The eyes are two important and clearly visible features of the human face, whose position can be used to deduce further relevant information like the head pose or size. Many face related applications rely on an exact and reliable eye localization, making this field an important research area with many different approaches.

Interestingly, most state-of-the-art systems combine an existing method for face localization with specialized techniques to determine the exact eye position in the given face. In addition, many systems make extensive use of prior knowledge about the shape and appearance of the eyes and their distance and can therefore not be easily generalized to other object detection tasks.

The most popular face detector, used by many eye detection frameworks, is the method developed by Viola and Jones [VJ04], which applies a combination of Haar-like features in a boosted cascade of classifiers to provide a face bounding box. Additionally, many techniques make use of a-priori knowledge about the localization task. This can, for example, be achieved by restricting the search space to certain areas of the previously localized face ([KHM08, KS10]) or by using a-priori constraints between facial landmarks, mostly both eyes like in [KS10]. Only in a few approaches, those constraints are learned from training data, like in [CCS04], where 15 facial features are combined to enhance the robustness of the localization procedure. Other approaches use a-priori knowledge about the typical shape and appearance of an eye, e.g. by searching for dark circular objects. In [TB11] a vector field of image gradients is analyzed to identify the center of circular structures in a previously localized face. An additional weighting of the gradients, depending on the gray values of the hypothesized target center, supports the detection of particularly dark circles like the pupil. An alternative to this approach is the application of the Hough Transform for the localization of circular objects, like in [DLCD04].

An extension of the Hough Transform, which can be applied to arbitrary shapes, is the Generalized Hough Transform (GHT) [Bal81]. This technique, which has already been applied to eye localization [Sch00], measures the match between a shape model and a feature image by using a simple voting procedure. The method is general but heavily relies on the quality of a given point model for the target shape.

The GHT is a well-known and widely applied approach for object localization, which is why many extensions have been developed. Some of the extensions produce a scale or rotational invariant GHT to reduce the necessary transformations. An overview about some of these extension can be found in [KTT99]. Other approaches (such as [GL09, Oka09, LLS08]) split the target object into parts and combine them in a discriminative manner in order to ease the localization of objects with a high variability.

To this end, we developed the Discriminative Generalized Hough Transform (DGHT) [RLS+11], which automatically learns discriminative shape models with individual positive and negative model point weights for usage in a standard GHT voting procedure. The technique has been successfully applied to various object detection tasks in medical image analysis and has recently been ported to the problem of eye detection [HRB+12]. In this work, the robustness of the DGHT framework is further improved by (1) utilizing a joint detection of both eyes with individual GHT models and (2) incorporating prior knowledge about the vector between the eyes, which has been obtained from training data.

This paper is organized as follows. In Section 2 we describe the used method, including the GHT (2.1), the discriminative training (2.2), a zooming strategy to improve the robustness and processing time (2.3), and the new extension of linear model interpolation (2.4). Section 3 contains a description of the used database and Section 4 describes the experi-

mental setup. The results are presented in Section 5, followed by a discussion in Section 6. Section 7 concludes our contribution.

2 Method

2.1 GHT

The GHT, introduced by Ballard in 1981 [Bal81], is a general and well-known approach for object localization, which belongs to the category of template-matching techniques. It uses a point model M to represent the structure of the searched-for object in relation to a reference point, which is the target point for the localization task.

The GHT transforms a feature image into a parameter space, called Hough space, utilizing a simple voting procedure. The Hough space consists of accumulator cells (Hough cells), representing possible target point locations and, potentially, shape model transformations. The number of votes per accumulator cell reflects the degree of matching between the (transformed) model and the feature image.

Since each additional parameter in a model transformation leads to an additional dimension in the Hough space, no model transformation, apart from translation, is considered in this work. Instead, moderate object variability with respect to shape, size, and rotation is learned into the model. Although in general arbitrary features could be used in the GHT, our framework applies the standard Canny edge detector [Can86] to extract simple edge features.

The voting procedure which transforms a feature image X_n into the Hough space H by using the shape model M can be described by

$$H(\mathbf{x}) = \sum_{\forall \mathbf{e}_i \in X_n} \sum_{\forall \mathbf{m}_j \in M} \begin{cases} 1, & \text{if } \mathbf{x} = \mathbf{e}_i - \mathbf{m}_j \text{ and } \arccos(< \mathbf{e}_i, \mathbf{m}_j >) \leq \vartheta_\varphi \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The combination of an edge point $\mathbf{e}_i$ with a model point $\mathbf{m}_j$ increments the appropriate accumulator cell at position $\mathbf{x}$ if the scalar product of $\mathbf{e}_i$ and $\mathbf{m}_j$ is below a given threshold ϑ_φ.

2.2 Discriminative Training of GHT Models

The usage of a representative and discriminative shape model in the GHT is crucial for a good localization result. Therefore, the discriminative training procedure, applied by our DGHT framework, aims at optimizing the model with respect to the error rate on training images. This includes the estimation of individual model point weights as well as the addition of relevant new structures to the model or the removal of unimportant points. The latter takes place during an iterative training procedure which is explained in Section 2.2.2.

2.2.1 Estimation of Model Point Weights

It is quite evident that the model points of a given shape model are of different importance for the localization problem. While points that allow for a good discrimination between the target object and other confusable structures are particularly useful, others may even mislead the detection procedure by fitting to wrong image parts. Therefore, the application of a model point specific discriminative weighting scheme appears to be a reasonable measure when using the GHT.

The estimation of the individual weights for the J points, contained in a given shape model, is achieved by (1) separating the Hough space votes coming from every single model point, (2) recombining those contributions in a weighted manner, and (3) optimizing the introduced weights with respect to an error measure.

The theory is based on describing the GHT as a probabilistic framework, in which the Hough space is interpreted as a posterior probability distribution $p(c_i|X_n)$. This distribution can be estimated, for example, from the relative frequencies of votes in each Hough cell c_i. The GHT-based localization task, which searches for the cell with the highest number of votes, can now be formulated as the Bayes classifier $\hat{c} = \arg\max_{c_i} p(c_i|X_n)$.

In order to identify the individual importance of each single model point, it is necessary to split the Hough cell votes into model point specific parts. To this end, the characteristic function

$$f_j(c_i, X_n) = v_{i,j} \tag{2}$$

is introduced, which denotes the number of votes $v_{i,j}$ from model point m_j in Hough cell c_i for a given feature image X_n.

Since the feature functions only consider the contributions of single model points they must be recombined in order to preserve the constraints from the GHT voting procedure for the entire model. In the DGHT framework this is achieved by using the maximum entropy distribution [Jay57], which assures maximum objectivity and introduces model point specific weights λ_j.

$$p_\Lambda(c_i|X_n) = \frac{\exp\left(\sum_j \lambda_j \cdot f_j(c_i, X_n)\right)}{\sum_k \exp\left(\sum_j \lambda_j \cdot f_j(c_k, X_n)\right)} \tag{3}$$

The estimation of the free parameters $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_J\}$ from the side conditions (2) leads to an optimal approximation of the training data distribution but not necessarily to a minimal error rate. Therefore, the parameter optimization in the DGHT follows a minimal classification error (MCE) training approach [JK92], first applied in the field of automatic speech recognition [Bey98]. This technique minimizes a smoothed error measure over a set of N training images and I Hough cells.

$$E(\Lambda) = \sum_{n=1}^{N} \sum_{i=1}^{I} \varepsilon(c_i, \widetilde{c}_n) \cdot \frac{p_\Lambda(c_i|X_n)^\eta}{\sum_k p_\Lambda(c_k|X_n)^\eta} \tag{4}$$

Here, η controls the influence of alternative location hypotheses on the error measure

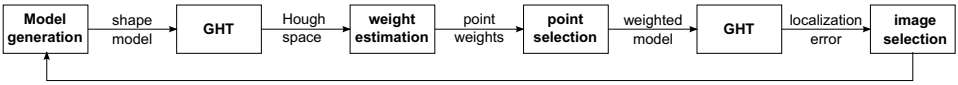

Figure 1: Illustration of the iterative training procedure.

and $\varepsilon(c_i, \widetilde{c}_n)$ denotes the error between the Hough cell c_i and the target cell $\widetilde{c}_n$ which is determined by the Euclidean distance $\|c_i, \widetilde{c}_n\|_2$.

The optimization of the model point weights Λ over the error measure $E(\Lambda)$ is achieved by applying the method of steepest descent. Although this technique does not guarantee to reach a global minimum, recent experiments [RLS+11, HRB+12] clearly demonstrate significant improvements over other state-of-the-art weighting strategies.

The estimated model point weights are directly incorporated into a standard GHT voting procedure by incrementing the value of a Hough cell c_i by $\lambda_j \cdot f_j(c_i, X_n)$ for each model point m_j.

$$H_\Lambda(c_i|X_n) = \sum_{j=1}^{J} \lambda_j \cdot f_j(c_i, X_n) \tag{5}$$

The localization result $\hat{c}$ is then given by $\hat{c} = \arg\max_{c_i} H_\Lambda(c_i|X_n)$. This leads to the same results as applying the log-linear feature combination (3), used for the training, since neither the normalization term in the denominator nor the exponential function has an influence on the result of the arg max function.

2.2.2 Iterative Training

Assigning optimized weights to a given static shape model can substantially improve the localization performance as shown in [RLS+11]. Since the assigned weights are directly incorporated into the GHT weighting scheme, it is furthermore obvious that the elimination of model points with a small absolute weight does not have a significant influence on the final localization result. Consequently, an initially given shape model can be substantially reduced, keeping only a small amount of the most relevant positively and negatively weighted structures. In an iterative procedure, illustrated in Figure 1, this technique can be applied in order to repeatedly expand the model with the shapes of unrecognized target objects and, for negative weighting, the most important confusing structures contained in the training corpus.

The framework of the training is described in more detail in the following. In the first step of the iterative training, a small subset of training images is randomly selected, and an initial shape model is created by overlaying edge points from a predefined region around the annotated target point. This model is subsequently used with equal point weights in a standard GHT procedure to localize the target points in all training images. The features $f_j(c_i, X_n)$ and error measures $\varepsilon(c_i, \widetilde{c}_n)$ are extracted from the resulting Hough spaces and utilized to compute the updated weights. In the next step, model points with a low absolute weight are removed from the model, which is afterwards tested on the whole training dataset. Since the estimation of this first shape model is based on very few images,

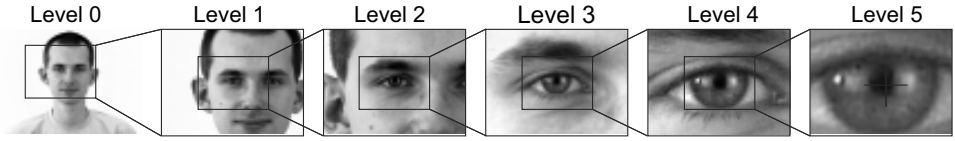

Figure 2: Image extracts with different resolutions in the multi-level approach.

it can most likely not cover the whole variability contained in the training data. Therefore, in our framework, the model is expanded by additional structures taken from images with high localization error. To this end, edge points from a region around the target object and the most confusable object are added to the model for the next iteration. The integration of structures from confusable objects into our shape model allows for the identification of *anti-shapes* since our weighting scheme is capable of assigning negative weights to those model parts, thus increasing the discrimination capabilities. In the next iteration the expanded model is again applied for target point localization on the training corpus, and new weights are estimated using the described method. The iterative training procedure stops when the localization error on all training images is below a given threshold or if all training images have been used for model generation.

2.3 Multi-Level Approach

A high performance detection of very small structures, like the pupil, can only be achieved in high resolution images. However, a clear drawback of using the highest available resolution level lies in a large processing time and memory demand, which hampers the utilization of the method in a practical application. Additionally, the feature extraction in high resolution images will produce many noisy details which may mislead the localization procedure. Therefore, a reasonable tradeoff between the level of detail, required for a reliable localization, and the necessary suppression of irrelevant structures is required. To this end, the DGHT framework uses a coarse-to-fine strategy based on a Gaussian image pyramid as described in [RKL+11].

The procedure begins with a low resolution image, which assures fast processing times and provides the most relevant structures for a coarse orientation with an appropriately trained GHT model. Since a reliable localization with high accuracy is not possible on this level, an image extract is cut around the detected point and is further processed on an increased resolution (see Figure 2). The refined search region has half of the size of the original image and twice of the resolution such that the number of pixels stays approximately the same. The procedure can be repeated several times, each time using a level-specific model for the object localization, cutting an image extract of half of the previous size, and doubling the resolution. Due to the gradual increase of the resolution, the method can be viewed as a zooming procedure.

2.4 Linear Model Interpolation

To further increase the robustness of the localization framework, different landmark detectors, e.g., for the right and left eye, can be combined with prior knowledge about their expected distance vector within the face. A straightforward technique for combining different knowledge sources into a single distribution, which has, for example, been used in the field of natural language modeling [KS93], is the linear model interpolation

$$p_\Omega(c_i|X_n) = \sum_{k=0}^{K} \omega_k \cdot p_k(c_i|X_n). \tag{6}$$

Here, a parameter ω_k from the set $\Omega = \{\omega_0, \omega_1, ..., \omega_K\}$ can be interpreted as the probability that the model k produces the observation c_i, whereas $0 \leq \omega_k \leq 1$ and $\sum_k \omega_k = 1$. Alternative combination schemes, like the log-linear interpolation [Kla98], may also be used for this purpose but will not be considered in this contribution. Note that the different probability models $p_k(c_i|X_n)$ address the same task, e.g., right eye detection, but may be obtained in different ways. They can for example be trained on different datasets or by using varying parameterizations. Another interesting way of obtaining alternative models, investigated in this contribution, is to (1) train detectors for other landmarks in the face l_k for $k > 0$ and (2) to employ prior knowledge about their distance to the target point c_i. Note that l_0 is used to denote the target landmark. To this end we assume for model k

$$p_k(c_i|X_n) = \sum_{j=1}^{I} p(c_i, c_j^k|X_n) = \sum_{j=1}^{I} p(c_i|c_j^k, X_n) \cdot p(c_j^k|X_n), \forall k > 0, \tag{7}$$

where c_j^k denotes a potential position c_j of landmark k in the Hough space. The distribution $p(c_j^k|X_n)$ is obtained by applying a GHT with a model, specifically trained for landmark l_k. The second distribution, $p(c_i|c_j^k, X_n)$, is assumed to be independent of the feature image X_n and therefore results in the prior probability $p(c_i|c_j^k)$ of the target point c_i, given the position of landmark l_k. This model can be obtained from the annotated landmark positions in the training dataset and has been estimated by $p(\mathbf{c}_i - \mathbf{c}_j^k)$ in order to compensate for different face positions. Here, $\mathbf{c}$ denotes the vector to the center of cell c.

For a first test of this approach, the detection of one eye l_0 is combined with a detector for the other eye l_1, resulting in the following simplified distribution.

$$\begin{aligned} p_\Omega(c_i|X_n) &= \omega_0 \cdot p_0(c_i|X_n) + \omega_1 \cdot p_1(c_i|X_n) \\ &= \omega_0 \cdot p_0(c_i|X_n) + \omega_1 \cdot \sum_{j=1}^{I} p(\mathbf{c}_i - \mathbf{c}_j^1) \cdot p(c_j^1|X_n) \end{aligned} \tag{8}$$

For example, if l_0 denotes the right eye, $p_0(c_i|X_n)$ is obtained by using the standard DGHT model for right eye detection, while $p(c_j^1|X_n)$ is a distribution obtained from a DGHT-based left eye detection procedure. Note that the free parameters Ω in Equation (8) should ideally be optimized on training data but have been set to equal values in this first attempt.

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

3 Data

For training as well as evaluation, the publicly available PUT face database [KFS08] is used, which provides 9971 images from 100 subjects, a sufficiently large amount of data for the experiments. The high image resolution of 2048×1524 pixels unveils various fine structures and therefore enables the modeling of minute details. The images were taken in front of a uniform background and under controlled illumination conditions in five different series with various head poses to provide a large variability (see Figure 3). Furthermore, an additional frontal view subset with 22 images per subject is also provided.

The database contains 30 landmarks amongst other both eyes, nose, mouth or lips, but these are only set if the corresponding landmark is visible. Consequently, only images with annotations for the left and right eye center were used in the described experiments.

The training was performed on a dataset of 600 images, which were randomly chosen by selecting 10 images from 60 subjects. The evaluation was performed on the images of the remaining 40 subjects. The whole evaluation corpus comprises 3830 images and the frontal view subset 869 images.

4 Experimental Setup

4.1 Baseline System

For the work presented here, a 64 bit system with Intel Xeon W3520 with 2.66 GHz and 24 GB Ram was used. For our experiments, we use the DGHT-Framework with a multi-level approach (see Section 2.3) with 6 zoom levels and the iterative training procedure described in Section 2.2.2. For each zoom level a constant image resolution of 64×48 pixels is used, from which feature images are generated by applying the Canny edge detection [Can86]. The localization of the left and right eye was performed by independently using two specifically trained GHT models [HRB+12]. The error measure used in the experiments takes the worse localization result of both eyes and normalizes it with respect to the eye distance as proposed by [JKF01]. Therefore, an error of less than 0.1 / 0.25 corresponds approximately to a location within the iris / eye.

Table 1: Experimental results for eye localization on the evaluation corpus comparing the baseline system with the model interpolation framework. Results are provided for different data sets and fault tolerances.

Setup	Test Set	$\mathbf{e < 0.1}$	$\mathbf{e < 0.15}$	$\mathbf{e < 0.2}$	$\mathbf{e < 0.25}$
Baseline system	frontal view subset	99.1%	99.1%	99.3%	99.4%
	large head pose variation	95.0%	95.4%	96.0%	96.5%
Model interpolation	frontal view subset	99.2%	99.2%	99.4%	99.5%
	large head pose variation	96.6%	97.1%	97.6%	98.1%

4.2 Model Interpolation

For a further improvement of the system performance, localization results of a right and left eye detector are combined by applying Equation 8. To this end, the required distribution over the distance vectors $p(\mathbf{c}_i - \mathbf{c}_j^1)$ was estimated from the annotated eye positions in the training corpus. In order to restrict the processing time, the model combination was performed only for a number of highest peaks in $p_0(c_i|X_n)$ and $p_1(c_i|X_n)$. Furthermore, no combination is done at zoom level 3 and higher since in these image extracts only a single eye and very few confusing structures are visible (see Figure 2). Consequently, approaches to increase the robustness have the highest influence in the first zoom levels.

5 Results

The results of our experiments are listed in Table 1. It can be seen that a comparably high iris localization accuracy of 95% was already achieved with the baseline system, which worked without additional a-priori knowledge or model interpolation. It is, however, noticeable that in 61% of an incorrect localization in the baseline the same eye was found by the left and right eye localization model. Such errors can be easily prevented by the integration of knowledge about the distance vector between both eyes, which is why we extended the baseline system with the linear model interpolation, described in Section 2.4. This measure significantly improved the iris localization accuracy to 96.6% for the corpus containing a large head pose variation. For further illustration of our results, Figure 4 presents an example from the test set with all zoom levels, applied models, feature images and the resulting Hough spaces.

Since after the first zooming step only image extracts are considered (see Figure 2), severe mislocalizations, which lead to a loss of the target point in the subsequent level, can not be corrected by a later step. Therefore, it is interesting to study the distribution of those errors over the different zooming levels, given in Table 2. It can been seen that the number of severe errors in the first three levels could be substantially reduced by the model interpolation.

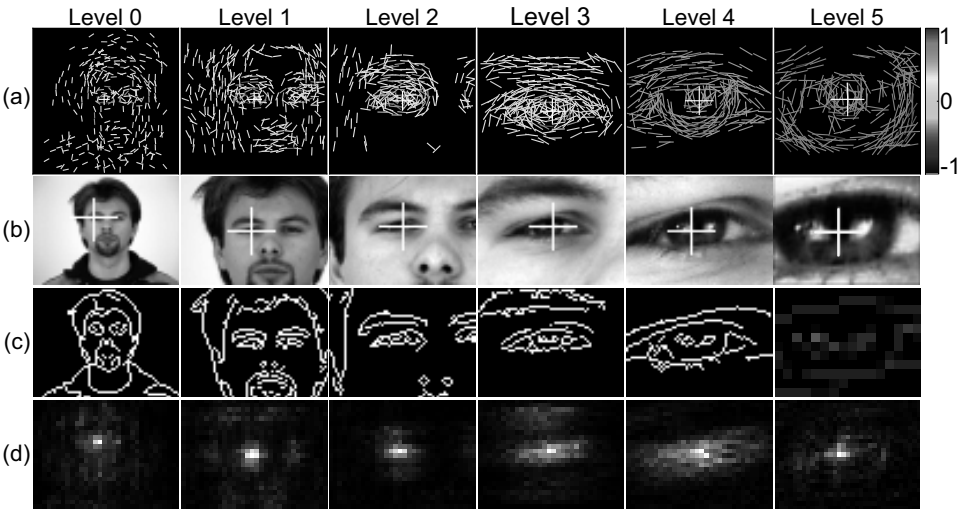

Figure 4: An example over all six levels, showing the model, the localization result, the feature image and the Hough space

Table 2: Number of image extracts with lost target points in each zoom level. Note that no errors may occur in level 0 which contains the whole image.

Setup	Zoom 0	Zoom 1	Zoom 2	Zoom 3	Zoom 4	Zoom 5
Baseline system	-	3	26	101	21	60
Model interpolation	-	0	14	54	21	59

6 Discussion

The experiments clearly show that in frontal view situations a high iris localization rate of 99% can be achieved by using the DGHT methodology. In case of facing the challenge of a large head pose variability, the combination of two localization models significantly improves the absolute accuracy by about 1.5% for all fault tolerances. This has been achieved by reducing the amount of lost target points in the first three zoom levels, as shown in Table 2). This table also demonstrates that zoom level 3 is the most error-prone level since both eyes but hardly any global head contours, supporting the orientation, are visible. The resulting confusion of both eyes could be substantially reduced by the applied right and left eye model interpolation.

The high amount of lost target points at level 5 is related to slightly inaccurate annotations (see figure. 5). Those inaccuracies naturally have the strongest impact on the smallest image extracts and play an important role already during the training phase. Here, the slight random character of the annotations prevents a good discrimination between important and unimportant structures which is why an optimal model cannot be generated.

For comparison of our system with a state-of-the-art method we refer to [KS10], where

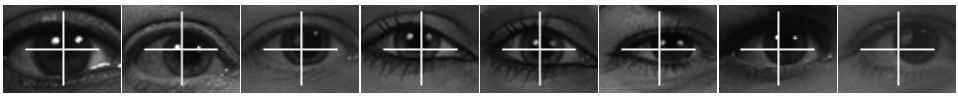

Figure 5: Illustration of image extracts and annotations in the last zooming step

the PUT face database has also been used for training and evaluation. Here, Haar-like features are utilized in a boosted cascade of classifiers [VJ04] for face detection as well as for the subsequent eye localization within the detected face region. This method reached a localization precision of 94% for a large head pose variation and 99% in case of the frontal-view subset. Especially for the images with a high variability of the head pose, the method of interpolated DGHT models performs significantly better without exploiting manually defined constraints and heuristics.

7 Conclusion

In this contribution, we propose the usage of the Discriminative Generalized Hough Transform (DGHT) for eye localization. The DGHT is a general method for the detection of arbitrary objects with well-defined shape, which was already successfully applied for medical image analysis. Since the iterative training method simply requires a set of images with annotated target points, the DGHT can be easily adapted to new localization tasks without the need of utilizing task-specific constraints or expert knowledge.

Additionally, a further extension was presented here to efficiently incorporate several specifically trained landmark detectors in a single DGHT framework by a robust linear model interpolation. A first experiment, combining a left and right eye model, significantly improved the localization rate and outperformed state-of-the-art experimental results, achieved by [KS10] on this dataset. Future work will focus on enhancing the robustness of the proposed method by incorporating a larger number of landmark detectors as well as on dealing with non-uniform backgrounds, which can be expected in real-world conditions.

Acknowledgments. This work is partly funded by the Innovation Foundation Schleswig-Holstein under the grant 2010-90H .

References

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

[Bey98] P. Beyerlein. Discriminative model combination. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 481–484, 1998.

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

[CCS04] D. Cristinacce, T. Cootes, and I. Scott. A multi-stage approach to facial feature detection. In *British Machine Vision Conference (BMVC)*, pages 277–286, 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)*, pages 278–281, 2004.

[GL09] J. Gall and V. Lempitsky. Class-specific hough forests for object detection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1022–1029, 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.

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

[JK92] B.H. Juang and S. Katagiri. Discriminative learning for minimum error classification. *IEEE Transactions on Signal Processing*, 40(12):3043–3054, 1992.

[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)*, pages 90–95, 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)*, pages 379–388, 2008.

[Kla98] D. Klakow. Log-linear interpolation of language models. In *Conference on Spoken Language Processing (ICSLP)*, 1998.

[KS93] R. Kneser and V. Steinbiss. On the dynamic adaptation of stochastic language models. In *Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 586–589, 1993.

[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.

[KTT99] A. Kassim, T. Tan, and K. Tan. A comparative study of efficient generalised Hough transform techniques. *Image and vision computing*, 17(10):737–748, 1999.

[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.

[Oka09] R. Okada. Discriminative generalized hough transform for object dectection. In *Conference on Computer Vision (ICCV)*, pages 2000–2005, 2009.

[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)*, pages 67–70, 2011.

[RLS+11] H. Ruppertshofen, C. Lorenz, S. Schmidt, P. Beyerlein, Z. Salah, G. Rose, and H. Schramm. Discriminative Generalized Hough transform for localization of joints in the lower extremities. *Computer Science-Research and Development*, 26(1):97–105, 2011.

[Sch00] A. Schubert. Detection and tracking of facial features in real time using a synergistic approach of spatio-temporal models and generalized Hough-transform techniques. In *International Conference on Automatic Face and Gesture Recognition (FG)*, pages 116–121, 2000.

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

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